C02029: Doctor of Philosophy
PhD Thesis: Analytics
December 2023

Digital Twin-Oriented Complex Networked Systems

Student: *Jiaqi Wen*

Principal Supervisor: *Katarzyna Musial*

Co-supervisor: *Bogdan Gabrys*

School of Computer Science
Faculty of Engineering & IT
University of Technology Sydney
NSW 2007, Australia

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Jiaqi Wen, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 21/05/2024

Abstract

This research thesis proposes a concept of Digital Twin-Oriented Complex Networked Systems (DT-CNSs) and offers a novel theoretical framework and its implementation for modelling of Complex Networked Systems towards an ultimate goal — a Digital Twin (DT) that can accurately reflect reality. As the first step towards DT-CNSs, a comprehensive literature review on Complex Networked Systems is conducted and a conceptual modelling framework that consists of five generations of DT-CNSs with increasing complexity levels towards a DT is proposed. Under the developed generations of modelling, we explore one of the possible development paths, Digital Twin-Oriented Social Network Simulator (DT-SNS) which is an instantiation of DT-CNS in the context of social networks. This is achieved by a topical review on the current state-of-the-art of Social Network Simulators (SNSs) and an experiment on developing the SNS towards the DT-SNSs. This review reveals a natural development path for DT-CNSs in the context of social networks — to increase the heterogeneity of nodes' features (characteristics of each specific node) and their preferences to create relationships while allowing the CNSs to evolve with the preference changes under the influence of epidemic transmission. Accordingly, we propose an extensible DT-CNS modelling framework based on heterogeneous node features and interaction rules that characterise nodes' preferences for connecting with others. We further improve the expressive power of node features in the proposed DT-CNS models by introducing heterogeneous node feature representation principles. This involves representing features with crisp feature values and fuzzy sets, each describing the objective and the subjective inductions of the nodes' features and feature differences. As the next step, we extend this modelling framework for evolutionary DT-CNSs by introducing changeable interaction rules. This involves the heterogeneous preference mutation mechanisms that model nodes' adaptive decisions on social contact under the impact of epidemic outbreak. Our empirical analysis of social network simulations and epidemic spreading process on these networks reveals the influence of nodes' features and interaction rules on the network formation and epidemic spread. The increasing complexity of process and structure while, at the same time, including interrelation between those two enables more robust modelling and better preserves necessary heterogeneous patterns observed in reality. This research project can be further extended towards a DT by incorporating more complex information extracted from the real world.

Keywords: Complex Networked Systems; Digital Twins; Social Networks; Network Dynamics; Heterogeneous Nodes Attributes and Features; Dynamic Processes on Networks

Acknowledgments

I would like to express my gratitude to all those who helped me during my PhD study.

I want to express my most profound appreciation to my supervisor, Professor Katarzyna Musial and co-supervisor, Professor Bogdan Gabrys, for giving me an opportunity to study in the Complex Adaptive Systems Lab and work on the exciting research topic, Digital Twin-Oriented Complex Networked Systems. I am incredibly grateful for their excellent guidance, comments, modifications and support during my PhD study. Their keen insights and kind help motivates the progress with my research journey.

I also want to thank my family, who supported me with love and care since I was born in all stages of my life. I am incredibly lucky to be surrounded by the most kind, lovely and caring people in the world. My true gratitude is beyond any word's description.

This research project was supported by the Australian Research Council, Dynamics and Control of Complex Social Networks under Grant DP190101087.

List of Publications

Below, a list of the publications created out of this research project is presented.

- J. Wen, B. Gabrys and K. Musial, "Toward Digital Twin Oriented Modeling of Complex Networked Systems and Their Dynamics: A Comprehensive Survey," in IEEE Access, vol. 10, pp. 66886-66923, 2022, doi: 10.1109/ACCESS.2022.3184801 - this paper includes the work on the review of the Complex Networked Systems modelling paradigms, proposition of the modelling framework for Digital Twin-Oriented Complex Networked Systems and assessment framework [1].

- J. Wen, B. Gabrys and K. Musial, "Review and Assessment of Digital Twin-Oriented Social Network Simulators," in IEEE Access, vol. 11, pp. 97503-97521, 2023, doi: 10.1109/ACCESS.2023.3312129.-this paper includes the work on the review and assessment of social network simulators, proposition of the Digital Twin-Oriented Social Network Simulator based on heterogeneous node features [2]

- J. Wen, B. Gabrys, and K. Musial, "Towards Digital Twin-Oriented Complex Networked Systems: Introducing heterogeneous node features and interaction rules," in Plos One, vol. 19(1), pp. e0296426, 2024, doi: 10.1371/journal.pone.0296426. -this paper includes the work on the extendable modelling framework for Digital Twin-Oriented Complex Networked Systems based on heterogeneous node features and interaction rules [3].

- J. Wen, B. Gabrys, and K. Musial, "Heterogeneous Feature Representation for Digital Twin-Oriented Complex Networked Systems," in arXiv preprint arXiv:2309.13229, 2023, doi: 10.48550/arXiv.2309.13229. -this paper includes the work on the feature representation in Digital Twin-Oriented Complex Networked Systems based on heterogeneous feature representation principles [4].

- J. Wen, B. Gabrys, and K. Musial, "Heterogeneous Feature Representation for Digital Twin-Oriented Social Network Simulators," Working Paper, 2024. -this paper

includes the work on the feature representation in Digital Twin-Oriented Complex Networked Systems based on heterogeneous feature representation principles. This work has been accepted by IEEE WCCI.

- J. Wen, B. Gabrys, and K. Musial, "Evolutionary Digital Twin-Oriented Complex Networked Systems Driven by Node Features and the Mutation of Feature Preferences," Working Paper, 2024. -this paper includes the work on the network evolution under the impact of dynamic process in Digital Twin-Oriented Complex Networked Systems based on heterogeneous feature preference mutation mechanisms. This work has been accepted by Plos One.

- J. Wen, B. Gabrys, and K. Musial, "DTCNS: A Python Toolbox for Digital Twin-Oriented Complex Networked Systems," Working Paper, 2024. -this paper includes the software work: "DTCNS" Python toolbox for Digital Twin-Oriented Complex Networked Systems (available at https://github.com/UTS-CASLab/DT-CNS). This work is now under review by SoftwareX.

Table of Contents

List of Publications vii

List of Figures xiii

List of Tables xxvii

1 Introduction 1

1.1 Background 1
1.2 Research Aim 3
1.3 Research Questions, Objectives and Methodology 4
1.4 Significance and Contributions 15
1.5 Thesis Organisation 17

2 Literature Review of Modelling Complex Networked Systems and Their Dynamics 19

2.1 Introduction 19
2.1.1 Background 19
2.1.2 Research design 21
2.2 The aim of modelling Complex Networked Systems 22
2.2.1 Prerequisites of modelling 23
2.2.2 Currently studied CNS modelling aims 24
2.2.3 Digital Twin: an ultimate goal 27
2.3 How to represent information in a form of a network 28
2.3.1 How to represent networked data 28
2.3.2 How to generate networks using different modelling paradigms . . 36
2.4 How to model the dynamics in networked systems 46
2.4.1 An overview of dynamics and their interrelations 46
2.4.2 Modelling Dynamics of Networks 51

2.4.3 Modelling Dynamic Processes 58
2.4.4 Combination of the network and process dimensions 63
2.4.5 Control mechanisms on CNS 65
2.5 How do we approach the ultimate goal? 67
2.5.1 What have we done so far? 67
2.5.2 How far are the CNSs, as they are currently modelled, from DTs? 71
2.5.3 How can we go further? 80
2.6 Conclusions 83

3 Review and Assessment of Social Network Simulators 87
3.1 Introduction 87
3.2 Current state-of-the-art of Social Network Simulators 90
3.2.1 Modelling prerequiste 90
3.2.2 Modelling generations 91
3.2.3 Complexity dimensions 93
3.2.4 Assessment of SNSs 95
3.3 Towards Digital Twin Oriented Social Network Simulator 96
3.3.1 An Extensible Social Network Simulator 96
3.3.2 Optimisation towards a Digital Twin 98
3.4 Results and Assessment 101
3.4.1 Performance of the proposed social network simulator 102
3.4.2 Complexity of the social network simulation 106
3.5 Conclusions 119

4 Heterogeneous Node Features and Interaction Rules for DT-CNS model 123
4.1 Introduction 123
4.2 Background 125
4.3 Digital Twin Orientated Complex Networked System 127
4.3.1 An extendable DT-CNS modelling framework 127
4.3.2 An Evaluation Protocol on Faithfulness 131
4.3.3 Optimisation towards a Digital Twin 134
4.4 Results 138
4.4.1 Simulation-based Networks 139
4.4.2 Simulation-based Processes 148
4.4.3 Faithfulness 160
4.4.4 Summary of results 163

4.5 Conclusion 164

5 Heterogeneous Feature Representation for DT-CNS model 167
5.1 Introduction 167
5.2 Digital Twin-Oriented Complex Networked Systems built on fuzzy representation of features 170
5.2.1 Feature representation 171
5.2.2 Feature preference 173
5.2.3 Network formation 174
5.2.4 Epidemic transmission 175
5.3 Empirical Analysis 176
5.3.1 Data Description 176
5.3.2 Network Models Initialisation 178
5.3.3 Model Selection considering Features and Rules 182
5.3.4 Model Extension With Flexible Fuzzy Representation Principles . 190
5.3.5 Network Representation 207
5.3.6 Disaster Resilience 226
5.4 Conclusions 232

6 Node Features and the Mutation of Feature Preferences for DT-CNS model. 235
6.1 Introduction 235
6.2 Modelling Framework 237
6.2.1 Complex Networked System Components 237
6.2.2 CNS Dynamics 242
6.3 Experiment Design, Results and Analysis 244
6.3.1 Initial Parameters 244
6.3.2 Infection Patterns 245
6.3.3 Interaction Patterns 247
6.3.4 Interaction Utilities 266
6.4 Conclusion 270

7 Discussion and Conclusion 273
7.1 Conclusion 273
7.2 Future Work 278

A Appendix 281

A.1 Extendable Social Network Simulators 281
A.2 Optimised Parameters for DT-SNSs 283

B Appendix **285**
B.1 Parameter set-ups for preferences and weights of preferences of the DT-CNSs in this study 285
B.2 Distance between the network patterns of simulated networks and the target network 286

C Appendix **289**
C.1 Parameter bounds' set-ups in the sensitivity analysis and optimisation process 289

D Appendix **291**
D.1 Networks and their statistics in the first 20 simulation steps 291
D.1.1 Social Capital Limit at 5 292
D.1.2 Social Capital Limit at 10 297
D.1.3 Social Capital Limit at 15 302
D.1.4 Social Capital Limit at 20 307

Bibliography **313**

List of Figures

Figure **Page**

1.1 An example of Complex Networked Systems and their constituent dynamics in social network space, including the social networks with evolving structures and attributes in Fig. (a) and the epidemic spreading process on social networks in Fig.(b). The social network evolution in Fig. (a) is driven by the mutation of nodes' preferences for heterogeneous features. In 2015, nodes prefer young, female and similar nodes; in 2020, they prefer young, female and dissimilar others; in 2025, the nodes prefer old, male and dissimilar nodes. The aim of this thesis is to model CNSs with increasing complexity levels concerning these CNS dynamics and their interrelations to better preserve necessary heterogeneous patterns observed in reality. 2
1.2 Digital Twin Oriented Modelling of Complex Networked Systems 7
1.3 Conceptual modelling framework for DT-CNSs 8
1.4 Conceptual modelling framework and research directions for DT-SNSs 9
1.5 Conceptual modelling framework and research directions for DT-CNSs 9
1.6 Thesis organisation considering the research motivations for each chapter. . . 17

2.1 Four dimensions of complexity related to a network's topology. 30
2.2 Observability and complexity with corresponding relations 33
2.3 Different types of networks and their components. 35
2.4 An overall view of modelling paradigms 44
2.5 Generations of modelling framework 48
2.6 Generation 5: the temporal dynamic processes on temporal networks, with interrelations and the real-time acquisition and feedback between the SNS and reality. 50
2.7 The plot of unified assessment criterion for CNS model and its distance to a DT 72
2.8 The plot of assessing the distance of a CNS model to a DT 76
2.9 The complexity metric of a CNS model 77

2.10 The model performance metric of a CNS model 78

3.1 The performance of optimised SNSs based on similarity and efficiency, including zero feature–based SNS (a), real feature–based SNS (b), simulate feature–based SNS (c) and hybrid feature–based SNS (d) 103
3.2 The performance of SNS based on the comprehensive composite performance index (a) and the efficiency (b) 104
3.3 The performance of SNS respectively based on the global and the local composite performance indexes. 105
3.4 The number of edges and the density of social network simulations over the iterations 107
3.5 The modularity and the assortivity of social network simulations 108
3.6 The degree distribution of networks 110
3.7 The average value and the JS divergence of degree distribution. 110
3.8 The shortest path length distribution of networks, where the path lengths value between 1 and 34 as we assume no self-links and replace the unavailable (infinite) paths with the lengths of 34. 112
3.9 The average value and the JS divergence of shortest path length distribution. 113
3.10 The clustering coefficient distribution of network simulations 114
3.11 The average value and the JS divergence of clustering coefficient distribution. 114
3.12 The four types of triadic closure in the network simulations 115
3.13 The number of the four types of triadic closure in the network simulations . . 117
3.14 Z–score values of triads in the social network simulations 118

4.1 The procedures of building Digital Twin-Oriented Complex Networked Systems.135
4.2 The age distributions used in the experiments. 140
4.3 The Hill number for each age distribution. 140
4.4 The age and degree distributions of social networks generated by DT-CNS models based different features and rules. 142
4.5 The age and clustering coefficient distributions of social networks generated by DT-CNS models based different features and rules 144
4.6 The age and shortest path length distributions between connected nodes of social networks based on different features and rules 145
4.7 The infection occurrence within certain distance to the seed given different transmissibilities. 151

4.8 The infection occurrence over time with different transmissibilities given an exposure. 152
4.9 The $PaR(1,1)$ and $PaR(2,2)$ values with different transmissibilities, age distributions and rules of network formation. 154
4.10 The PaR(1,1) for each age group given different transmissibilities, age distributions and the $P+$ rule. 155
4.11 The PaR(1,1) for each age group given different transmissibilities, age distributions and the $P-$ rule. 157
4.12 The PaR(1,1) for each age group given different transmissibilities, age distributions and the $H+$ rule. 158
4.13 The PaR(1,1) for each age group given different transmissibilities, age distributions and the $H-$ rule. 159
4.14 The PaR(1,1) for each age group given different transmissibilities, age distributions and the PH rule. 161
4.15 The runtime of simulation-based DT-CNS models given different features and rules. 162

5.1 The age & sex-structured population profiles in 2005 of countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy (e) Luxembourg and (f) Poland. 177
5.2 The simulated age and sex-distributions in 2005 for countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. . 180
5.4 The fuzzy representation of age (age difference). 181
5.3 The diversity of the population in each country, as measured by the Hill numbers based on age, sex and age&sex groups. 181
5.5 The target social contact matrix ((a) target) and the recreated social contact matrices of Belgium, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 184
5.6 The target social contact matrix ((a) target) and the recreated social contact matrices of Finland, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 185
5.7 The target social contact matrix ((a) target) and the recreated social contact matrices of Germany, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 186
5.8 The target social contact matrix ((a) target) and the recreated social contact matrices of Italy, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 187

5.9 The target social contact matrix ((a) target) and the recreated social contact matrices of Luxembourg, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 188

5.10 The target social contact matrix ((a) target) and the recreated social contact matrices of Poland, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models. 189

5.11 The searching space of fuzzy sets for age and age difference. In the experiments, both the age and age difference initialise the fuzzy sets based on the same parameter set-ups. Therefore, this figure 192

5.12 The EU distance achieved with different settings of fuzzy sets (Fuzzy Set Number (H) for age difference on the y-axis and Fuzzy Set Number (P) for age on the x-axis) in experiments on (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. 193

5.13 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$ (a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Belgium. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h) and parameters of the third fuzzy set of age difference (μ_3^h and σ_3^h) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 197

5.14 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Finland. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h) and parameters of the fourth fuzzy set of age difference (μ_4^p and σ_4^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 199

5.15 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$ (a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Germany. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameters of the third fuzzy set of age (μ_3^p and σ_3^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 201

5.16 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Italy. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the first fuzzy set of age (μ_1^p and σ_1^p) and parameters of the second fuzzy set of age (μ_2^p and σ_2^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 202

5.17 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Luxembourg. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameter for the first fuzzy set of age (μ_1^p), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameter for the fifth fuzzy set of age (μ_5^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 204

5.18 The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Poland. In (a), parameter for the first fuzzy set of age difference (σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameter for the fifth fuzzy set of age (μ_5^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e). 206

5.19 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Belgium based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature). 209

5.20 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Finland based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature). 210

5.21 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Germany based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within [0,1]) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).. 211

5.22 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Italy based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within [0,1]) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).. 212

5.23 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Luxembourg based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within [0,1]) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).. 213

5.24 The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Poland based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).. 215
5.25 The target social contact matrices of (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. 217
5.26 The social contact matrices recreated with the $HN-A_f-S$ model built on calibrated fuzzy sets for (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. 218
5.27 The age and degree distributions in Belgium for (a) females and males, (b) females and (c) males. 219
5.28 The age and degree distributions in Finland for (a) females and males, (b) females and (c) males. 219
5.29 The age and degree distributions in Germany for (a) females and males, (b) females and (c) males. 219
5.30 The age and degree distributions in Italy for (a) females and males, (b) females and (c) males. 220
5.31 The age and degree distributions in Luxembourg for (a) females and males, (b) females and (c) males. 220
5.32 The age and degree distributions in Poland for (a) females and males, (b) females and (c) males. 220
5.33 The age and clustering coefficient distributions in Belgium for (a) females and males, (b) females and (c) males. 221
5.34 The age and clustering coefficient distributions in Finland for (a) females and males, (b) females and (c) males. 222
5.35 The age and clustering coefficient distributions in Germany for (a) females and males, (b) females and (c) males. 222
5.36 The age and clustering coefficient distributions in Italy for (a) females and males, (b) females and (c) males. 222

5.37 The age and clustering coefficient distributions in Luxembourg for (a) females and males, (b) females and (c) males. 223

5.38 The age and clustering distributions in Belgium for (a) females and males, (b) females and (c) males. 223

5.39 The age and shortest path length distributions in Belgium for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 225

5.40 The age and shortest path length distributions in Finland for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 225

5.41 The age and shortest path length distributions in Germany for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 225

5.42 The age and shortest path length distributions in Italy for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 225

5.43 The age and shortest path length distributions in Luxembourg for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 226

5.44 The age and shortest path length distributions in Poland for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males. 226

5.45 The proportion of infections to the overall population (90 nodes) within the specific distance (a) to the seeded node within specific simulation steps (b). . 228

5.46 The PaR within the specific distance to the seeded node over the specific time steps in (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. 230

5.47 The simulated age & sex & infection-distributions in 2005 of countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. The arrow points to the specific age & sex group that includes the seed node for epidemic spread (Tab. 5.7). We also denote each age&sex group with an infection ratio in the first step of epidemic spread to indicate the infections risk $PaR(1,1)$. 231

6.1 The infection occurrences in 100 iterations (a) and first 20 iterations (b) in the evolving social networks that are respectively driven by inactive, ignorant, egocentric, cooperative and collaborative mutation styles. 246

6.2 The number of interactions in the 100 simulations steps (a) and first 20 iterations (b) in the evolving social networks that are respectively driven by inactive, ignorant, egocentric, cooperative and collaborative mutation styles. 248

6.3 The interaction numbers' changes between nodes in each health statuses in evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 250

6.4 The evolving social networks driven by collaborative preference mutation styles under the impact of different social capital limits. 251

6.5 The evolving social networks driven by ignorant preference mutation styles under the impact of different social capital limits. 252

6.6 The evolving social networks driven by different mutation styles under the social capital limit at 10. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 253

6.7 The degree distribution of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 254

6.8 The average node degree for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 256

6.9 The Euclidean distances between the degree distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance. 257

6.10 The clustering coefficient distribution of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 258

6.11 The average node clustering coefficient for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 260

6.12 The Euclidean distances between the clustering coefficient distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance. 261

6.13 The shortest path length distributions of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 263

6.14 The average shortest path length for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 264

6.15 The Euclidean distances between the shortest path length distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance. 265

6.16 The total interaction utilities of nodes in the evolving social networks (a) and their cumulative values over time (b). 267

6.17 The average utility of infected nodes and healthy nodes in evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles. 269

7.1 Discussion of the research motivations and the research contributions of each chapter 274
7.2 The research directions and research gaps. 275

D.1 The evolving social networks driven by inactive nodes in an epidemic outbreak.292
D.2 The evolving social networks driven by ignorant nodes in an epidemic outbreak.293
D.3 The evolving social networks driven by egocentric nodes in an epidemic outbreak.294
D.4 The evolving social networks driven by cooperative nodes in an epidemic outbreak. 295
D.5 The evolving social networks driven by collaborative nodes in an epidemic outbreak. 296
D.6 The evolving social networks driven by inactive nodes in an epidemic outbreak.297
D.7 The evolving social networks driven by ignorant nodes in an epidemic outbreak.298
D.8 The evolving social networks driven by egocentric nodes in an epidemic outbreak.299
D.9 The evolving social networks driven by cooperative nodes in an epidemic outbreak. 300
D.10 The evolving social networks driven by collaborative nodes in an epidemic outbreak. 301
D.11 The evolving social networks driven by inactive nodes in an epidemic outbreak.302
D.12 The evolving social networks driven by ignorant nodes in an epidemic outbreak.303
D.13 The evolving social networks driven by egocentric nodes in an epidemic outbreak.304
D.14 The evolving social networks driven by cooperative nodes in an epidemic outbreak. 305
D.15 The evolving social networks driven by collaborative nodes in an epidemic outbreak. 306
D.16 The evolving social networks driven by inactive nodes in an epidemic outbreak.307
D.17 The evolving social networks driven by ignorant nodes in an epidemic outbreak.308
D.18 The evolving social networks driven by egocentric nodes in an epidemic outbreak.309
D.19 The evolving social networks driven by cooperative nodes in an epidemic outbreak. 310

D.20 The evolving social networks driven by collaborative nodes in an epidemic outbreak. 311

List of Tables

Table **Page**

1.1 Research challenges, questions, objectives and contributions 5
1.2 Summary of CNS modelling generations 8

2.1 Different goals of building CNS models 25
2.2 Summary of CNS modelling generations 51

3.1 Current SNSs and the observability of network components in their simulations 91
3.2 Current SNSs and their generations 92
3.3 Current SNSs and the components of their simulations 94
3.4 Small steps of building up the complexity. 99
3.5 The complexity of the extended SNSs. 102
3.6 The number of nodes and edges, as well as the density of the target network. 108
3.7 The degree distribution of the target network. 109
3.8 The shortest path length distribution of the target network. 111
3.9 The clustering coefficient distribution of the target network. 113
3.10 The definition, numbers and significance profiles of triadic closure of the target network. 116

4.1 Measures of complex patterns emerging from CNS representation. 131
4.2 Measures used to assess performance of simulation, prediction and/or control models of CNS dynamics. 132
4.3 Requirement of a faithful CNS representation and modelling under observability constraints. 133
4.4 Examples related to features and the conditions in the DT-CNS representation and modelling. 136
4.5 The examples of possible rules governing the interactions, seed selection and epidemic transmission in DT-CNS representation and modelling. 137

4.6 Topological information of the networks generated with the DT-CNS models. 147
4.7 The single seed selection based on the largest node degree. 149
4.8 The similarity of degree distributions with the target degree distribution, as measured by the Jensen–Shannon divergence. 162

5.1 Number (No.) of nodes and edges. 178
5.2 Model description. 179
5.3 The similarity and the efficiency of network representation with the network models, as measured by the Euclidean distance (EU distance) between social contact matrices and the Runtime (second). 183
5.4 The EU distance achieved with the respective fuzzy sets selected for each country. 194
5.5 The parameter ranking and selection for the selected fuzzy sets and the resulted model performance evaluated with EU distance. The parameter ranking is determined based on the aggregated ranks of sensitivity indices including $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$. 196
5.6 Topological information of the generated networks. 208
5.7 The starting node of the epidemic spreading process in each country. 227

6.1 The initialised parameters for the DT-CNS models. 245

7.1 Development path of DT-CNS model from generation 1 to generation 5 (the ultimate goal of a DT) with increasing complexity levels, concerning the interaction rules, transmission rules, and the real-world decision-making dynamics that are linked to the generation 5 DT-CNSs in a real-time feedback loop. 279

A.1 SNSs in the current studies 282
A.2 Best solution for the optimisation of zero feature–based SNS considering similarity and efficiency. 283

B.1 The parameters of the DT-CNSs. 286
B.2 The JS divergence between the network patterns of simulated networks and the target network. 287

C.1 Parameter bounds' set-ups in the sensitivity analysis and optimisation process289

D.1 Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5. 292
D.2 Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 5. 293
D.3 Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 5. 294
D.4 Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 5. 295
D.5 Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 5. 296
D.6 Topological information of the network simulations driven by inactive mutation style under a social capital limit at 10. 297
D.7 Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 10. 298
D.8 Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 10. 299
D.9 Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 10. 300
D.10 Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 10. 301
D.11 Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5. 302
D.12 Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 15. 303
D.13 Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 15. 304
D.14 Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 15. 305
D.15 Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 15. 306
D.16 Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5. 307
D.17 Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 20. 308

D.18 Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 20. 309
D.19 Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 20. 310
D.20 Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 20. 311

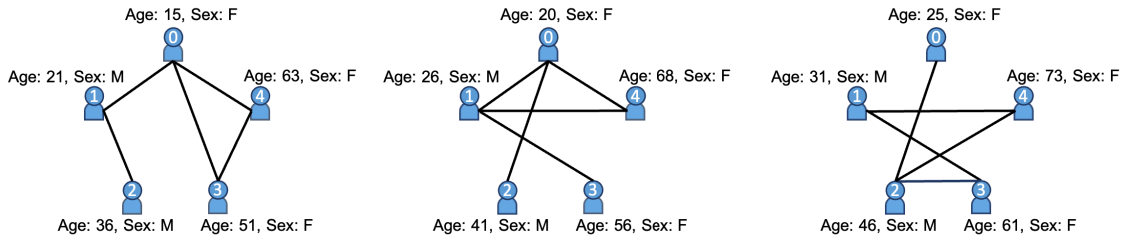
CHAPTER 1

INTRODUCTION

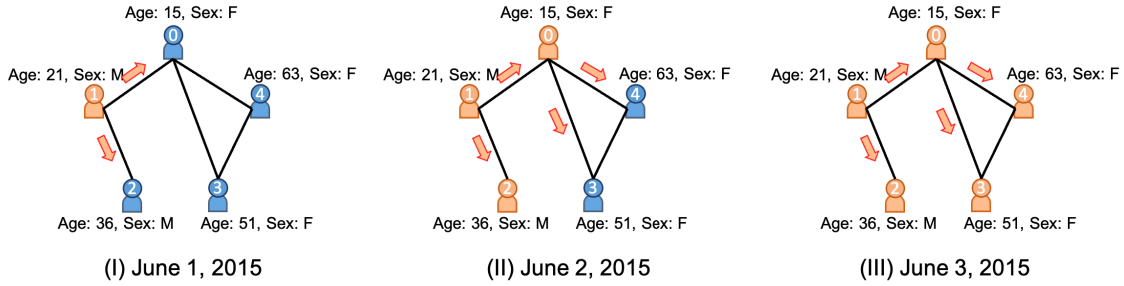
1.1 Background

A complex network is a universal concept used for representation and analysis of complex systems, which can be seen in various application scenarios and studied across disciplines, such as cellular networks in biology [5], supply chain networks in engineering [6] and social networks in sociology [7, 8]. Complex Networked Systems (CNSs) are modelled with the aims of simulation, prediction and/or control, which include solving a wide range of tasks such as link prediction [9–11], anomaly detection [12, 13], mimicking of real systems [14, 15], etc. The CNS models, together with their respective primary aims, are also typically developed within different disciplinary focuses such as epidemiology [16], sociology [7], microeconomics [17]. The choice of modelling paradigms, which can be understood as the approaches employed to model the real state or dynamics of the system, largely depends on the model's aims determined based on the research questions. Therefore, we need to understand what is expected from modelling CNSs, and this would serve as the basis for further discussions of detailed modelling paradigms.

CNS models are composed of networks and dynamics of and on the networks [1]. The networks can be represented as a set of nodes connected with each other via edges where both nodes and edges can have attributes (Fig. 1.1, an example of social networks). The CNS dynamics can be either considered as: (i) dynamic processes on the networks (Fig. 1.1 (b), an example of epidemic spreading process on social networks), which involve studies on spreading phenomena including epidemic processes and information spreading



(a) Social networks composed of nodes (people), edges (social contact) and node attributes (age and sex features), which evolve over time with changeable network structures and attributes (age).



(b) Epidemic spreading process on social networks, which propagates over time from the seed node (first infectious node, denoted by 1 and marked by the red colour) to its neighbours via social contact.

Figure 1.1: An example of Complex Networked Systems and their constituent dynamics in social network space, including the social networks with evolving structures and attributes in Fig. (a) and the epidemic spreading process on social networks in Fig.(b). The social network evolution in Fig. (a) is driven by the mutation of nodes' preferences for heterogeneous features. In 2015, nodes prefer young, female and similar nodes; in 2020, they prefer young, female and dissimilar others; in 2025, the nodes prefer old, male and dissimilar nodes. The aim of this thesis is to model CNSs with increasing complexity levels concerning these CNS dynamics and their interrelations to better preserve necessary heterogeneous patterns observed in reality.

processes [15, 18], or (ii) dynamic networks with evolving structures and attributes (a.k.a. features; Fig. 1.1 (a), an example of evolving social network structures and attributes)) [19, 20]. Studies on CNS models generally start from the representation of a network topology, which describes how nodes relate and connect to each other via edges [21, 22]. Multiple modelling paradigms, such as inner rule-based [23], event-based [24], and agent-based paradigms [25], have been developed and utilised to represent network topologies and sometimes, at a higher complexity level, model the associated network dynamics that drive network topology changes over time. Based on the network representation, some studies on CNS models further investigate how network topologies impact the dynamic processes on the networks [15, 18, 26–30]. This leads to an increasing complexity of modelling CNS considering the interactions and interrelations between a network and

a dynamic process, which involves discussion on CNS modelling towards the highly complex reality and requires a detailed research on the currently existing modelling paradigms in the CNS space.

Over the years, CNSs have been modelled with increasing complexity levels which resulted in more realistic network topologies and evolving dynamics. Current studies change their focus from the least complex CNS models where a single dynamic process takes place on a static network without changes in the network components or the process parameters [31, 32] to more realistic CNS models by representing and modelling not only the structure but also attributes (e.g. node or edges attributes) as well as temporal changes (e.g. addition/removal of nodes and edges) of networks and dynamics associated with networks [33–35]. These research efforts aim to build models that accurately reflect and extend reality, which can be described with a Digital Twin (DT) approach [1].

Digital Twin (DT) is a linked collection of digital artefacts that evolves with the real system along the whole life cycle and integrates currently available knowledge with the purpose of describing behaviour and deriving solutions for the real system [36]. A DT is characterised by the digital (virtual part), the real physical product and the connection between them [37]. DTs, aiming at an accurate reflection and extension of reality, allow us to compare current conditions with historical data to provide meaningful information to assist in decision-making and enable forecasting and feedback on eventualities that have never happened before [38]. DTs are expected to provide an accurate reflection and extension of reality, fulfilling simultaneously multiple modelling aims. There are studies of the Digital Twining of real systems across disciplines, but complex networks area has been marginalised in this development space [39–41]. Therefore, the convergence of Digital Twinning and Complex Networked Systems emerges as an exciting research focus —- Digital Twin-Oriented Complex Networked Systems (DT-CNSs), with the potential to develop CNS models towards a DT with increasing complexity levels.

1.2 Research Aim

In this thesis, a Digital Twin is understood as an ultimate goal for the representation and modelling of the CNSs, which has not been realised yet but is expected to be steadily approached over next years with the ability to take more complex information into account, capture and represent dynamics of and on the networks as well as simulate and interact with the modelled reality in real-time [1]. As presented in section 1.1, CNSs, over the years, have been modelled with increasing complexity levels to approach a DT

of real systems.

DT-CNS models proposed in this thesis model real-world systems and enable to test different what-if scenarios that have never happened before. For example, in the space of social networks, the aim of the DT-CNS is to model interactions between people and dynamics of any processes propagating over the resulting from the interactions social network (Fig. 1.1). The information that is utilised to build DT-CNS includes people's characteristics (e.g. age, gender) and preferences when it comes to who they want to interact with (i.e. preference of people towards specific age of others). The DT-CNS models of social networks in an epidemic outbreak would enable the simulation of real systems impacted by different social intervention measures such as vaccination, social distancing and self-isolation. However, the DT-CNS establishment, considering heterogeneous network components and their interrelated changes over time as well as the DT-CNS model development towards a DT, which incorporates increasing complexity levels, still remain research challenges to be addressed. Therefore, the aim of this thesis is:

- **Aim**: to propose, develop and assess a modelling paradigm called Digital Twin Oriented Complex Networked System (DT-CNS) that incorporates increasing complexity levels dependent on heterogeneous network components and their changes over time.

Under this goal, we identify the research questions and objectives and propose research methodologies to answer these questions in the next section.

1.3 Research Questions, Objectives and Methodology

Overall, we develop research questions related to DT-CNSs and the node features that drive the DT-CNS formation and evolution (Tab. 1.1). This research project addresses the complexity of DT-CNS models that results from heterogeneous features and rules related to those features. Given the research questions, we propose the Research Objectives (**RO**s) for the systematic execution of the research towards the project's goal – to build and assess the DT-CNS. In addition, we discuss the project's Research Methodology (**RM**), which enables to answer the Research Questions (**RQ**s) and achieve the Research Contributions (**RC**s) (as presented in Table 1.1 and section 1.4). Throughout the research project, we employ the Design Science Research Methodology (DSRM) approach to describe the the research methodology.

Table 1.1: Research challenges, questions, objectives and contributions

Research Challenges	Research Questions	Research Objectives	Contributions
RCh1: Compare and assess existing CNSs modelling paradigms in the context of a DT paradigm.	RQ1: How to compare modelling paradigms of CNS and measure their distances to a DT?	RO1.1. Conduct a comprehensive critical review and comparative analysis of CNS models.	RC1.1: A comprehensive and comparative review of existing CNSs modelling paradigms.
		RO1.2: Propose a conceptual modelling framework for DT-CNS.	RC1.2: New conceptual modelling framework for DT-CNSs.
		RO1.3: Propose an assessment approach to conceptually compare and evaluate CNS models.	RC1.3: Novel assessment framework for evaluation of models developed within DT-CNS modelling framework.
RCh2: Build and evaluate DT-SNSs based on heterogeneous node features for a faithful network representation.	RQ2: How to build and evaluate DT-SNS that accommodates heterogeneous features of nodes?	RO2.1: Develop a DT-Oriented Social Network Simulator (DT-SNS) that enables social network simulation based on heterogeneous node features.	RC2.1: DT-SNS that enables to faithfully represent static network topology with a given complexity level.
		RO2.2: Evaluate proposed DT-SNS.	RC2.2: A case study on the modelling and assessment of DT-SNS with different levels of structural complexity.
RCh3: Build a DT-CNS modelling framework based on node features and feature preferences, as well as develop evaluation framework for its assessment.	RQ3: How to build and evaluate DT-CNS considering node features and feature preferences?	RO3.1: Develop an extensible DT-CNS based on heterogeneous node features and rules of feature preferences.	RC3.1: Construction of DT-CNS based on heterogeneous node features and feature preference principles.
		RO3.2: Develop evaluation protocol for DT-CNS based on node features and feature preferences.	RC3.2: Evaluation protocol to assess the faithfulness of DT-CNS, considering similarity, efficiency and reproducibility.
		RO3.3: Evaluate the proposed DT-CNS considering node features and feature preferences	RC3.3: A case study on DT-CNS models built on different node features, feature preferences and transmissibility of epidemic spread.
RCh4: Build a DT-CNS based on uncertain node features and feature preferences.	RQ4: How to represent uncertain features with DT-CNS models?	RO4.1: Develop flexible rules of feature representation for the DT-CNS models.	RC4.1: Feature representation with fuzzy sets for uncertain features.
		RO4.2: Validate and evaluate the influence of different ways of feature representation on DT-CNS models.	RC4.2: DT-CNS models built on different feature representation principles.
		RO4.3: Evaluate the proposed DT-CNS models built on different feature representation principles	RC4.3: A case study on DT-CNS models built with increasingly flexible rules of feature representation.
RCh5: Build an evolutionary DT-CNS driven by preference mutation, where both network and process can change and those changes are interrelated.	RQ5: How to build and evaluate the evolutionary DT-CNS models driven by mutation of feature preferences?	RO5.1: Build an evolutionary DT-CNS with changeable feature preferences and interrelated DT-CNS components.	RC5.1: New feature preference mutation mechanism considering heterogeneous mutation styles
		RO5.2: Build an evolutionary DT-CNS based on the changes of feature preferences	RC5.2: Evolutionary DT-CNS models with evolving networks, dynamic processes and the interrelations between their dynamics.
		RO5.3: Evaluate the proposed evolutionary DT-CNS.	RC5.3: A case study on the evolutionary DT-CNS models in an epidemic outbreak based on different preference mutation styles.

There is a considerable literature about CNS models and various surveys reviewing and exploring the topic from different perspectives, including application areas [42], CNS structures related to network topologies [21, 22] and the dynamics [15, 18–20], and modelling techniques that have been applied to build CNS models [19, 43, 44]. However, none of these studies integrate and overview the current, relevant state-of-the-art from multiple disciplines and inform future research directions and challenges in DT-CNSs space. In addition, the comprehensive review of modelling paradigms of CNSs with the ultimate goal of building a DT requires to conceptually compare the current studied CNS models, while having in mind how far they are from the DTs. How to build a universal modelling framework that enables the review and assessment of any CNS, especially in the context of a DT, remains an unstudied area. We also need a unified criteria to assess the DT-CNS while considering both the requirements of building a good CNS and the conditions of a ready DT. Therefore, to address the first research challenge (**RCh 1**) of comparing and assessing CNS modelling paradigms in the context of a DT paradigm, we propose the **RQ1** and identify three **RO**s concerning the review and evaluation of DT-CNS models.

- **RQ1**: How to compare modelling paradigms of CNS and measure their distances to a DT?

    1. **RO1.1**: Conduct a comprehensive critical review and comparative analysis of CNS.
    2. **RO1.2**: Propose a conceptual modelling framework for DT-CNS.
    3. **RO1.3**: Propose an assessment approach to conceptually compare and evaluate CNS models.

- **RM1**: We answer **RQ1** in chapter 2 by a comprehensive survey on CNS models as they approach their ultimate goal of creating a DT (Tab. 1.1). In this literature review, we propose four complexity dimensions for the network representation and five generations for the dynamics modelling (both on and of CNS, referring to Tab. 1.2) to systematise and describe the increasing complexity levels of the CNS that has been/will be developed towards a DT. Based on that, we create unified assessment criteria to assess their respective capabilities of reaching the ultimate goal of building DT of a networked system. Using the proposed criteria, we evaluate how far the reviewed current state-of-the-art approaches are from the DTs. We also identify and propose potential directions and ways of building a DT-CNS

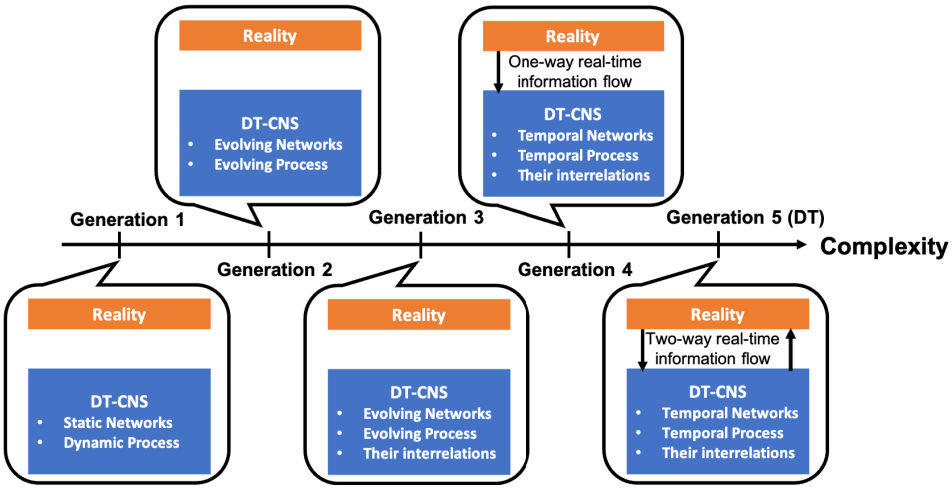


Figure 1.2: Digital Twin Oriented Modelling of Complex Networked Systems

based on the convergence and integration of CNS and DT utilising a variety of cross-disciplinary techniques (Fig. 1.3). Results of this part of the project have been presented in [1].

More specifically, Tab. 1.2 and Fig. 1.3 give us an overview of the generations of modelling framework and the potential directions of developing DT-CNS models. In Tab. 1.2, the first generation of DT-CNS models (generation 1) focuses on static networks and dynamic processes on these networks. The second generation of DT-CNS models (generation 2) investigates evolutionary DT-CNS models by modelling the evolving networks or dynamic processes without considering their interrelations. The third generation of DT-CNS models (generation 3) progresses the generation 2 models forward by incorporating interrelated component changes in the DT-CNS evolution process. In the fourth generation (generation 4), DT-CNS models get more complex by modelling the temporal dynamics of networks and processes, together with their interrelations, based on real-time data acquisition. The DT-CNS models reach the ultimate goal of a DT in the fifth generation (generation 5) when introducing the generation 4 models with a closed feedback loop, which enables the two-way real-time information flow between DT-CNS models and reality. As shown in Fig. 1.2, the conceptual modelling framework for DT-CNSs progresses towards a DT with increasing complexity levels regarding the evolvability of DT-CNS dynamics, the interrelations within DT-CNS models and the interplay between DT-CNS models and the reality.

An example of CNS models are social networks that represent connections between

Table 1.2: Summary of CNS modelling generations

Generation		Brief description
First generation	(generation 1)	Dynamic process on static networks
Second generation	(generation 2a)	Evolving dynamic process on static networks
	(generation 2b)	Dynamic process on evolving networks
Third generation	(generation 3)	Evolving dynamic processes on evolving networks with interrelations between them
Fourth generation	(generation 4)	Temporal dynamic processes on temporal networks with interrelations between them and the acquisition of real time information
Fifth generation	(generation 5)	Temporal dynamic processes on temporal networks with interrelations between them, as well as the real time two-way feedback between the reality and the CNS models, enabling an idealised state required by a DT

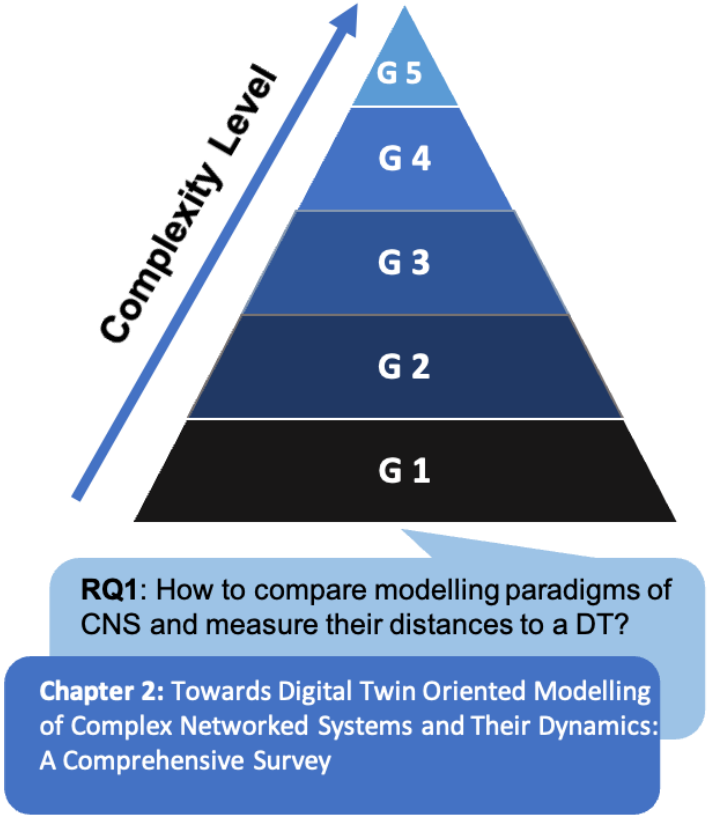


Figure 1.3: Conceptual modelling framework for DT-CNSs

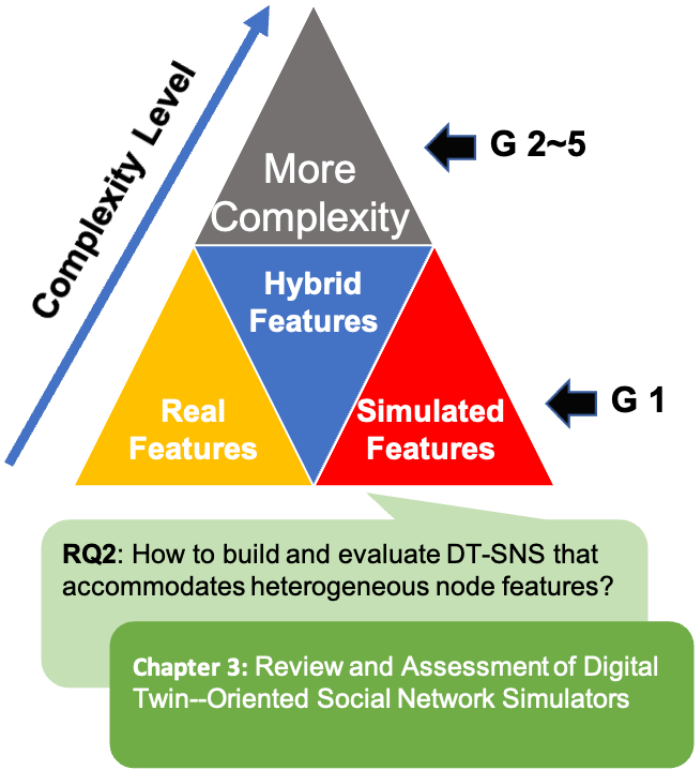


Figure 1.4: Conceptual modelling framework and research directions for DT-SNSs

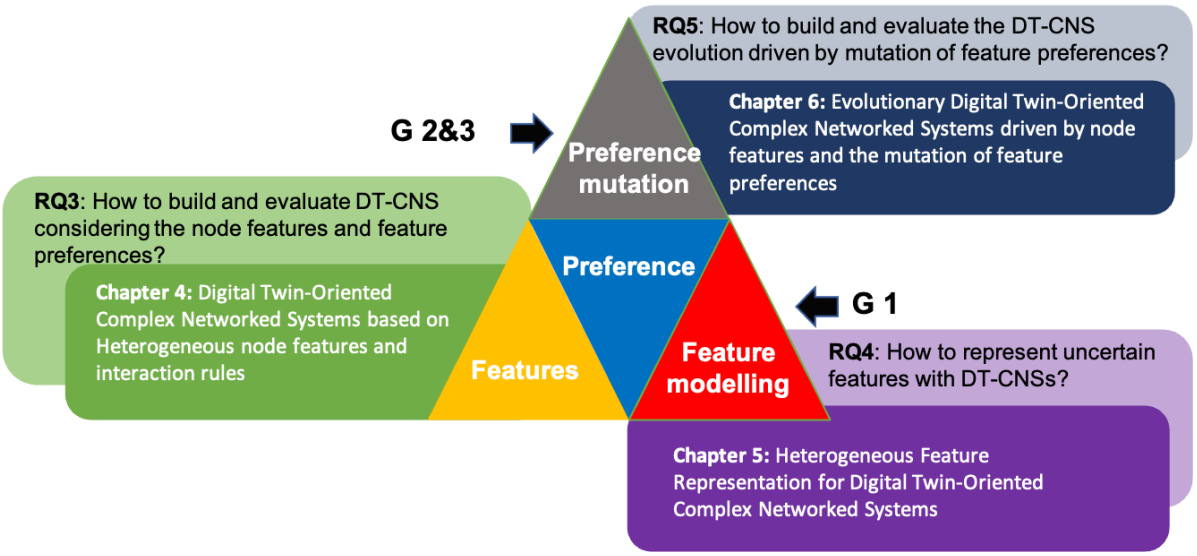


Figure 1.5: Conceptual modelling framework and research directions for DT-CNSs

people and they are used in this thesis as a primary complex system to be investigated due to their high complexity and non-deterministic dynamics. After the review of the existing CNSs modelling paradigms, the thesis moves to reviewing and assessing existing simulators in the space of social networks. In terms of building a Social Network Simulator (SNS) that generates faithful network representations of the real systems, current studies on SNSs have made good progress in the simulation of static networks [24, 45–48], where, however, only few of them consider the attributes of nodes and edges [34]. Under the ultimate goal of building a DT of reality, we need to build and assess Digital Twin-Oriented Social Network Simulators (DT-SNSs) that can be extended towards a DT by incorporating realistic attribute information and generating more complex social networks. This reveals the second research challenge (**RCh 2**) concerning the construction of DT-SNSs based on heterogeneous features of nodes. Therefore, to explore one of the possible development paths for future research on DT-SNS under the conceptual ideas and goals of DT-CNS models, we propose **RQ2** and identify two **RO**s related to the establishment and further development of DT-SNSs based on heterogeneous node features.

- **RQ2**: How to build and evaluate DT-SNS that accommodates heterogeneous features of nodes?

    1. **RO2.1**: Develop a DT-Oriented Social Network Simulator that enables social network simulation based on heterogeneous node features.
    2. **RO2.2**: Evaluate the proposed DT-SNS.

- **RM2**: We answer **RQ2** by a topical review on the current state-of-the-art of Social Network Simulators (SNSs) and an experiment on developing the SNS towards DT-SNSs. In chapter 3, we conduct a critical review and experimental evaluation of the SNSs under the Digital Twin Oriented modelling framework concerned with the proposed complexity dimensions and the assessment criteria based on similarity and efficiency, as proposed in our study [1] (chapter 2). We subsequently extend one of the most promising simulators from the evaluated ones [34], to facilitate generation of social networks of varied structural complexity levels. This extension brings us one step closer to a Digital Twin Oriented SNS (DT-SNS) by introducing heterogeneous features of nodes. We also propose an approach to assess the similarity between real and simulated networks with the composite performance indexes based on both global and local structural measures, while taking runtime of the simulator as an indicator of its efficiency. We illustrate

various characteristics of the proposed DT-Oriented SNS using a well known Karate Club network as an example. In addition, we consider the real feature, the simulated feature and the hybrid features that include both real and simulated features in the construction of DT-SNSs to investigate their respective impact on the composite performance (Fig. 1.4). While not considered to be of sufficient complexity, the simulator is intended as one of the first steps on a journey towards building a Digital Twin of a social network that perfectly mimics the reality. Results of this part of the project can be seen in chapter 3 and have been presented in [2].

The complexity of CNS models results from the heterogeneity of CNS models' components and interactions between them as well as dynamic processes (e.g. epidemic spread) in those systems. Current studies on CNS models focus on the least complex scenario where a single dynamic process takes place on a static network without changes in the network components (nodes and edges) or the process parameters [31, 32, 49–52]. However, there are no current realisations of the conceptual ideas for DT-CNS models. Therefore, to increase the complexity of DT-CNS modelling towards a DT, a natural way forward is to increase the heterogeneity of nodes' features (characteristics of each specific node) and their preferences to create relationships while allowing the CNS models to evolve with the preference changes. As shown in Tab. 1.1, our third research challenge (**RCh 3**) is related to the construction of DT-CNSs modelling framework based on node features and feature preferences and the creation of evaluation protocol for DT-CNS models. We propose **RQ3** and identify three **RO**s, which enable to progress the development and evaluation of DT-SNSs to DT-CNS models by modelling heterogeneous node features and feature preferences.

- **RQ3**: How to build and evaluate DT-CNS considering node features and feature preferences?

    1. **RO3.1**: Develop an extensible DT-CNS based on heterogeneous node features and rules of feature preference.
    2. **RO3.2**: Develop evaluation protocol for DT-CNS.
    3. **RO3.3**: Evaluate the proposed DT-CNS considering feature preferences and features.

- **RM3**: We answer **RQ3** in chapter 4 by building the extensible modelling framework on heterogeneous node features and interaction rules that characterise nodes' preferences for connecting with others. There are two widely employed principles for network formation and growth: preferential attachment [45, 53–55] and homophily [56–59]. By incorporating these two principles, we propose an extensible DT-CNS modelling framework for generation-1 DT-CNS models, which are composed of static networks and spreading processes on these networks without considering their interrelations. In addition, we conduct experiments on simulation-based DT-CNS models that incorporate various age features and rules about network growth and different transmissibilities related to an epidemic spread on these networks. The results can been seen in chapter 4 and have also been included in [3]. Finally, we present a case study on disaster resilience of social networks given an epidemic outbreak by investigating the infection occurrence within specific time and social distance.

As presented in the abovementioned **RO**s, our research project aims to build a DT-CNS modelling framework based on heterogeneous node features and feature preference principles. However, all the features currently studied in the CNS space are specific qualitative features [34]. The uncertain, subjective and vague inductions of qualitative and quantitative features, such as different views of the linguistic descriptions for height (short, tall, very tall, etc.), emerge as another source of complexity for DT-CNS models and a research gap to be addressed in this research project. This is our fourth research challenge (**RCh 4**): to represent uncertain features in DT-CNS models, and therefore, we propose **RQ4** and two **RO**s that aim to improve the expressive power of the proposed DT-CNS modelling framework.

- **RQ4**: How to represent uncertain features with DT-CNS models?

1. **RO4.1**: Develop flexible rules of feature representation for the DT-CNS models.
2. **RO4.2**: Validate the influence of different ways of feature representation on DT-CNS models.
3. **RO4.3**: Evaluate the proposed DT-CNS models built on different feature representation principles.

- **RM4**: We improve the expressive power of node features in generation-1 DT-CNS models by modelling uncertain features with heterogeneous feature representation principles. Current studies generally use fuzzy sets to express vague qualitative or quantitative data based on various Fuzzy Membership functions, such as Gaussian Membership functions [60]. Our study introduces such fuzzy sets to describe the nodes' subjective inductions of the uncertain features and feature differences. We also investigate the influence of the number of fuzzy sets on the expressive power and optimise their parameters to improve expressiveness as part of sensitivity checks and model validation. As a natural extension of our previous study on simulating social networks with simulated age features, our empirical analysis builds DT-CNS models to recreate realistic physical contact networks in different countries from real age and gender feature distributions based on various representation principles and an optimised feature preference. The results can be seen in chapter 5 and have also been presented in [4].

As mentioned in section 1.1, considerable studies in the CNS space focus on the simulation and modelling of static networks [23, 24, 45–48, 61]. Some CNS studies progress towards a more realistic representation of real systems by modelling network topology changes [34, 62–66]. Recent studies also introduce the epidemic spreading processes on dynamic networks and investigate the nodes' infection occurrences impacted by the evolving network connectivity patterns [67, 68]. However, only a few studies investigate the influence of an epidemic outbreak on network evolution [69, 70]. Therefore, an undeveloped research space remains where nodes make heterogeneous decisions about their network evolution in response to the connectivity patterns and the infection occurrences for positive feedback. The interrelated changes of networks and processes, motivated by the nodes' purposeful behavioural changes, add complexity to the CNS dynamics and pose the the fifth research challenge (**RCh 5**): to build an evolutionary DT-CNS models based on the mutation of feature preferences to be solved in our research project. In this context, we propose **RQ5**.

- **RQ5**: How to build and evaluate the DT-CNS evolution driven by mutation of feature preferences?

    1. **RO5.1**: Build an evolutionary DT-CNS with changeable feature preferences and interrelated DT-CNS components
    2. **RO5.2**: Build an evolutionary DT-CNS based on the changes of feature preferences
    3. **RO5.3**: Evaluate the proposed evolutionary DT-CNS models.

- **RM5**: To answer **RQ5**, we propose an evolutionary DT-CNS framework related to heterogeneous node features, changeable connection preferences and their interrelations with the epidemic risks in chapter 6. In addition to the preferential attachment [45] and homophily [56] principles, this study introduces transitivity effects to characterise nodes' preferences for connecting common neighbours in the evolving social networks [71]. Moreover, to facilitate the DT-CNS models with the capability to evolve over time under the impact of the dynamic process, we create heterogeneous preference mutation mechanisms to characterise nodes' adaptive decisions on preference mutation in response to the changes in interaction patterns and epidemic risks. In this space, we use nodes' interaction utilities to characterise the positive feedback from interactions and the negative impact of epidemic risks. We also introduce social capital constraint to harness social connections' density. The nodes' heterogeneous preference mutation styles include the (i) inactive style that keeps initial social preferences, (ii) ignorant style that randomly mutates preferences, (iii) egocentric style that optimises individual interaction utility, (iv) cooperative style that optimises the total interaction utilities by group decisions and (v) collaborative style that further allows the cooperative nodes to transfer social capital. The DT-CNS models, driven by the inactive style fall in generation 1 category by modelling dynamic process on static networks. The evolutionary DT-CNS models driven by the ignorant style fall in generation 2 category by modelling the dynamic process on evolving networks. The DT-CNS models, driven by egocentric, cooperative and collaborative styles, progress to generation 3 by modelling dynamic process on evolving networks with considering their interrelations. Finally, we conduct DT-CNS simulations based on different social capital constraints and preference mutation styles and compare the dynamic patterns regarding the interactions, infections and node utilities. The relevant results are included in chapter 6 and can also be found in [72].

1.4 Significance and Contributions

Below the research contributions (RC) of the thesis are enumerated and their numbering is kept consistent with the one presented in Table 1.1.

In section 1.3, we propose **RQ1** and three **RO**s concerning (i) the review and comparative analysis of CNS models, (ii) proposal of a conceptual modelling framework for DT-CNSs and (iii) creation of an assessment criteria to compare and evaluate CNS models (section 1.2). Corresponding to each research objective, our research contributions (**RC**s) lie in three aspects.

- **RQ1**: How to compare modelling paradigms of CNS and measure their distances to a DT?

    1. **RC1.1**: A comprehensive and comparative review of existing CNSs modelling paradigms.
    2. **RC1.2**: New conceptual modelling framework for DT-CNSs.
    3. **RC1.3**: Novel assessment framework for evaluation of models developed within DT-CNS modelling framework.

For the second research challenge (**RCh 2**) concerning the construction of DT-SNSs based on heterogeneous features of nodes, we propose **RQ2** and two **RO**s, related to building and assessing DT-SNSs. The corresponding **RC**s are as follows:

- **RQ2**: How to build and evaluate DT-SNS that accommodates heterogeneous features of nodes?

    1. **RC2.1**: DT-SNS that enables to faithfully represent static network topology with a given complexity level.
    2. **RC2.2**: A case study on the modelling and assessment of DT-SNS with different levels of structural complexity.

We propose **RQ3** and three **RO**s to address our third research challenge (**RCh 3**), related to building and assessing DT-CNS models. The corresponding **RC**s are as follows:

- **RQ3**: How to build and evaluate DT-CNS considering node features and feature preferences?

    1. **RC3.1**: Construction of DT-CNS based on heterogeneous node features and feature preference principles.

2. **RC3.2**: Evaluation protocol to assess the faithfulness of DT-CNS, considering similarity, efficiency and reproducibility.
3. **RC3.3**: A case study on DT-CNS models built on different node features, feature preferences and transmissibility of epidemic spread.

Regarding the fourth research challenge (**RCh 4**) of representing uncertain features in DT-CNS models. We propose **RQ4** and two **RO**s, which aim to improve and validate the expressive power of the DT-CNS models by introducing flexible rules of feature representation. The corresponding **RC**s are as follows:

- **RQ4**: How to represent uncertain features with DT-CNS models?
  1. **RC4.1**: Feature representation with fuzzy sets for uncertain features.
  2. **RC4.2**: DT-CNS models built on different feature representation principles.
  3. **RC4.3**: A case study on DT-CNS models built with increasingly flexible rules of feature representation.

Finally, to address the fifth research challenge (**RCh 5**) of building an evolutionary DT-CNS models based on the mutation of feature preferences, we propose **RQ5** and three **RO**s. We aim to build and assess the evolutionary DT-CNS models considering heterogeneous preference mutation styles and the interrelations between the dynamic processes and the networks. The corresponding **RC**s are:

- **RQ5**: How to build and evaluate the DT-CNS evolution driven by mutation of feature preferences?
  1. **RC5.1**: New feature preference mutation mechanism considering heterogeneous mutation styles.
  2. **RC5.2**: Evolutionary DT-CNS models with evolving networks, dynamic processes and the interrelations between their dynamics.
  3. **RC5.3**: A case study on the evolutionary DT-CNS models in an epidemic outbreak based on different preference mutation styles.

Overall, this thesis contributes to the DT-CNS space by a conceptual modelling framework and an evaluation protocol for DT-CNS models. In addition, we realise these conceptual ideas by proposing and progressing a DT-CNS model with an increasing complexity level towards a DT. We also create what-if scenarios using the DT-CNS

models to investigate the social networks impacted by an epidemic outbreak. The DT-CNS models in this thesis focus on the heterogeneous node features and the dynamics related to feature representation, feature preferences and preference mutation. They are first proposed to model static networks and dynamic processes on the networks and, afterwards, developed with higher complexity levels by modelling uncertain features and mimicking evolutionary networks and their interrelations with the dynamic process.

1.5 Thesis Organisation

In this thesis, we first introduce the background, aim, research questions and objectives, research methodology as well as the thesis structure in chapter 1. We organise the following thesis structure based on the research questions proposed in section 1.3 and outline the motivations for proposing and answering the respective research question in each chapter as per Fig. 1.6.

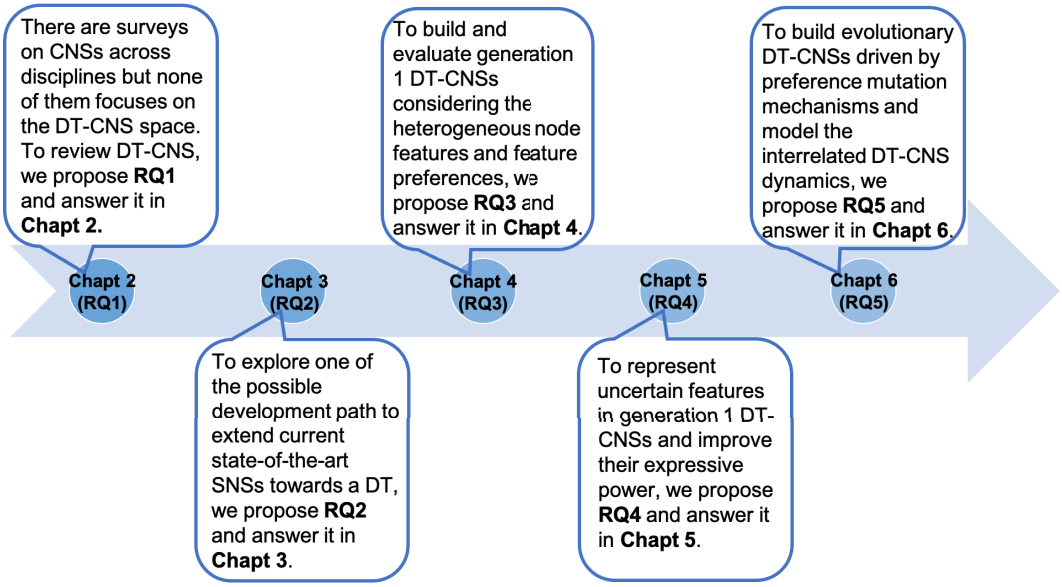


Figure 1.6: Thesis organisation considering the research motivations for each chapter.

In chapter 2, we answer **RQ1** by a comprehensive overview on CNS models as they approach their ultimate goal of creating a DT. In this literature review, we conceptualise five generations of modelling framework for DT-CNSs (referring to Tab. 1.2) to systematise and describe the increasing complexity levels of the CNS that has been/will be developed towards a DT. The results can also be also found in [1].

Under the conceptual framework and goals of DT-CNS models proposed in chapter 2, we explore one of the possible development paths for future research on DT-SNS in chapter 3. We propose **RQ2** and answer this question with a topical review of the current state-of-the-art Social Network Simulators (SNSs) and an experiment on developing the SNS towards DT-SNSs based on heterogeneous node features. The results have been also presented in [2].

Based on the topical review in chapter 3, we can find that, to increase the complexity of modelling towards a DT, a natural way forward is to increase the heterogeneity of nodes' features (characteristics of each specific node) and their preferences to create relationships while allowing the CNS models to evolve with the preference changes.

Therefore, as shown in Fig. 1.5, we propose **RQ3** and answer this question in chapter 4 by building the extensible modelling framework on heterogeneous node features and interaction rules that characterise nodes' preferences for connecting with others. The research results can be found in [3].

To represent uncertain features in DT-CNS models, we propose **RQ4** and answer this question in chapter 5 by introducing feature representation principles (Fig. 1.5). This improves the expressive power of generation 1 the DT-CNSs by introducing more complex dynamics related to feature representation.

To progress the generation 1 modelling framework towards a higher complexity level concerning the interrelated DT-CNS component changes, we propose **RQ5** and answer this question in chapter 6 by proposing an evolutionary DT-CNS models driven by heterogeneous preference mutation mechanisms.

Overall, this thesis contributes to the topic, Digital Twin-Oriented Complex Networked Systems with five separate studies corresponding to the five research questions (chapter 2 for **RQ1**, chapter 3 for **RQ2**, chapter 4 for **RQ3**, chapter 5 for **RQ4** and chapter 6 for **RQ5**). Finally, we conclude the thesis in chapter 7 by discussing research contributions and providing ideas for further work in this space.

CHAPTER 2

LITERATURE REVIEW OF MODELLING COMPLEX NETWORKED SYSTEMS AND THEIR DYNAMICS

2.1 Introduction

2.1.1 Background

A complex network can be seen as a universal concept used for representation and analyses of complex systems. Given the growing interest in real complex systems and a fast development of modelling techniques, the complex networked system (CNS) area has become a highly cross-disciplinary field that involves multiple modelling approaches with various research aims posed and achieved over the years.

There is a considerable literature about complex networks and various researchers have published several surveys reviewing and exploring the topic from different perspectives and application areas. Those include works on complex networks and their applications covering multiple application areas [42] or oriented towards specific topics such as networks of cryptocurrency transactions [73], vehicular networks [74], supply chain networks [75], internet of things [76] and networks of short written text [21].

When it comes to the models of complex networks in the context of their structure, there is also a body of work surveying a variety of network topologies [21, 22] and dynamics. The networks dynamics can be either considered as: (i) dynamic processes over networks, that involve surveys on spreading processes like epidemic processes, information spreading processes [15, 18], or (ii) dynamic networks with evolving structures and

features [19, 20].

When it comes to modelling techniques that have been applied to build complex networked systems, there are also surveys that review certain types of modelling approaches for complex networks, including Graph Neural Network [19], game theory [43] or non-parametric Bayesian modelling [44].

Complex networked systems are modelled with the goal of accurate reflection of reality with the aim of simulation, prediction and/or control. Over the years, proposed CNS models have become more and more accurate with more realistic networks topologies, characteristics and evolving dynamics modelled. Increasingly, they can capture features of real world scenarios and, in some cases, behave like their twins. Researchers have already focused on the studies of Digital Twining of real systems across disciplines, which involves a wide range of applications but so far complex networks area, and especially social network space, has been marginalised in this development space. Digital Twins are expected to provide an accurate reflection and extension of the reality fulfilling simultaneously multiple aims of modelling [1]. In our review of the current state-of-the-art of the modelling of CNS, a DT is treated as an ultimate goal for the representation and modelling of the Complex Networked Systems, which has not been realised yet but is expected to be steadily approached with the ability to take more complex information into account, capture and represent dynamics of and on the networks as well as simulate and interact with the modelled reality in real time.

Therefore, the convergence of Digital Twinning and Complex Networked Systems emerges as an exciting research focus with a potential to address some of the outstanding modelling and representational challenges ultimately leading to the establishment of a Digital Twin-Oriented Comple Networked System (DT-CNS) area as its main goal. And in this space, the modelling of CNS in the context of DT is to represent the observable information faithfully in a form of a network together with dynamics of and on this network, while enabling the fulfilment of model's aim and, in its ultimate realisation, seamlessly intertwining and interacting with reality in real time.

Multiple modelling paradigms, with their ways of data analysis, representation and modelling, have been applied to answer various questions about CNSs. By modelling paradigm we understand a family of approaches, which are built around common principle (e.g. rule–based systems), that can be used to build a model of a system. There is a need to review and explore where, how and why complex networked systems are modelled across many different disciplines. However, most of the surveys only account for certain types of complex networked systems from a specific perspective or a single

discipline (e.g. supply chain networks [75], network topology [21, 22], etc.), except for this work on complex networks and their applications covering multiple application areas [42].

2.1.2 Research design

To fill in the gap mentioned in section 2.1.1, we propose the **RQ1** and in this chapter, answer **RQ1** by a comprehensive overview on CNS models as they approach their ultimate goal of creating a DT. We structure this literature review using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISM) approach. More importantly, we propose our research foci on DT-CNS models and provide the research agenda for building and developing DT-CNS models towards a DT.

- **RQ:** How to compare modelling paradigms of CNS and measure their distances to a DT?

    1. **RO1.1**: Conduct a comprehensive critical review and comparative analysis of CNS models.
    2. **RO1.2**: Propose a conceptual modelling framework for DT-CNS.
    3. **RO1.3**: Propose an assessment approach to conceptually compare and evaluate CNS models.

In this chapter, we review conceptually and compare research on complex networks from a holistic point of view while trying to deal with the questions concerning diverse modelling paradigms and their distance to the idealised DTs (**RO1.1**). We devise a framework that enables to compare various modelling paradigms (**RO1.2**) and evaluate their respective distances to DTs (**RO1.3**), which involves answering the following four fundamental questions: (1) What is the aim of the modelling?; (2) How to represent information about a system in a form of a network?; (3) How to model the dynamics in a networked system?; and (4) How do we approach the ultimate goal of building a CNS model that models the reality at a Digital Twin level?

To answer these questions, we found current up-to-date studies on CNSs and DTs by searching keywords on google scholar. Regarding the CNS space, the keywords include complex networks, network dynamics, dynamic processes, network simulations, network representation and spreading process on the networks. Regarding the DT space, the keywords include digital twin and digital twin networks.

Through answering the above questions, this chapter mainly contributes in four aspects:

- an overview of the current, relevant state-of-the-art CNS models from multiple disciplines to inform future research directions (**RC1.1**);
- proposing four complexity dimensions regarding the complexity of network representation (**RC1.2**);
- constructing five generations of models to describe the increasing complexity level of CNS model that has been/will be developed towards a DT in terms of the dynamics modelling (**RC1.2**);
- creating a unified assessment framework to evaluate how far the reviewed CNS models are from the idealised DTs (**RC1.3**).

This chapter is organised as follows. Section 2.2 illustrates the prerequisites of setting and fulfilling model's aims, discusses different model's aims across different disciplines and summarises their research foci. This is followed by two sections concerned with complex networked systems with section 2.3 defining and reviewing different network representations and modelling paradigms and section 2.4 focusing on modelling dynamics in networked systems. In section 2.5 a new framework is proposed for evaluating models of complex network systems given the ultimate goal of achieving DTs faithfully representing the reality. The research gaps and future research directions are also identified and discussed in this section. Finally, the Conclusions are provided in the last section.

2.2 The aim of modelling Complex Networked Systems

There is always a need to determine the aim of developing a new model at the start of the modelling process, where, under the constraints of observability, we model a Complex Networked System in a way to fulfill this aim. The choice of modelling paradigms, which can be understood as the approaches employed to model real state or dynamics of the system, is largely dependent on the model's aims determined based on the research questions. The aims of modelling Complex Networked Systems include solving a wide range of tasks such as link prediction, network detection, mimicking of real systems, etc.

The CNS models together with their respective primary aims are also typically developed within different disciplinary focuses such as epidemiology, sociology, microeconomics, etc. We categorise the modelling paradigms based on their concrete model's aims, which helps us to understand what is expected from modelling complex network systems and serves as the basis for further discussions of detailed modelling paradigms. In this section, having in mind various model's aims that influence how we build such a model, we also set an ultimate goal of modelling complex networked systems, which is to create a Digital Twin (DT) of a real world system. Achieving such a goal would enable the fulfilment of multiple aims simultaneously through accurate reflection and mimicking of the reality.

2.2.1 Prerequisites of modelling

Selecting the right modelling paradigm together with the availability of the needed input data are the prerequisites of setting and fulfilling model's aims. Given varying complexity of reality which can only be partially observed and modelled (e.g. instantaneously emergent patterns motivated by intractable dynamics), there are no concrete standards or measures to ensure and prove the implementation of these prerequisites. Thus the model is always built on the strong assumptions of appropriate variable and model selection. The discrepancy between these assumptions and the truth may result in incapability of predicting dramatic changes due to partially observable input space [77] or overfitting problem that is typical for overly complex models [78]. Therefore, researchers start to evaluate and reduce their distance to these prerequisites by considering observability.

Observability, in structural way, is about the ability of reconstructing the state of a system from a limited set of measured variables in finite time [79], while from the perspective of dynamics, observability is related to a deeper understanding about how the selected variables interact and evolve to change the states of the system. The structural observability is dependent on the available data set while the dynamical observability is analysed based on a model of dynamics, which is inevitably wrong to some extent but may be useful either in reflecting the reality or diagnosing an input space that is partially observable. Linear time-invariant system proposed by [80] is one typical modelling paradigm widely used in the studies on observability of dynamical systems [81].

In complex networked systems (CNS), structural observability determines the number of nodes to be measured to render a network observable, while dynamical observability considers node dynamics and coupling when selecting the best driver (sensor) node to modulate (observe) the whole network activity [79]. The design and analysis of CNS

for each model's aim involves the observability of topology [81], variables for coupling nodes [82] and node dynamics [83]. These structural observability and dynamical observability are closely related to variable and model selection, and serve as the prerequisites of representing and modelling reality under each model's aim in different application scenarios.

The measures of observability vary depending on the linearity of systems. Linear networks without symmetries have been well studied research objects with respect to the topic of observability, where the observability matrix based on the dynamic model for a linear (time-invariant) system proposed by [80] is widely used. Nonlinearity of dynamical networks have been recently considered in the studies on dynamical observability. For example, [84] quantify the observability and controllability of nonlinear networks with explicit symmetries that shows the connection between symmetries and nonlinear measures of observability and controllability. [85] propose a nonlinear graph-based theory for dynamical network observability from the Jacobian matrix of the governing equations of nonlinear systems.

2.2.2 Currently studied CNS modelling aims

The model's aims can be categorised as: (i) specific and (ii) abstract goals. Specific model's aims, such as community discovery, link prediction, anomaly detection, synchronization and controllability of networked systems, focus on specific external tasks for observable research object (i.e. CNS in our case) with measurable model outputs. Abstract model's aims, like topological feature analysis and the mimics of real life systems, approach the inner rules of real dynamics, analyse observables and simulate unobservables for further research on specific model's aims. Examples of different CNS models' aims together with the relevant references are shown in Tab. 2.1.

Community discovery aims to decompose complex networks into meaningful sub-networks that better describe local phenomena [20]. The local phenomena refers to a set of entities that share some closely correlated sets of actions with the other entities of the community [20, 128]. This has been explored and discussed in a wide range of applications, including the detection of community structure hidden in real social networks [86–88], collaboration network analysis like detecting citation patterns [89], improving routing of telecommunication network [90], reconfiguration of the brain network [91] or political affiliation [92].

Table 2.1: Different goals of building CNS models

Model's aim	References
Community discovery	[20], [86], [87], [88], [89], [90], [91], [92]
Link prediction	[9], [10], [11], [93], [46], [94], [95], [96], [97]
Node classification	[98], [34], [99], [100]
Synchronization	[101], [102], [103], [104], [105], [106], [107], [108], [109]
Controllability	[20], [86], [110], [89], [90], [91], [92]
Anomaly detection	[12], [13], [111], [112], [113], [114], [115], [116], [117]
Topological analysis	[21], [118], [119], [104], [17], [120], [32], [109]
Mimics of reality	[14], [15], [121], [122], [123], [124], [109], [125], [116], [126], [127].

Link prediction aims to infer the behaviour of the network link formation process by predicting missed or future relationships based on observed links and the attributes of both nodes and relationships [10, 11]. Link prediction involves questions of dealing with missing links or link labels of networks and predicting links in changing networks, including social networks [9, 46, 93], food webs [94], networks in collaborative recommendation tasks[95], knowledge graphs [96] and biochemical networks of protein interaction [97] and metabolism [94].

Node classification aims to provide a labeling for unlabeled nodes in a network composed of partially labelled nodes and edges [98]. Node classification, as an important way to explore node features and links, has been widely studied in social networks [34, 98], citation networks [99] and co-author networks [100].

Synchronisation of complex networks implies that the states of two or more interacting nodes in a network with different initial conditions gradually approach each other and finally reach the same state [101]. The applications of synchronisation in complex networks range from the stability of power grids [102, 103], controllablity of neuronal networks [104–106], optimising timetables for transportation [107] to the synchronization patterns affected by network topology in chemical systems [108] and IoT systems [109].

Controllability of networks represents the ability of controlling the networks, which is independent of the way that the outputs are formed, while its related concept of observability, depends only on the outputs but not on the inputs [129]. The studies on controllability are often combined with observability, which range from social net-

works [130], protein interaction networks [131], brain networks [132] to transportation networks [133].

Anomaly detection of networks is about finding objects, relationships, or points in time that are unlike the rest [12]. There are many studies on anomaly detection in various application scenarios, ranging from anomaly detection in social networks [13, 111, 112], public health [111], IP networks [113], wireless sensor networks [114] to intrusion of networks [115]. Similarly to the principles of anomaly detection, pattern recognition is also used for diagnostic analysis [116] and heterogeneous component detection in knowledge networks [117].

Topological feature analysis is a very popular model's aim in both real-life networks and artificial networks. This is a research topic that probes network topological features from real data for applications across disciplines based on network-based models, such as the probed topological features of text network for language organisation [21, 118], probed boundary features within the small world complex networks for image classification [119] and organisation of brain network [104]. Local network topological features like three-node motif [134], directed closure [135] and quadrangle structure [136] are researched through topological feature analysis of social networks. There are also discussions of artificial network topological effects on network based models, which involves studies of topological effects on P2P trading in financial market [17], artificial neural networks in computation tasks [120], epidemic spreading [32] and enhancing synchronisation of IoT systems [109].

Mimics of reality helps to deal with questions about analyses of dynamics over networks or features of networks [14]. There are studies of spreading processes in artificial networks including epidemic dynamics [15, 121], opinion dynamics [122, 123] and meme diffusion [124]. There are also studies that simplify the real complex systems as data-driven networks to assist further analyses, which involves applications ranging from Digital Twins of IoT systems [109, 125], image information representation [116], mimics of transportation systems [126], design of supply chain networks [137, 138, 138–141] and the representation of human-object interactions via networks [127].

2.2.3 Digital Twin: an ultimate goal

Researchers focus on the studies of twining real systems across many disciplines and those efforts have already resulted in a development of a field on its own known as Digital Twins (DT). DT serves as an "almighty" paradigm of mimics across spatial and temporal scales. It has also grown to become an ultimate goal of modelling complex networked systems due to its reality-friendly nature, integration of model functions and the wide range of applications.

Digital Twin is a virtual extension of reality, which not only allows to compare current conditions with historical data to provide meaningful information to assist in decision-making, but also enables forecasting and feedback of eventualities that have never happened before [38]. Researchers have defined DTs from different perspectives across the application scenarios. In a fully digitalized product life cycle, DT is a comprehensive virtual product model with the features of real-time monitoring, simulation and forecasting [142]. For mechanical and cyber-physical systems, DT is a linked collection of digital artefacts that evolves with the real system along the whole life cycle and integrates currently available knowledge with the purpose of describing behaviour and deriving solutions for the real system [36].

DTs have three elementary components repeatedly emphasised: the digital (virtual part), the real physical product and the connection between them [37], while there are also other imperative components added with the accumulation of practice, including data, service, machine learning, and DT Performance evaluation [39–41]. DTs also resemble a series of models with integrated functions like simulation, optimization and data analytics [143] and features of real-time processing and continuous updates [144]. This makes DTs impossible to be replaced by any single tool and ideal modelling paradigms for health monitoring [145] , planning of manufacturing [146], management of smart city [147], accurate healthcare [148] and anomaly detection [149] within a wide range of complex systems, including complex networked systems like IoT systems [150, 151] and blockchain-encapsulated systems [152, 153].

The above mentioned components, integrated functions and universal application of DTs differentiate them from any other simulation tool or modelling paradigms by emphasising the properties of real-time data acquisition of observations and feedback, and self-evolution through continuous machine learning analysis. They contribute to DTs' status as a powerful tool for the mimicking of a series of realities and an ultimate goal for modelling complex networked systems across disciplines.

Modelling complex networked systems using a Digital Twin paradigm has a potential

to build a universal model that can be adapted to fulfil multiple, different models' aims discussed in this section. But before this can be attempted, we need to review and assess how the modelling of complex networks, their dynamics as well as dynamics on those networks is approached currently and this is the focus of the following sections.

2.3 How to represent information in a form of a network

This question involves two important issues in building complex networked systems, which are: (1) what types of networks are needed in modelling certain phenomena including the ways of building the required network topology with appropriate complexity using real data and simulations, and (2) how to obtain these networks from different perspectives of obtaining and processing data to realise such complexity. To achieve a faithful representation of a network that preserves as much information as needed in the modelling of networked systems (section 2.3) for certain models' aims (section 2.2.2), one needs to collect and process observable data and information about the network structures and associated dynamics with modelling paradigms that minimise the information loss in the process of a network generation.

2.3.1 How to represent networked data

According to [154] network topology is a representation of the physical connections that exist among entities in a communication network. This definition can be and has been easily expanded across disciplines as network topology describes how the entities of any type relate to each other in any type of network. The establishment of network topology is a crucial step for modelling complex networked system, and its effect on the dynamics has been a popular research focus.

2.3.1.1 Complexity dimensions

Network topologies vary in complexity when they represent networked information that can vary in data availability and modelling necessity. Complexity of the topology results from different types of nodes, edges and their attributes. As shown in equation 2.1, the network at time t can be represented as G_t, with its complexity level dependent on the heterogeneous network components and their changes over time: the set of nodes $V_t = \{v_{1,t}, \cdots, v_{N,t}\}$, the set of edges $E_t = \{(v_{i,t}, v_{j,t}) | v_{i,t}, v_{i,t} \in V, i \neq j\}$, the attributes of each

node $A_t = \{a(v_{1,t}), \cdots, a(v_{N,t})\}$ and the attributes of each edge $W_t = \{w(v_{i,t}, v_{j,t}) | v_{i,t}, v_{i,t} \in V, i \neq j\}$ [87],

$$G_t = (V_t, E_t, W_t, A_t) \tag{2.1}$$

In Fig. 2.1, we propose to describe this complexity alongside and using the following four dimensions: (i) a structural dimension connected to the scale and diversity of the topology, (ii) a temporal dimension concerned with time-to-live of different components of the network, (iii) a spatial dimension connected with the space in which the topology can be embedded, and (iv) a dynamics dimension connected to the topology's exerted or encapsulated dynamics. The complexity in structural, temporal and spatial dimensions describes the necessary reality required to be represented via networks, while the dynamics complexity depends on the models selected to explore the complex reality. These modelling paradigms are shown in Fig. 2.1. Given available networked data set that is observable from the perspective of the above mentioned four complexity dimensions, in topological feature analysis, only structural information about networks need to be represented and analysed based on a model that reveals the inner rules of network formation, while for prediction of link formation over time, temporal complexity is further incorporated and a complex model that performs well in such an external task is employed.

Structural complexity of the scale and diversity of nodes, edges and their attributes involves discussions of non-attributed and attributed networks in terms of their diversity, as well as small/big networks and sparse/dense networks in terms of their scale, where structural complexity increases with more detailed information needed to be represented.

Research studies usually start from the exploration of non-attributed networks that are built only with homogenous nodes and edges. Such networks represent simplified real world scenarios and are thoroughly studied in terms of their topological features, like the analysis of potential energy landscape with both a small-world and scale- free character [48] or the conduction of specific model's aim e.g. community discovery which only uses network topology to find partitions [155].

Attributed networks can better represent real-world networked interactions and information as they introduce the auxiliary information via node or edge attributes [156, 157]. The node attributes describe the features of nodes within interactions or relations, while the edge attributes capture information about how the adjacent nodes interact with

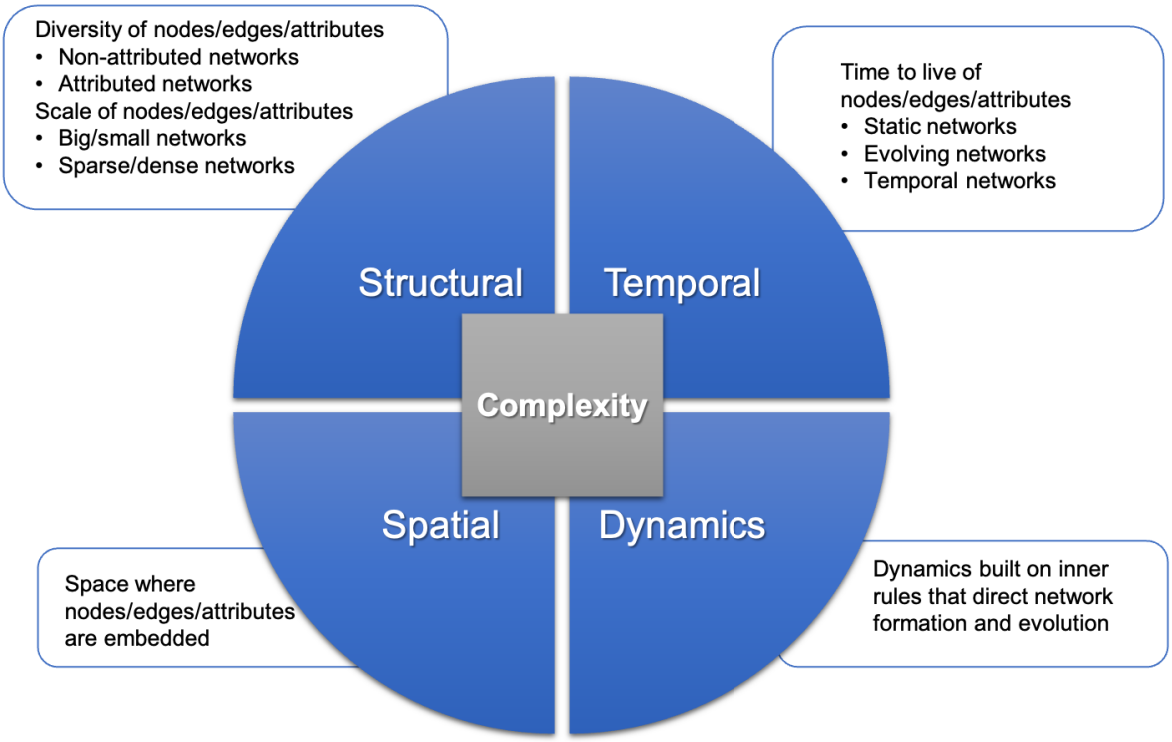


Figure 2.1: Four dimensions of complexity related to a network's topology.

others in the networks [156]. These attributes vary with application scenarios. Taking online social networks as an example, nodes represent users and are attributed with user profiles [158], while edges represent online relationships and have attributes like nature of relation, direction, intensity and durability [159]. Many researchers study the structural complexity of node-attributed networks having in mind such modelling aims as community discovery [87, 160], link prediction [161], anomaly detection [158], controllability [162], topological feature analysis and mimicking of reality [163]. However, few researches focus on the representation and modelling of generalised edge-attributed networks. Edge-attributes of such networks within most researches typically take the numerical or categorical forms [156, 164], where directions [165] and edge weights [166, 167] are often studied. Compared with the large number of studies on node-attributed networks with various model's aims, only a small number of studies focuses on community discovery [168] and anomaly detection [156] of edge-attributed networks.

The structural complexity also increases with the requirements of representing more information within large-scale networks and the difficulty of processing networks with sparse edge information. Structural complexity of large-scale networks with thousands and millions of nodes results from a complicated and higher-order inner structure [169],

which are common in DTs like city IoT [163] and DT of manufacturing with big data [170]. Structural complexity of sparse networks, given fewer edges, lies in their restrictions of attributes processing [157, 171] and optimal modelling [172]. There is an even more complex case for large sparse networks where both complicated large-scale inner structure and problematic sparse edges are involved [172].

Temporal complexity of networks increases when more temporal information about nodes, edges and their attributes can be captured and modelled meaning that less information is lost. Networks can be conceptually described as static, edge-weighted, evolving and temporal [19, 20] as they are transforming to be instantaneous and approaching the state of continuity without temporal aggregation.

The basic modelling of real-world phenomena and systems using CNS models starts with a static network topology where nodes and edges are fixed and they are assumed to be "frozen" in time. Such an assumption greatly simplifies the modelling process but fails to capture the evolving features of real-world systems.

The attempts of accounting for temporal information in the network modelling process have gained attention as they usually improve the model performance. For example, social network analysis initially viewed networks as static rather than changing over time [173, 174]. As the field developed, social interactions started being represented using temporal networks to capture dynamics and instantaneous character of the contacts [167, 175]. In terms of biomolecular networks, studies on protein-protein networks first employed static data-driven networks to represent and analyse the protein-protein interaction [176] until the dynamic protein-protein networks have been found to benefit the study on the molecular systems of protein complexes [177]. The analyses of transportation networks also initially used static networks [178] and then shifted to dynamic management of transportation systems for better model performance [179]. However, the transition from the static networks under the most stringent assumptions to the temporal networks that result in smaller loss of information can not be achieved overnight, while the scale of involving time dimension also differs from case to case.

Some of the researches focus on edge-weighted networks with temporal weights in the analysis of social relations [167] and the mobility in wireless networks [166]. Other researches focus on evolving networks where network topology changes slowly over time so that its instantaneous snapshot yields a well defined network [19]. Some evolving networks just represent more stable relations, rather than instantaneous interactions between the nodes, which can be captured with durable edges like citations [180] and

friendships [159, 181], while some researches use evolving networks built with snapshots to aggregate the temporal information of interactions within a time window for a more stable representation in the analysis of instantaneous features [89, 90, 110, 181]. There are also studies on temporal networks that have non-trivial topology changes and can not be represented via instantaneous snapshots. They preserve all the temporal information and build networks in a more faithful way, such as instantaneous contacts of communication via e-mail, text messages, or phone calls with temporal edges of networks [182].

Spatial complexity involves discussion of spatial networks, which are proposed and defined as networks where nodes are located in a space equipped with a metric of the usual euclidean distance [183]. For example, spatial networks that represent urban street patterns can be built with a metric of distance, which is measured not just in topological terms (steps), but in properly spatial terms (meters, miles) [184]. Considerable applications of spatial networks involve modelling of human activities that take place on a spatial matrix obtained from largely three types of transportation networks including matter (streets, roads, highways, railways, airport networks), energy (the power grid) and information (Internet, telephone networks) [185, 186].

Researchers have been characterising and understanding the structure and the evolution of networks under the impact of spatial constraints [183], where topology alone does not contain all the information to understand the dynamics of networks and the spatial information is needed. Some social-spatio networks consider the interplay of social networks and locations, such as social interactions under the impact of transportation [187], employment outcomes in a labour market driven by social contacts under an explicit geographic structure [188] and the social properties of Twitter users,Äô networks with the spatial proximity of the networks [189]. Traffic under the constraints of transportation networks also involves studies on commuter flows constrained in large transportation networks [190] and planning of unobstructed paths in traffic-aware spatial networks [191].

Dynamics complexity is about the dynamics of networks that enables to develop a deeper understanding of temporal complexity based on structural and spatial complexity by investigating the rules of network evolution via simulation or modelling. Temporal and dynamics complexity, which are respectively derived from structural and dynamical observability, represent varying degrees of reality. The temporal complexity describes

which components of a network change over time, while the dynamics complexity, as a deeper understanding of temporal complexity, considers why and how these components change. The relations between observability and complexity dimensions are shown in Fig. 2.2.

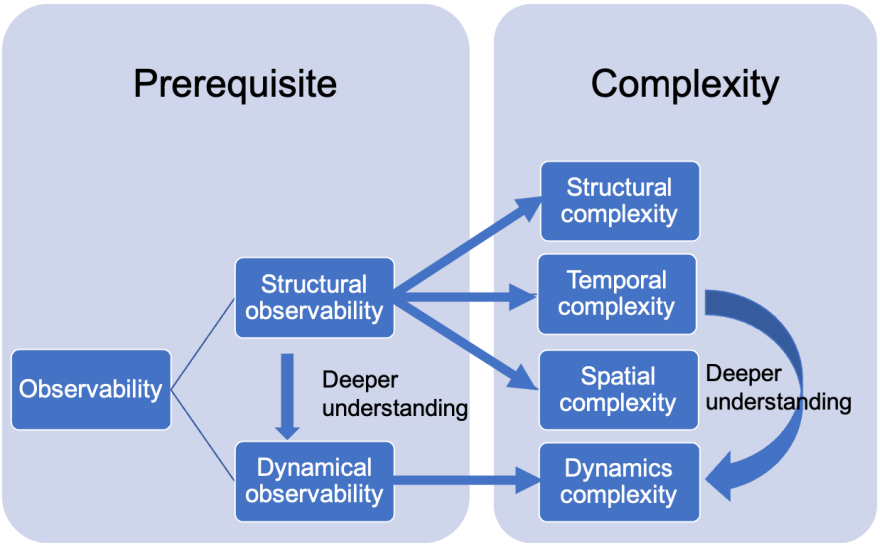


Figure 2.2: Observability and complexity with corresponding relations

For dynamics built on inner rules that direct network formation and evolution, the dynamics complexity increases as networks are generated with less human involvement based on more realistic rules ranging from statistical relations [192, 193] to realistic principles like homophily [47, 56]. This dynamics can either be exerted dynamics that controls the network changes to a desired state of CNS model [64, 66, 194], or encapsulated dynamics that motivates the network formation and evolution with higher degrees of automation [34, 47, 56]. For example, exerted dynamics used for social network intervention can control each step of network evolution by changing attributes of nodes that are identified based on different man-made strategies [66]. Encapsulated dynamics mainly focuses on the edge formation mechanisms, e.g. based on preferential attachment principles [34, 47] which are typical examples of highly-automated self-evolution without involvement of human decision.

The synergy effects of complexity dimensions describe the "1+1>2" effect of the combined complexities from different dimensions, which are more complex but closer to reality. Spatial temporal networks are typical examples for combined complexity of temporal and spatial dimension, which are proposed for a more faithful representation of reality with the influence of space on constraining the structure of temporal networks

considered. Some researches employ temporal networks to capture and process temporal information under consecutive frames, while they construct spatial networks to extract certain static features. Such spatial-temporal networks have been widely used in computer vision like facial expression [195], video-based person reidentification [196] and identification of human-object interaction [127]. Some researches introduce temporal information to spatial networks. Taking recommendation task as an example, [175] build spatial temporal networks with temporal edge-weights by incorporating time dimension to user-location graphs and using sessions that capture the co-locations among two or more users during a time window.

It is clear that different complexity dimensions are intertwined. They influence and build on each other and this is one of the challenges that need to be considered when modelling of complex networked systems is attempted. Research has been conducted in each of the complexity dimensions but building overarching framework that would enable to flexibly adjust a level of complexity of each of the dimensions and simulate various what if scenarios is still an outstanding challenge. With the recent developments in the Digital Twin space, modelling CNS using DT paradigm is a promising way forward.

2.3.1.2 Data-driven vs simulation-based networks

Researchers have made a lot of effort to faithfully represent the information from real world systems by developing variety of modelling approaches. This involves data-driven networks, simulation-based networks and networks that are built by combining these two approaches. These networks are featured with varying degrees of complexity. Components of different network types are shown in Fig. 2.3.

Data-driven networks built on rich real data sets may not necessarily capture all the information as they may be confined by relatively simple model's aims under strict assumptions, like the ignorance of temporal information for temporal networks analysed in a "static" time scale [173]. As researchers explore the real data sets with more complex modelling aims and advanced techniques, data-driven networks gradually approach reality as more complexity is allowed to be introduced to network topology with more relaxed assumptions. For example, the analysis of social networks started from statistical analysis of static networks and their topology [173] and then evolved into a more complex modelling aims such as community discovery in networks that are increasingly more and more complex, starting from static [197] to evolving [86, 198], attributed [171] and both evolving and attributed networks [20].

When real data is not available, simulated (synthetic) networks can be generated

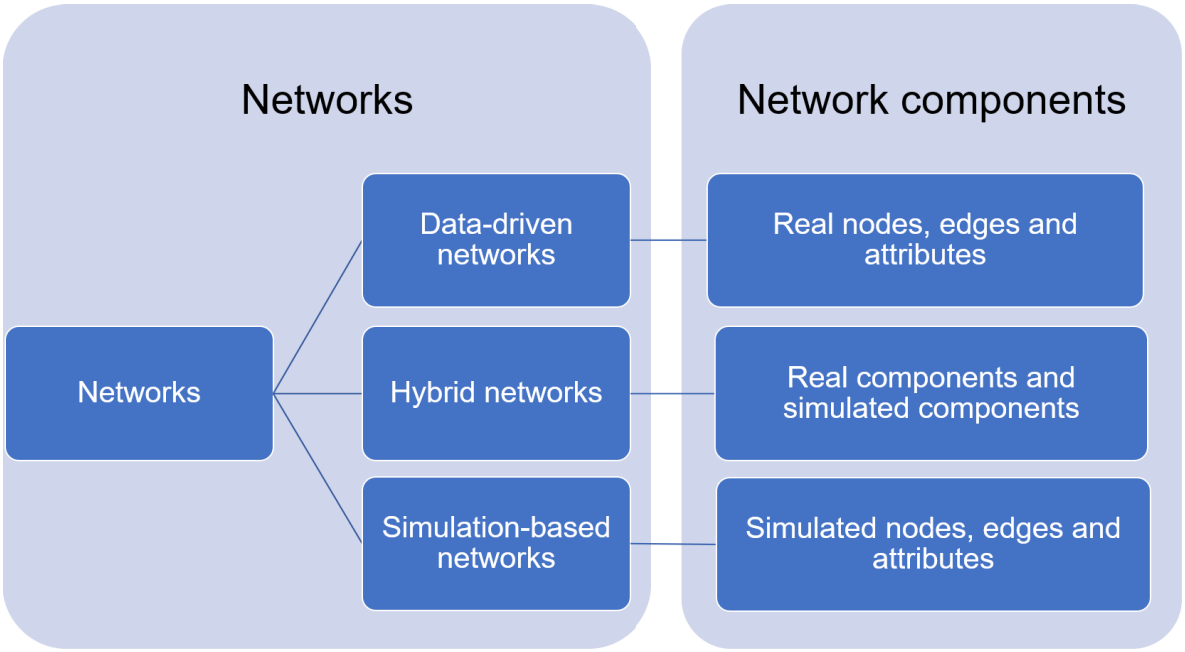


Figure 2.3: Different types of networks and their components.

and used to analyse various phenomena [18]. These simulated networks enable to model network phenomena with different levels of complexity. The statistics-based simulations based on predetermined network statistics involve classical examples like the Barabasi-Albert model for the scale-free network [45], the Watts-Strogatz small-world model for the small-world network [61] or the Erdos-Renyi model for the random graph network [192, 193]. The simulation-based approaches increase their complexity and flexibility as more complex rules governing creation of structure and its dynamics are introduced. The capability of generating networks with distinctive features enabled the similutaion-based approaches to become a universal tool for topological feature analysis. They also help with assessing the impact of network topology on dynamic processes, involving researches like synchronization of IoT systems using networks ranging from scale-free to small world models [109].

Another type of simulation-based networks, the principle-based simulations, are built according to different connection principles like homophily [47, 56], triadic structure [47], geographic proximity [64, 199]. These networks have typically higher degrees of temporal and dynamics complexity than statistics-based network simulations as they self-evolve with highly autonomous and interpretable edge formation process and generate temporal networked information. However, few researches involve principle-based network simulations with various edge attributes [34] or network simulations embedded in spa-

tial dimension [68]. They are more complex but closer to reality, which calls for future research on network simulations with the combination of different levels of complexity from structural, temporal, spatial and dynamics space.

There are also networks built using combination of real data and simulations. They are featured with the mix of different levels of complexity either captured from real data or represented via simulation. Such networks consist of real attributed nodes and connections between the nodes are created using network simulation dynamics. For example, [66] build social networks by extracting necessary information about nodes' attributes from real database and simulate the edges using scale-free networks based on a network density from referenced literature. Some researches simulate missing edges for real networks, where the single imputation methods including null-tie imputation and reconstruction summarised by [200] are typical examples. The complexity of networks built using this hybrid approach increases as more data-driven features are captured by having the network simulation dynamics trained to fit observable real network components. For example, as an improvement of single imputation, researches propose a multiple imputation method that fits an exponential random graph model (ERGM) to the real data and simulates missing ties via inference [200].

2.3.2 How to generate networks using different modelling paradigms

Networks, no matter to what extent real data or simulation is used, are created to faithfully represent information about a given system in preparation for further analyses. Depending on the modelling goal, networks used for analysis will differ with respect to the four introduced above complexity dimensions and will have varying settings of nodes, edges and attributes. There are several modelling paradigms that are used to obtain desired networks and those include rule–based, agent–based and event–based approaches that focus on fundamental generative process from a local perspective of network formation, as well as basic graph–based, probabilistic graph–based and network–embedding based approaches that aim at condense network representation from the global perspective. These modelling paradigms, of local or global level, differ in ways of observing and processing networks but converge to a faithful representation of reality that aims at minimising the information loss between reality and the model. Observed local interactions lead to global emerging characteristic behaviours observed and analysed via graphs, while graphs lose less information as the four complexity dimensions

are introduced via local level observations.

2.3.2.1 A Local view

Modelling paradigms of networks from a local view focus on mechanisms ruling the network formation. They take local level observations as a starting point and explore how they construct the characteristic CNS at a global level, corresponding to the discussion of observability considering the structural, temporal, spatial and dynamics complexity. The rule–based, agent–based and event–based approaches at a local level are introduced below with more information observed or simulated for the four complexity dimensions.

Rule–based paradigm generates networks under explicit dependence laws from predetermined assumptions or rules detected from real world cases, which not only focuses on network topology, but also investigates dynamics of network structures via simplified rules of edge formation and evolution.

Rule–based paradigms are controlled by a limited number of parameters according to a rule-based mathematical function, where randomness is introduced via certain probability distributions involved in edge formation or variable changes, like the scale-free network generation via a scale-free power-law distribution [23], edge formation with a bias probability dependent on the similarity of node attributes [47], and the changes of node property according to a uniform distribution [34]. For rule-based paradigms, there is inevitable information loss resulting from the divergence of rules and reality, as the rules greatly simplify the partially observable and complex real world scenarios. Researchers have made much effort to bridge this gap, involving studies on rule-based dynamics transforming from simulated [45, 61, 192, 193] to trained to better fit real networks [47], as well as rules evolving from statistics-based [200–204] to principle-based [47, 205]. They include temporal complexity by incorporating temporal changes of network topology into the rule–based network generation process, and sometimes, to better characterise network generation, introduce the impact of node attributes with increasing structural complexity.

Rule–based paradigm started from simulated dynamics of networks either based on statistical rules [45, 61, 192, 193] or principles [47, 56, 206](as is mentioned in section 2.3.1.2.), where they can be tuned to approach reality via seeking optimal parameters for the rule-based mathematical functions to fit real data and make inference [47]. For example, the ERGM involved in multiple imputation of networks can fit real data and simulate missing ties via inference [200], where the involved dynamics represent data-

driven features to some extent, but neither preserve rich information about node/edge attributes nor explore topological information other than typical interactive patterns of ERGM rules.

Statistics-based network generators are able to fit real networks via statistical inference based on an explicit likelihood function, like scale-free model [201], ERGM [200, 202] and geometric branching growth model [204]. As statistics-based dynamics on typical interactive patterns can hardly represent diverse real-world networks, principle-based network generators have been studied with their flexible and adaptable design of principles, which fit the statistics of real networks via likelihood-free inference such as approximated Bayesian computation [47]. For example, scale-free structure is empirically rare in social networks [201], while a principle-based simulator based on cumulative effects of triadic closure and homophily is able to reveal social network dynamics [47].

Event–based paradigm refers to the network representations with two elementary components: nodes and their local pairwise interactions referred to as events [24] or nodes of events and their logical relationships [207–209], which involves event-based network representation and analysis [208–210] as well as stochastic point processes for events that can perform network inferential tasks [33].

Event–based paradigm started from a simple network representation that captures richer heterogeneous information of interactions related to events with increasing structural complexity. As a typical example, event-based social networks (EBSNs) are enabled to further capture and use the information of offline social interactions, in addition to the online social relationships included in conventional online social networks [210]. The enriched information observed about events assist further analysis and modelling tasks. For example, the consideration of both online and offline interactions for EBSNs provides adequate information for global level analysis and modelling via graph-based models [209] and improves the prediction power for event recommendation tasks [210].

Then it comes to stochastic point processes that are enabled to make inference about nodes [33] or edges (events) [35] via estimating dependencies between events or dependencies between events and latent space models with increasing dynamics complexity. Taking Hawkes process as an example, [33] employ a self-exciting point process on the edge to perform an identity-inference task, considering the effect of available observations of geographically distributed interactions (edges). [35] apply mutually exciting point process on the edge that includes the effect of node-specific latent vectors to assess the significance of previously unobserved connections for anomaly

detection tasks.

Agent–based paradigm applied in the context of complex networked system refers to interactive structures that are typically composed based on three basic components: agents, interaction rules, and space (this could be geo-space or some other abstract space) [211], which enables more degrees of freedom in building networks from rich data or simulations as they can select more detailed information about interactions from microscopic perspective of agents.

Agent–based paradigms vary in degrees of information loss with different dimensions and degrees of complexity required for various research goals. The interactive structure can either be static with neighbour sets determined once and for all, or dynamic with evolution along time depending on model assumptions [25], while the interaction rules may either be simulation-based under certain constraints of space [17, 51, 212] or data-driven based on realistic scenarios [52, 213, 214].

The information about interactions of agents is initially represented in a fixed network topology, where decisions of agents are affected by network structures over which they interact. Such networks serve as an environment and a constraint for agents' behaviours [25]. Most static networks involved in the researches on agent-based models are simulation-based, focusing on the network effect on interactions in agent-based systems like trading behaviours in double auction [17, 50] and tax compliance and evasion [212]. Recently, these agent-based networks approach reality by introducing real data with features of time and space. [213] proposed a data-driven agent-based model for forecasting emerging infectious diseases, where data-driven networks are built with spatial information and social contact of realistic synthetic population. [214] also proposed an agent-based computational model under a data-driven decision-making framework for supply chain networks given their complicated micro structures, macro emergencies and dynamic evolution.

Multi-agent systems are also important simulation tools for modelling evolving networked systems with multiple agents interacting under certain constraints. [110] proposed a multi-agent system that replays the evolution of a network and reproduces the rise and fall of communities with the strength of adapting to real-time, changing problems. [215] researched on the constrained concensus and optimisation of multi agent networks, where multiple agents align their estimates with a particular value over a network with time-varying connectivity in different local constraint sets.

To better capture the structural patterns and instantaneous dynamics of networks,

event-driven models (also named as activity-driven models) have been proposed with an activity potential, a time invariant function characterising the agents‚Äô interactions and encoding the instantaneous time description of the network dynamics [216]. These paradigms include the rich information observed or simulated from microscopic views of agents and model their connections activated by an event trigger, involving observations varying from binary interactions [216] to simplicial complexes [63], as well as applications ranging from event-based consensus [217] to contagion problems [63, 218].

2.3.2.2 A Global view

Modelling paradigms of networks from a global view aim at representing high-dimensional and heterogeneous networked information in a way that can be easily analysed. More information can be observed at a global level as more information about interactions is captured via local methods, which involves discussion on how global performance can change through controlling and modifying those interactions, corresponding to the research aim about controllability. The basic graph–based, probabilistic graph–based and network embedding–based methods at a global level are introduced below with increasing complexity as they are able to preserve more observable information in network representation and modelling.

Basic graph–based paradigm is based on the graph theory and it can be seen as a set of selection principles for microscopic laws of behaviour in network science [219] which typically involves a simplified graph representation and analysis of networks just concerned with nodes and their connections, e.g. [220, 221].

Graph theory began when, in 1735, Leonhard Euler presented the first mathematical demonstration based on geometry of position to solve the seven bridges of Königsberg puzzle[222, 223]. Graph theory focuses on providing rigorous proofs for graph properties, such as graph enumeration, coloring, and covering [219, 224], while the evolution of random graphs motivated graph theory to generate a new branch of network science for a separate direction: quantifying the structure and dynamics of real-world complex systems [219].

Graph-based paradigm simply represents networks as basic graphs composed of nodes and connections, enabling readily available graph analysis but taking the cost of information loss to varying degrees especially in terms of structural complexity [220, 221, 225]. For example, from an accumulation of experimental data on biomolecules, graph-based models for cell biology build the graph only with cellular components (nodes)

and their interactions (edges), which allows for network topology analyses using graph-theoretical concepts but lose information other than the graph-structure [220].

Graphs are one of the widely studied data structures in computer science and discrete mathematics [226], while the graph-based models are also widely applied in the modelling of CNS across disciplines, such as analyses of graphic characteristics for networks in cell biology [220], anomaly detection of computer networks using protocal graphs [225], graph representation of vulnerability relations for industry IoT network [221].

Probabilistic graph–based paradigm models networks with uncertainties on the relationships between nodes [227], which has two elementary components: a graph that defines the network structure and a set of local functions whose product is the joint probability of this compact representation [228]. The network representation with probabilistic graph–based paradigm is typically flexible in directed or undirected, static or dynamic dimensions, each corresponding to varying degrees of structural complexity and temporal complexity. The exploration of the local functions for dependence rules or cause-effect relationships enables the inference and learning of real networks in sophisticated models [229].

The probabilistic graphs can either be undirected with symmetric relations like conditional random fields [230] and Markov networks [231, 232], or directed with cause-effect relationships between the nodes, such as sigmoid belief networks [233], Bayesian networks [234] and hidden Markov model [235]. The usage of directed or undirected graphs depends on the application scenarios. For example, Markov networks, as a typical undirected graph, represent relational dependencies without the hindrance of acyclicity constraint and thus are well suited for discriminative training [231]. They have been widely used in argumentation tasks like finding labellings or probabilistic inference tasks by deciding credulous and sceptical acceptance [236]. While Bayesian networks, as a directed acyclic graph with explicit cause-effect assumptions for interactions [234], can handle problems including missing data, prediction and data over-fitting [237]. Some researchers also employ Bayesian networks in DTs due to efficient linear computation [150, 238, 239], while nonparametric Bayesian networks are more flexible in capturing time varying features with computation efficiency [44, 240].

Probabilistic graph–based paradigm can either be static or dynamic based on whether the probabilistic graphs represent variables at a time point or across different times [228], such as the Bayesian networks [234] versus the dynamic Bayesian networks [238]. They are transforming from static to dynamic with the trend of modelling networks with

more temporal complexity. For example, continuous time Bayesian networks [241] and continuous time Markov networks [242] are proposed to capture changeable variables given the evolving features of structured stochastic processes and the graphical structure of Markov networks and Bayesian networks.

Probabilistic graph–based paradigms are able to conduct inference that helps in answering different probabilistic queries based on the model and some evidence as well as learning that estimates the graph structure and parameters of probabilistic graph–based paradigms' local functions [228]. As exact inference is often intractable, researchers tend to use approximation algorithms to find distributions that are close to the correct posterior distribution [243], like Gibbs sampling [244] and belief propagation [245]. While maximum likelihood estimation methods and the expectation maximization (EM) algorithms [246] are respectively employed for learning problems with or without hidden variables. To deal with data uncertainty and to integrate imprecise data, fuzzy probabilistic regression models are employed to model the functional relationship between the entities [247, 248]. However, probabilistic graph–based paradigms have difficulty representing, inferring and learning high-dimensional, heterogeneous networks, which calls for combined application of network embedding methods.

Network embedding–based paradigm aims at network construction and network inference via embedding node information in the network into low-dimensional space [249], which is featured with a concatenation of an encoder and a decoder [19, 250, 251].

Network embedding starts from dimensionality reduction techniques that are also applicable in scenarios other than networks, which include stochastic multidimensional scaling [252], isometric mapping (ISOMAP) [253], principle component analysis (PCA) [254], linear discriminant analysis (LDA) [255], stochastic neighbor embedding (SNE) [256] and t-distributed stochastics neighbor embedding (t-SNE) [257]. There are also basic models that merely focus on network embedding tasks, which, as summarised by [258], are respectively built upon the skip-gram models [259] and matrix factorization models [260]. These models are able to encode network information into a low-dimensional space, but do not focus on decoding information to reconstruct networks or perform inferential tasks.

To preserve more information in the modelling process and perform inferences, deep learning (DL) techniques have been widely utilised to embed highly diverse, heterogeneous, high-dimensional network information into a low-dimensional latent space. DL-based network embedding, also referred to as a network representation learning,

is able to make inferences and assist network analytic tasks including node classification, link prediction, clustering, recommendation, similarity search and visualization, involving unsupervised and semi-supervised learning methods [261]. As it is categorised by [262], DL based graph embedding methods are either random walk based like Skip-Gram based deep learning models [263–265], or without random walk but directly utilising DL methods on a whole graph or its proximity matrix via autoencoder [266], deep neural networks [267] and graph convolutional networks [268].

Graph Neural Networks (GNN) refer to a widely used DL-based embedding method that encode graph structures via a neural networks architecture, which is able to decrease information loss by aggregating features of neighbouring nodes together [19]. This motivates the combination of GNN with other models, which is typically composed of an encoder, a generative model and a decoder. For example, graph Bayesian networks [269] and graph Markov Neural Networks [270, 271] can be employed to infer graph parameters (statistical relational learning) based on encoder/decoder - graph convolutional neural networks = and generative model = the involved PGM. The GNN combined with ordinary differential equation (ODE) is able to infer the latent states of irregular observations (encoding), learn the state transition in latent space via generative model = ODE, and make predictions continuously (decoding) [272].

2.3.2.3 An Overall view

Three trends can be discovered if we look through the modelling paradigms of generating networks from either local or global view: (i) increasing complexity of network topology and dynamics, (ii) decreasing interpretability with the increase of complexity, (iii) models' aims transforming from abstract to specific. These trends are shown in Fig. 2.4:

Complexity of networks increases when they are represented and generated more faithfully with less information loss.

Models at a local level focus on the network formation process. Rule–based paradigms are usually statistics-based and generate static networks that are just composed of fixed nodes and edges [23]. They recently turned to be principle-based while allowing for the introduction of node attributes with increasing structural complexity and edge addition over time with increasing temporal complexity [34, 47]. Given the above complexity dimensions, rule–based paradigms generate networks under the simplest rules of edge addition with lowest level of dynamics complexity. To incorporate more information about attributes and temporal changes of network topology including edge addition and

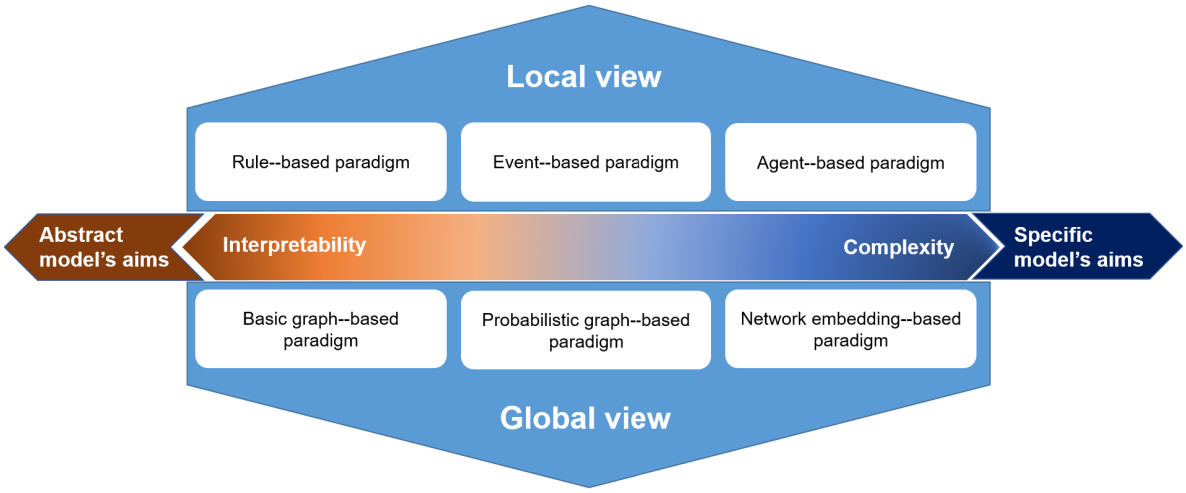


Figure 2.4: An overall view of modelling paradigms

removal, event–based paradigms based on more complex rules of local pairwise interactions, such as the stochastic point processes, are employed with increasing dynamics complexity [33]. Agent–based paradigms introduce more complex interaction rules to account for more detailed information of agents (nodes), including their attributes and various actions that result in edge addition or removal over time [110].

Models at a global level aim at a condense representation of high-dimensional and heterogeneous networked information. Basic graph–based paradigms focus on a simplified graph representation of networks that are just about nodes and their connections [220]. Probabilistic graph–based paradigms further incorporate information of network components with increasing structural and temporal complexity via the introduction of node attributes, edge directions or addition of edges over time [11, 228]. These models, with increasing dynamics complexity, also enable the inference and learning of real networks via modelling the uncertainties on the relationships between nodes [227]. Network embedding–based paradigms can conduct network construction and network inference via embedding highly diverse, heterogeneous, high-dimensional node information into a low-dimensional space. This is characterised with the highest level of complexity in terms of network representation and dynamics.

Interpretability represents the ability to explain or to present in understandable terms to a human [273]. The interpretability of networks is about the understanding of network representation and the corresponding network dynamics. It decreases as the network complexity increases with more information represented and modelled. From a local view, more complex rules of network formation enable smaller information loss but

may result in less interpretable global emerging characteristic behaviours observed in networks. For example, compared with rule–based paradigms, agent–based paradigms may generate less interpretable networks as they incorporate the impact of various observables from microscopic perspective of agents [110]. From a global view, more complex models can embed more complex networked information but are characterised by less interpretable embedding process and inference results, such as the network embedding paradigms which are less interpretable in encoding or decoding but can represent highly diverse and heterogeneous networks.

Measurability represents the ability to measure a characteristic of a class of objects [98]. Measurability of networks generated for abstract model's aims is about measuring the similarity between the inner rules of a networked system and that of real networks, while the measurability of networks under specific model's aims focuses on the measurable output for external tasks. Networks of higher interpretability but lower complexity, like non-attributed static networks generated by rule–based or graph–based paradigms, can be easily analysed, measured and compared in terms of network components and dynamics. These networks are usually involved in researches with abstract model's aims including topological feature analysis [23, 45, 49, 274] and mimics of reality [202, 222]. Networks of high complexity but low interpretability like the attributed temporal networks are always generated by event–based, agent–based, probabilistic graph–based or network embedding–based paradigms. They incorporate rich information especially in structural and temporal dimensions while preserve enough dynamics complexity for the fulfilment of specific model's aims including link prediction, node classification, community discovery, anomaly detection, synchronization and controllability.

To conclude, modelling paradigms of network representation, either from a local or a global perspective, get more complex with compromised interpretability while transferring from the abstract models' aims to the specific model's aims. Such a trade-off between complexity and interpretability in the current CNS models does not indicate their incompatibility but reveals a research direction where new CNSs models can be both complex and interpretable. And the division between the abstract and the specific model's aims will also be eliminated when CNSs are modelled as a Digital Twin and are enabled to undertake multiple external tasks thanks to faithful representation of real systems.

2.4 How to model the dynamics in networked systems

In this section, we explore the answers to the question "how?" about the modelling of dynamics in complex networked systems. This involves a discussion about two key concepts of complex networked systems: (i) dynamics of networks and (ii) dynamics over networks. Dynamics of networks considers the network's generation process and network's changes over time (evolution), while the dynamics over networks refers to the dynamic processes that occur on networks over time. In addition to the process and network complexity, we also emphasise the complexity resulting from various interrelations between dynamic processes and network dynamics, where unilateral or mutual influence from one-way or two-way interactions are reviewed and discussed. Given complex networks (obtained via methods from section 2.3) and observable information of dynamic processes, their dynamics and interrelations should be modelled in a way that models' aims (section 2.2.2) are fulfilled.

2.4.1 An overview of dynamics and their interrelations

Taking the networks obtained via modelling paradigms introduced in section 2.3 as a starting point, the modelling of dynamics within the CNS model, at the appropriate complexity level as per complexity dimensions introduced in section 2.3.1, enables a deeper understanding of changeable network structures and processes over them with and contributes to achieving certain models' aims under the constraints of observability.

The complexity of modelling CNS dynamics comes from three primary elements: the modelling of networks, processes over networks, as well as the interrelation between the two. Their states can be respectively represented as G_t, P_t and I_t, with the dynamics of and over the networks modelled as $F_g(x_t)$ and $F_p(y_t)$. The parameter x_t and y_t of these two dynamics and their interrelation I_t, with the capability to change over time, result in the complexity of modelling CNS.

$$\begin{aligned} & \begin{cases} G_t &= F_g(x_t) \\ P_t &= F_p(y_t) \end{cases} \\ s.t. \quad & I_t(x_t, y_t) = c \\ & x_t, y_t \in R \\ & c \in \{R, \mathrm{NA}\} \end{aligned} \tag{2.2}$$

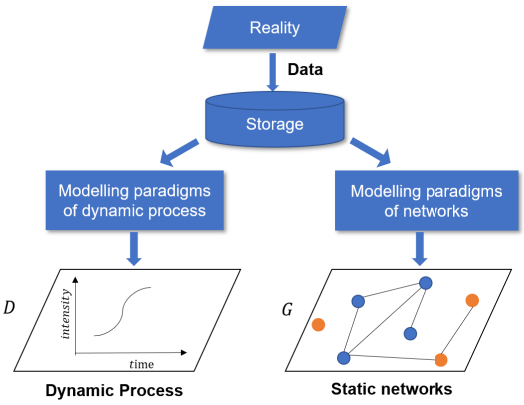
To have an overview of various CNS models resulting from those three elements, we introduce a 5-generation modelling framework (Tab. 2.2 for brief descriptions and Fig. 2.5 for details) to navigate a pathway through different levels of complexity of CNS models.

As is shown in Fig. 2.5(a), in the first generation (generation 1), the research concentrates on static networks, where spreading processes are introduced without changing their parameters. The static networks with fixed nodes and edges just provide necessary spatial or structural information rather than one-way influence that might trigger the parameter changes (evolution) of dynamic processes. Some studies employ spatial networks that utilize a metric to embed spatial information [183, 185], where the dynamic processes are modelled under the constraints of space, like the human activity on transportation networks [185]. Majority of research on dynamic processes over complex networks can be classified as the first generation of models [18].

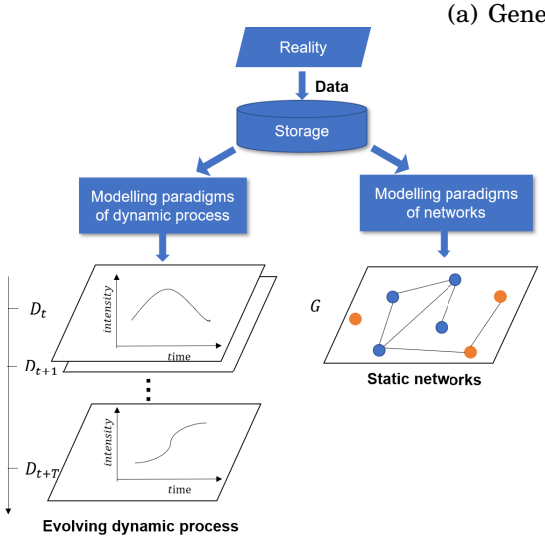
Assuming a virus a spreading in the society where all relevant information is observable, in the generation 1, we can simplify this scenario into an epidemic spreading process of virus a with fixed parameter (e.g. infection rate) on a static non-attributed networks that are just built with fixed nodes (e.g. people) and edges (e.g. social contact). This simplified CNS model increases complexity in structural and spatial dimensions once we incorporate more information about node attributes (e.g. age, gender, location, etc.) and edge attributes (e.g. direction, weights, distance, etc.).

With the temporal complexity introduced in CNS model in the second generation (generation 2), parameter changes of dynamic processes or network topology changes over time can be observed via snapshots in discrete time steps. In the generation 2a, CNS model represents the interactions via evolving dynamic processes (captured in snapshots) and static networks (Fig. 2.5(b)), where the spreading process on static networks can evolve through parameter changes [275]. In the generation 2b, CNS model represents the interactions directly via evolving networks (captured in network snapshots) and dynamic process with no parameter changes (Fig. 2.5(c)). In this context, researchers aim to transform the snapshots of CNS into latent states and model their transition process discretely, involving the networks that switch arbitrarily between different adjacency matrices according to stochastic mechanisms like positive linear switched systems [276] and Markov switching rules [29, 30].

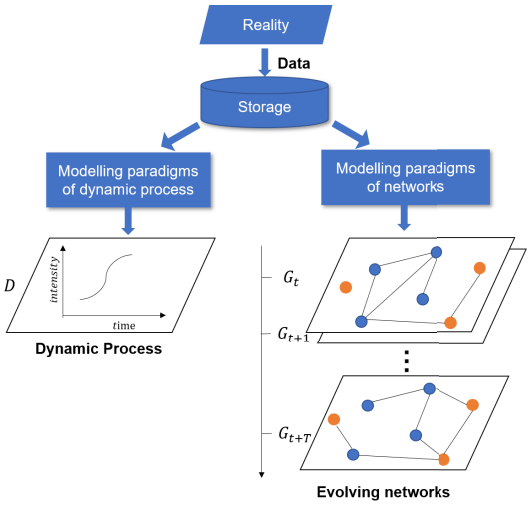
Taking the above mentioned spreading of virus a as an example, in the generation 2a, a parameter (e.g. infection rate) of virus a can change just arbitrarily or according to an external factor outside of CNS (e.g. temperature, etc.), while in the generation 2b, edge



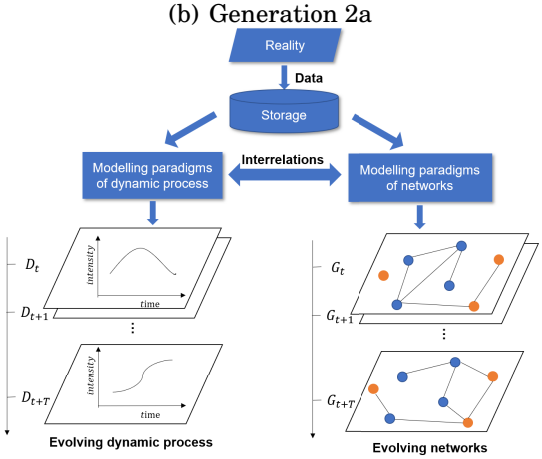
(a) Generation 1



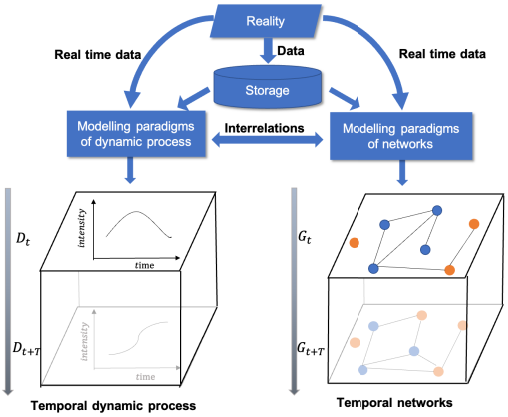
(b) Generation 2a



(c) Generation 2b



(d) Generation 3



(e) Generation 4

Figure 2.5: Generations of modelling framework

addition (e.g. social contact over time) or edge removal (e.g. enabled by the implemented policies of social distancing) can naturally lead to observations of evolving networks.

The modelling framework steps into the third generation(generation 3) as the dynamic process starts to co-evolve with the evolving networks (Fig. 2.5(d)). The involved evolving dynamic processes are captured via evolving snapshots with discrete parameter changes, where the changes of parameters and the transition of the latent states of evolving networks can either be independent or interrelated. The interrelations, including unilateral and mutual influences, enable the modelling of co-evolution of dynamics and greatly increase the complexity of modelling CNS. For example, [277] model the co-evolution process of dynamic social networks and the opinion migration on networks via introducing mutual influence between these two dynamics, where opinion migrates based on social structure and social networks evolve considering the similarity of opinions. Generation 3 can easily collapse into the generation 2a when there is only one snapshot of dynamic processes and no interrelation is considered and vice versa for the generation 2b.

To have a better illustration of the interrelation in the generation 3, we also start with the simplest scenario where the virus a spreads on a static social network (a single snapshot of evolving networks). In the generation 3, we can allow the network topology to change over time in response to the spreading of virus a (e.g. people die of virus a and are removed from the social networks). These evolving networks are interrelated with dynamic processes and are characterised with increasing temporal complexity. The parameter (e.g. infection rate) of virus a can also change according to the network topology (e.g. number of social contacts, etc.) or node attributes (e.g. age, gender, etc.). In this way, the complexity of the CNS model also increases with an introduction of the evolving dynamic process and interrelations. The CNS model will become even more complex if we further incorporate a vaccination b in the modelling framework of the generation 3, where the spreading of the vaccination b can affect the node attribute (e.g. vaccinated or not) and directly change the parameter (e.g. infection rate) of the virus a.

The modelling framework then goes through a fundamental change in the fourth generation (generation 4) as the time gaps between the CNS snapshots are narrowed, in the limit, to zero and the co-evolving dynamics are represented and modelled continuously with the introduction of real time data. As is shown in Fig. 2.5 (e), based on both historical data from a data storage and the real time data, the temporal networks capture and model all the instantaneous interactions, while the temporal dynamic processes are modelled with continuous parameter changes. The complexity of the CNS model increases as

more complex interrelations between continuous dynamics are introduced and modelled. There are studies on modelling continuous changes of networks via introducing ordinary differential equations (ODE) to the GNN methods [272], but none of them involve the most complex case of modelling continuous co-evolving dynamics of interrelated dynamic processes and dynamic networks. CNS models in the generation 4 can approach the ultimate goal of DTs when the model output of CNS models is additionally fed back to and can influence the reality in a real time manner as a reference for practice (Fig. 2.6). We refer to this scenario as the fifth generation of modelling framework (generation 5), where a closed feedback loop of real time monitoring, simulation, forecasting and deriving solutions for reality is formed and enables the CNS models to approach DTs as an extension of reality.

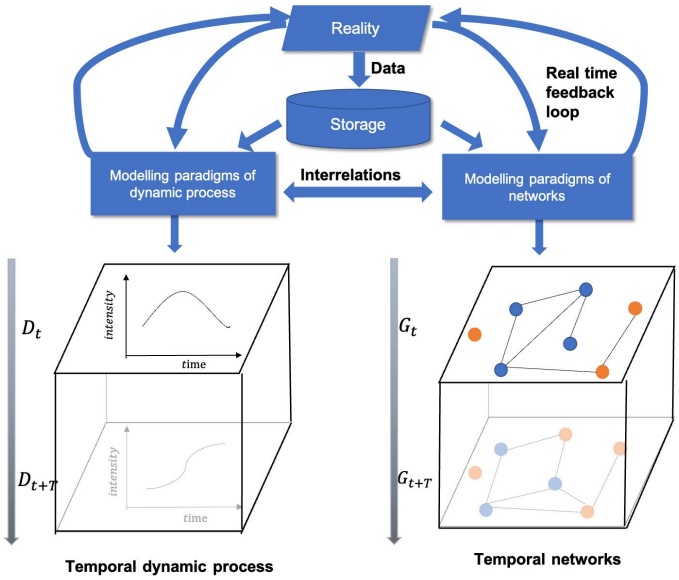


Figure 2.6: Generation 5: the temporal dynamic processes on temporal networks, with interrelations and the real-time acquisition and feedback between the SNS and reality.

To be more specific and continuing with our illustrative example, in the generation 4, the constant event streams about the spreading of the virus a and the temporal networks can be observed and modelled simultaneously. The instantaneous social contact and the virus a spreading can be captured instantly. CNS models in this context can react to sudden changes observed in reality and evolve in real time. Any changes to the temperature, recorded all the time, can trigger continuous parameter (e.g. infection rate) changes of the spreading virus a. The vaccination b can also be introduced at any time and trigger network attribute (e.g. vaccinated or not) change right away. In generation 5, the simulation result of CNS models about the spread of the virus a can be fed back to the

Table 2.2: Summary of CNS modelling generations

Generation		Brief description
First generation	(generation 1)	Dynamic process on static networks
Second generation	(generation 2a)	Evolving dynamic process on static networks
	(generation 2b)	Dynamic process on evolving networks
Third generation	(generation 3)	Evolving dynamic processes on evolving networks with interrelations between them
Fourth generation	(generation 4)	Temporal dynamic processes on temporal networks with interrelations between them and the acquisition of real time information
Fifth generation	(generation 5)	Temporal dynamic processes on temporal networks with interrelations between them, as well as the real time two-way feedback between the reality and the CNS models, enabling an idealised state required by a DT

policy makers and trigger, for instance, a launch of a "promote vaccination b" campaign, where the spread of vaccination b will be simultaneously monitored and modelled by CNS models with a real time output about social networks fed back to the policy makers.

The above mentioned modelling framework shows with examples of real scenarios that CNSs can be represented and modelled with increasing complexity through generations and finally reach the goal of a DT in generation 5. To be more aware of the progress of studies on CNS models under this modelling framework, we need to review and discuss what kind of model's aims can be achieved by modelling the dynamics in networked systems, and how these dynamics are modelled in network dimension, process dimension and both of these dimensions.

2.4.2 Modelling Dynamics of Networks

Dynamics of networks can result in networks with different structural characteristics and their changes over time. Their modelling fulfils models' aims via analysing and learning these resulting patterns and changes of nodes, edges and attributes. There are overlapping areas between modelling dynamics of networks (in a way that models' aims can be fulfilled) and modelling paradigms of network generation (in a way that minimises information loss), as a good model that approaches real network dynamics can naturally accomplish these two tasks simultaneously. Studies on modelling dynamics of

networks start from patterns of non-attributed static networks and then turn to dynamic networks that consist of nodes and edges that change over time. Those networks include temporal information about characteristics and behaviours of their components. Time dimension is introduced to break the assumption of static networks with fixed nodes and edges [12, 19, 20].

Modelling of the network dynamics, considering the patterns and changes of nodes, edges or attributes from a structural view, can be categorised as models that allow for: (i) no changes, (ii) topology changes, (iii) attribute changes, (iv) both attribute and topology changes and (v) structural pattern changes. They differ in whether network structures change and how they change over time, each to varying degrees coping with the models' aims including prediction and classification of network components, as well as the pattern discovery of network structures. As networks with only attribute changes are only involved in studies of dynamic processes on networks, we only discuss models of network dynamics for category (i), (ii),(iv) and (v) just from the perspective of a network dimension.

2.4.2.1 No changes

The modelling of network dynamics starts from static networks without any changes of structural components over time and focuses on the prediction of unobservable network components.

Networks built only with fixed nodes and edges are of the least complexity within this no change category, which involves missing link prediction fulfilled with approaches that are applicable for static networks. Some techniques take a link prediction of these networks as an unsupervised ranking problem based on a score for each non-observed link, which can be calculated either with a structural similarity index or probabilistic and statistical functions [11]. Structural similarity–based methods assume that nodes tend to form links with other similar nodes, which solely consider the local or global topological information of networks [10]. Probabilistic and statistical function–based models abstract the network structure and then predict probabilities of the missing links using the learned model [11, 278], where rule–based modelling paradigms in section 2.3.2.1 and probabilistic graph–based paradigms mentioned in section 2.3.2.2 can be employed to fit network topology for a probability score of each link. There are also techniques that take link prediction as a supervised binary classification problem about whether each pair of nodes are connected or not, where machine learning models like logistic regression and decision tree can be applied to account for the effects of different topological similarity

metrics [279, 280]. Another universally applied supervised approach is the network embedding modelling paradigm mentioned in section 2.3.2.2, where the low-dimensional latent space representation of nodes is learned and their connections can be inferred via dependencies of latent space.

Networks increase in structural complexity when attributes are gradually introduced for nodes and edges. Some approaches use node attributes to assist prediction of links [97], where above introduced basic approaches for link prediction have been altered to incorporate node features. For unsupervised ranking algorithms, as is summarised by [279], the similarity score between two attributed nodes can be calculated to assist a link prediction using methods including vertex feature aggregation [281], kernel feature conjunction [282], extended graph formulation [283] and generic SimRank method [284]. In addition, probabilistic and statistical function–based models can also incorporate attributes to model the probability of links between each pair of nodes [279]. Supervised models that take link prediction as a binary classification problem can also be improved based on the above mentioned information about attributed nodes. For example, [283] introduces node attributes via vertex feature aggregation to the machine learning algorithms like a decision tree or SVM in the link prediction tasks. There are also improved network embedding methods extensively reviewed by [249] for the link prediction task considering an effect of both network topology and node attribute. An example of such an approach is the deep attributed network embedding method designed by [285] using a deep neural network based on the topology proximity and attribute proximity.

Some studies additionally introduce attributes to the edges and transform the link prediction into a classification task for multi-relational networks. Once the transformation is completed, this problem can be approached by the above mentioned probabilistic and statistical function–based approaches and supervised learning approaches considering the similarity of nodes and dependencies of inner principles. For example, [286] uses relational Markov network to investigate the probability of link labels given the known node attributes. [287] performs link prediction in multi-relational networks using a non-negative matrix factorization algorithm based on relational similarity.

Other methods introduce attributes to the networks and focus on the node classification tasks, which employ models considering the effect including node attributes [285, 288, 289] and edge attributes such as weights [285, 285, 290–292]. Given the sparsity of graphs with fully labelled nodes and the time consuming manual labelling, most studies employ partially labelled graphs and train a classifier for the prediction of unlabelled nodes. Referring to the models for a node classification, very well summarised by [293],

we further categorise these modelling approaches into three types: (i) unsupervised learning approaches including probabilistic and statistical relational learning [290], metric modelling [291], spectral partitioning [292] and graph clustering [294]; (ii) supervised learning approaches that build classifiers like logistic regression model given features of labelled nodes [288]; and (iii) the semi-supervised learning approaches given the labels of few nodes, where network embedding methods like random walk–based [295] and GCN–based Network embedding [285, 296] are employed to account for both topological and attribute proximity effect [285].

2.4.2.2 Topology changes

The modelling of network dynamics becomes more complex as network components start to change over time. The changes of topology include the edge addition and removal, node addition and removal. As the network formation and evolution can be captured in a time series of static snapshots, modelling approaches of link prediction for static networks mentioned in section 2.4.2.1 can be extended and improved to learn one or several types of topology changes.

Studies on network topology changes start from networks built with fixed nodes and addition of edges, where the edges can be predicted and inferred via edge formation process extracted from networks over time using structural similarity–based models or probablistic and statistic function–based models discussed in section 2.4.2.1. For example, [297] employ a Markov model based on both topological and semantic features similarity between two nodes to evaluate the probability of a link. [46] predict formation of new links based on a combined popularity and similarity measure, which incorporates both global and local topological information via the introduction of Newton's gravitational law.

Modelling network dynamics becomes more complex when it comes to networks built with fixed nodes and temporal edges that are added or removed over time. Link prediction tasks for these networks focus on the temporal topology information and its evolution, where modelling approaches for prediction of missing links in static networks have been improved to account for the effect of historical topology. For unsupervised similarity ranking–based models, the link prediction task is based on the predicted similarity scores calculated using past structural similarity scores via a time series forecasting model such as ARIMA [62]. To deal with the model capacity and computational efficiency problem for probabilistic and statistical function–based models, efficient learning algorithms can be introduced to account for influence of topology, such as the proposed neighbour

influence clustering algorithm in a conditional temporal restricted Boltzmann Machine for a prediction of temporal edges [298]. As for supervised learning approaches, graph convolution network (GCN) is widely used to learn a node structure of a network snapshot for each time slide and LSTM is employed to performs a temporal feature learning for all the network snapshots [99, 299]. The discussed so far approaches only use network topology, but the heterogeneous prior information, such as node attributes, is suggested to be used to further improve the accuracy [300] but is not commonly explored in the research community. [301] propose a nonparametric link prediction algorithm that can use both topology and node labels for the calculation of linkage probability with seasonality linkage patterns.

There is an even more complex case when the networks shrink or grow in size as they evolve with temporal nodes and edges. However, as the networks can be modified to include all observed nodes from all the snapshots of temporal networks, it is always assumed that there is a fixed set of nodes for all the networks at different time points [302], where the above mentioned modelling approaches for link prediction of networks built with fixed nodes and temporal edges can be used.

To conclude, the existing studies focus on modelling topology changes where nodes are fixed. The current methods consider one of the two scenarios: (i) edge addition and (ii) edge addition or removal. There is space for further research on modelling networks with temporal nodes and edges, where addition and removal of nodes, as well as the resulting change of network sizes should be considered.

2.4.2.3 Attribute and topology changes

The modelling of network dynamics becomes more complex as we allow, on top of the topology changes, for network attributes to change over time. As current models of topology evolution are limited to networks built with a fixed set of nodes and edge changes over time, the models of networks where both attributes and topology change also have the same limitation. These models, given an addition or removal of edges, consider: (i) edge attributes changes, (ii) node attributes changes and (iii) both of these changes. Modelling approaches for the link prediction and node classification of static networks presented in section 2.4.2.1 and the network topology changes introduced in section 2.4.2.2 can be extended to learn both attribute changes and topology changes.

The topology change, linked to the edge addition or removal, can be accompanied with changes of edge attributes. There are studies on link prediction in temporal networks that have edge weights [99, 300, 302] and directions [302]. The typical supervised learning

approach for topology changes mentioned in the section 2.4.2.2, where GCN explores the local topology of each snapshot and LSTM characterises the evolving features of dynamic networks, can be improved by introducing the generative adversarial network (GAN) to tackle the sparsity and the wide-value-range problem of edge weights [99]. Network embedding methods based on matrix factorisation can also be improved to include the information about edge weight or direction into adjacency matrix for a prediction of temporal edges [300, 302]. Further research is needed for a variation of other edge attribute changes.

When node attributes change in networks built with fixed nodes and non-attributed edges that change over time, modelling approaches of node classification can be employed to learn the changes of the node attributes by considering the evolution of attributes and topology. [34] use GCN to conduct node classification task on the social networks built with a fixed number of attributed nodes, changeable node labels and non-attributed edges that are added over time. GCN implemented in this research not only considers the local topology and its attributes, but also uses the similarity–based matrices to account for patterns of high-order neighborhood. Further research is needed for node classification in networks built with dynamic labelled nodes and temporal edges. There also comes the most complex case where node attributes, edge attributes and network topology all can change at the same time. Currently, there is no research in this space and further studies are required to model this very complex scenarios.

2.4.2.4 Structural pattern changes

The above mentioned structural components and their changes result in various structural patterns and corresponding changes. Structural patterns refer to the correlated combination of nodes, edges and attributes within a community or a network. Research in the space of structural patterns includes patterns and their dynamics discovery, analysis and prediction. The common models' aims here are e.g. a community discovery and an anomaly detection.

A community discovery starts from defining a community. This characterises the structural patterns of the sub-networks to be discovered via generally unsupervised way of modelling. A community in a complex network, as is defined by [20, 128] in a generic way, is a set of entities that share some closely correlated sets of actions with the other entities of the community. To fulfil the requirement of reflecting certain features of reality, the closeness within each community can be measured based on density, vertex similarity, actions of nodes or influence spread, all corresponding to different types of

community discovery algorithms well summarised by [128].

Considering the increasing temporal and structural complexity of networks resulting from the topology variations and an introduction of attributes, the community discovery approaches have also been further developed and categorised from the perspective of process that is seen as the inner structures of algorithms. To deal with the community instability problem in dynamic networks, the temporal smoothing operations have been included in community discovery models to smooth-out the evolution of communities, which accordingly involve a new category based on varying extent of temporal-smoothness [20]. For attributed networks, built with attributed nodes and edges, a fusion procedure has been introduced to community discovery models to account for both effects of topology and node features, where another category based on when and how they use and fuse network structure and attributes is proposed [171]. To deal with the directed networks that are featured with edge directions and asymmetrical matrices, different community discovery models have been proposed and well summarised depending on the way directed edges are treated [303].

However, the perfect community discovery algorithm does not exist despite of all the above mentioned attempts, as each of them performs well on one specific declination of the general problem and can achieve different partitions even for the same networks [20]. Based on that, [304] further categorise the community detection algorithms according to the similarity of their results, which attempts to confirm the valid definitions of a community and help with the choice of algorithms for future research. There is still a need for further research on the community discovery of network variations with varying degrees of temporal complexity and structural complexity, given the superposed challenges of community definition, temporal smoothness as well as topology and attribute information incorporation.

Studies on anomaly detection focus on the rare occurrences of structural components, patterns as well as their changes, involving detection of anomalous nodes, edges, subgraphs, events and graphs [12, 112, 305]. They start from static networks built with fixed nodes and edges, where anomaly detection of nodes, edges and subgraphs can be realised via traditional non-deep learning approaches based on the network statistical features or using representation learning methods, as is summarised by [305, 306]. These methods have also been improved respectively when it comes to attributed networks with richer information about network structures [305, 306].

As temporal complexity is introduced to networks, two types of anomaly detection methods can be distinguished. More specifically, there is a two-stage approach that

maps networks into a vector of real numbers and then employs an anomaly detector on it for a node, edge or subgraph anomalies, involving well categorised community, compression, decomposition, distance, and probabilistic model–based models [12]. There are also deep learning approaches respectively applied to the anomaly detection of nodes, edges, subgraphs and graphs [305]. There is also research gap that calls for further study on anomaly detection models incorporating both attribute and temporal information of networks.

2.4.3 Modelling Dynamic Processes

In this subsection, we focus on the dynamic processes that can take place over networks. These dynamic processes interact with the dynamics of networks and vice versa. Dynamic processes can result in three types of changes: (i) a network topology change; (ii) a network attribute change; and (iii) a parameter change of dynamic processes [307]. The network topology change and the attribute change can also lead to the parameter change of dynamic processes [308]. There are many interesting variations of the combination of dynamic processes and dynamic networks, which differ depending on the research objectives and application scenarios. Corresponding studies start from single parameterized dynamic processes on the static networks and then turn to multiple dynamic processes with dynamically changing parameters on the dynamic networks, which is featured with increasing process complexity, network complexity, as well as the complexity resulting from various interrelations between dynamic processes and network dynamics

Modelling of the dynamic processes research can be categorised into the following groups: (i) a single dynamic process, (ii) independent multiple processes and (iii) interrelated multiple processes. They differ in a number of dynamic processes and how they interact, each to a varying degree mimicking the real world. Within each category, researchers either focus on parameterized dynamics or dynamically changing dynamics.

2.4.3.1 Single dynamic processes

Spreading dynamics on complex networked structures obtained via methods in section 2.3, as a typical dynamic process, is covered by considerable literature on the spread of phenomena/medium ranging from virus [309–312], meme [124], opinion [122, 123, 313], decision making [314], nutrition [315, 316] and social contagion [317].

The modelling of spreading dynamics on networks starts from classic population models with the simplest analysis that considers the evolution of the state for all individuals

rather than the state of each individual [15]. They include the stochastic population model that describes the evolution of the population state via a Markov process, as well as its approximation, deterministic population model with deterministic definitions of the population state [15, 26]. To model the states of each individual independently and allow for arbitrary interactions among them, as is summarised by [15], these population models are extended and improved to faithfully learn spreading processes on networks.

A spreading process on static networks is an example of a simplified representation of various real world scenarios where networks underlying the dynamic processes are simulated or represented under the most stringent assumption of nodes and edges that do not change. When a dynamic process evolves much faster than the network of interactions, static networks can serve as accurate proxies of slowly switching topologies. This real world situation can be approximated as dynamic processes over static networks [26].

The modelling of dynamic processes on static networks start from analysing the impact of pairwise interactions using the extended classic models, including stochastic network model and deterministic network model [15, 26], where networks with varying features and structures can be introduced to the studies. Stochastic network models assume the state transition for each node as a Markov jump process or its extension under relaxed Markovian assumptions [26, 318–320]. Deterministic network models, as approximations of stochastic network models with deterministic definitions of node states, are also widely used [321, 322]. A threshold model is one example of a typical deterministic model for an epidemic process [317]. The modelling of spreading dynamics becomes more complex and realistic as complex contagion is also considered with the incorporation of group interactions, which currently has been realised via simplicial complexes [323].

A spreading process on dynamic networks refer to one of the most common real-world application scenarios. The underlying networks of interactions are not static, but dynamically change while co-evolving with and influenced by the dynamic processes over the networks [324]. Models that capture their co-evolution at comparable time-scales have been well categorised into temporal-switching, activity-driven, and edge-Markovian networks [26]. For dynamic processes on dynamic networks, temporal switching networks model them as snapshots switching arbitrarily between a set of topologies according to stochastic mechanisms such as Markov switching rules [29, 30]. Activity-driven approaches focus on the networks' interactions generated according to a time invariant function characterising individual properties, which involves a series of extensions with the introduction of epidemic threshold due to its analytical tractability, as is summarised

in detail by [325]. Edge-Markovian dynamic graphs can model stochastic evolution of dynamic networks, which also involves analytically tractable extensions with spreading dynamics [326, 327].

Spreading dynamics on dynamic networks can also be modelled with data-driven machine learning approaches. They focus on transforming spatial information and other temporal features involved in a spreading process to well handled temporal information, where deep learning–based predictive models like Recursive Neural Networks and Convolutional Neural Networks can be employed to predict spreading process, as is well summarised by [328]. Researchers also start to employ network embedding approaches, as is mentioned in section 2.3.2.2, to incorporate network information into the predictive systems, which involves typical examples of predicting epidemic spreading with graph neural networks [27] or using node regression based on transfer learning [28].

Parameter changes of processes are discussed in terms of the above mentioned single dynamics. The existing modelling approaches to a single dynamic process propagating over the network are generally parameterized without any change of parameters implemented. Only few studies focus on the evolution of spreading processes on networks. [275] incorporate mutation of pathogen strains and corresponding changes of epidemic transmission probability, which trigger the evolutionary adaptations of the spreading processes with dynamically changing epidemic threshold. For this example, transmissibility changes between limited number of fixed values and is controlled via mutation and transition probabilities.

2.4.3.2 Independent multiple processes

Dynamic processes can result in three types of changes, including: (i) a network topology change; (ii) a network attribute change; and (iii) a parameter change of dynamic processes [307]. As almost all the models of multiple processes in section 2.4.3.1 are based on probability of state transition, the parameters of multiple processes in section 2.4.3.2 and section 2.4.3.3 refer to the transmissibility, adoptability or probability of an entity being infected or activated or an entity adopting a given behaviour or state.

Independent multiple processes take place independently without direct influence on their parameters as they ignore the dependence of multi-spread in co-infected status or just exclude the concurrent infections of multi-spread. Since no research can be found on independent multiple processes where co-infected status is allowed, we mainly focus on

independent multiple processes that exclude co-infected status and interact via changes of network topology, attributes or structural patterns.

Mutually exclusive processes refer to the multiple processes propagating over the network where the concurrent infection by more than one spread is not possible. The goal is to investigate circumstances under which the dominance of a single spread can emerge [329–331].

These dynamic processes interact on the same network while preserving their independence via assumptions including temporal separation [330], structural separation [308], cross immunity [329, 332, 333] and cross adoption [334, 335]. Under temporal separation, two pathogens can spread independently in separate time steps and interact via network topology changes, such as node removal as a result of death or immunity [330]. Studies on concurrent multiple processes generally adopt cross immunity or cross adoption assumption to deal with the concurrent infection that is not allowed for mutually exclusive competing processes. Under cross immunity, the states of an infected/recovered network vertex are changed to be immune to any other infections [329, 332, 333]. Under cross adoption, the infected network nodes can transit to be just infected by another spread with a specified probability [334–336]. There is also research on concurrent multiple processes that use structural separation, where network nodes are grouped and are only available for specific spreading dynamics, such as the simple contagion and the complex contagion that work out for predetermined vertex groups [308].

Under the above mentioned assumptions and basic settings, the models widely employed in studies of single spreading dynamics in section 2.4.3.1 can be extended and used to model independent, multiple processes. In addition, there are already studies using SI [308], SIS [332], SIR [333, 334], independent cascade model [331], percolation model [330, 333] and Reed–Frost model [329]. For example, [308] propose dynamic message-passing equations for two SI-type competing processes to incorporate the message-passing into the parameters. [332] use the extended model, SI_1I_2S, to model the propagation of two concurrent epidemic spreading processes.

Parameter changes of dynamic processes are discussed in terms of independent multiple processes. There are scenarios (e.g. election information spread and entertainment information spread in the same social networks) concerning parameterized independent multiple processes where co-infected status is allowed under the ignorance of the depen-

dence of multi-spread without changing their parameters, though no research can be found here.

Most of the mutually exclusive processes involved in the existing studies do not consider dynamically changing parameters of the process [329, 330, 332, 333]. A small number of studies introduce dynamics with parameter changes over time under the impact of node states, where transmissibility varies from nodes' groups (structural patterns) [334, 335], states of neighboring nodes [336], message-passing [308]. The parameters can either change between limited number of fixed values [334, 335] or change continuously according to the network attributes resulting from another spreading process [308].

2.4.3.3 Interrelated multiple processes

Interrelated multiple processes are characterised by direct unilateral or mutual influence of processes on their parameters. They can interact not only by changing network topology or attributes, but also via parameter changes.

Partially inclusive processes refer to the multiple processes that allow the concurrent infections of nodes while also incorporate the dependence between spreads themselves.

Interrelated multiple processes can have suppressing [16, 337] or supporting [308] relations, which are involved in only limited number of studies and so far all of them consider static networks [16, 308, 337]. These spreading dynamics change parameters with the transition of concurrent infection states. For example, epidemic spreading process, under the suppressing impact of awareness spreading, has different infection probabilities given nodes' different levels of awareness [16, 337, 338]. Similarly, collaborative multiple processes also have different probabilities under their supporting impact [308].

Under the above mentioned assumptions of concurrent infection, the models mentioned in section 2.4.3.1 and section 2.4.3.2 can also be extended and used, where some of them have already been used to model interrelated multiple processes, including SI [308], SIS [338], SIR [337] and $SIS-SIRS$ [16].

Parameter changes of processes here are discussed in the context of interrelated multiple processes. For almost all the existing research, at least one of the considered spreading dynamics has dynamically changing parameters in response to the impact of another spreading dynamics [16, 337, 338]. [308] further introduce collaborative multiple

processes that all change parameters with the message passing of nodes. Similarly to the case of independent multiple processes, parameters of interrelated multiple processes can either change between limited number of fixed values [16, 337, 338] or change continuously according to the network attributes resulting from another spreading process [308].

2.4.4 Combination of the network and process dimensions

In this section, we focus on the superposition of the network dimension and the process dimension, as well as the increasing complexity of modelling CNS considering the interactions and interrelations between the network and a dynamic process.

2.4.4.1 Superposition of networks and processes

Propagation process dynamics, either on static networks or dynamic networks, has been extensively studied using non-machine learning approaches [15, 26], where data-driven machine learning approaches have recently been a popular choice of incorporating more structurally and temporally complex network information [27, 28]. Dynamic networks involved in these studies only allow for topology changes [27–30] which influences the result of spreading dynamics.

The majority of the independent multiple processes considered in the existing and current research take place on static networks [308, 329, 331–334] and only a small number of studies can be found for those on dynamic networks [330, 335]. As a typical example of competing epidemic processes, [330] allows removal of nodes and their edges over time as a representation of death or immunity and this results in the topology changes. However, interrelated multiple processes are involved in only a limited number of studies and so far all of them take place on static networks [16, 308, 337]. Modelling approaches on multiple processes generally employ non-machine learning approaches to model single dynamics on networks, where further research is needed for modelling multiple processes using both non- and data-driven machine learning approaches.

2.4.4.2 Interactions of networks and processes

Network dimension and dynamic process dimension can either be interrelated or independent based on whether one dimension can trigger the dynamics of another dimension to change. Parameter changes of dynamic processes as well as network changes of (i) topology, (ii) attributes and (iii) structural pattern, each indicates the changes of dynam-

ics in the process dimension or the network dimension. Thus, the interrelations exist in two scenarios: (a) certain states of networks trigger the parameter changes of a dynamic process, (b) a dynamic process results in one of the above mentioned three types of the network changes. The interrelations can either be described as a one-way influence that is just about (a) or (b), or a mutual influence that refers to both (a) and (b).

Independent VS Interrelated relations between a network and a dynamic process are discussed considering their changes and the corresponding causes of change within the CNS model.

An independent relation between the network and a dynamic process is common in terms of dynamic processes on networks and the research space is dominated by this approach [15, 26, 30]. In this case, a dynamic process only influences and causes changes of the node attributes connected with the process itself (e.g. whether a node, as a result of the process, has been infected or adopted new behaviour) rather than a change of the network structure or dynamics. In this scenario parameters of a dynamic process are not altered by the changes in the network.

Networks and dynamic processes with interrelated relations between them consider their mutually or unilaterally triggered changes. Multiple processes that interact with each other via changing network attributes are typical examples of a one-way influence of the type (a), where parameters of spreading dynamics can change with network attributes [16, 308, 334–338]. For example, in a rumour-truth mixed spreading scenario, the truth spreading rate gets lower when nodes are attributed as rumour-believers [336]. There are also cases for a one-way influence of the type (b), where networks change topology in response to the spread. For example, disease spreading through the network can leave some nodes dead and get them removed [330]. Currently no research is found on interrelation about a mutual influence between the process dimension and the network dimension so the situation where a closed feedback loop between the process and the network is considered.

Parameter changes of dynamic processes are here discussed in terms of interactions and interrelations between networks and processes. All the available examples of processes on networks that are dynamically changing under the impact of networks are about multiple processes, where networks serve as a media for their interactions [308, 308, 334, 334, 335, 335, 336]. Further research is required in the space of dynamic processes that are dynamically changing under the impact of network changes.

Network topology that changes in response to the dynamically changing dynamics is another interesting and not addressed research gap.

2.4.5 Control mechanisms on CNS

Control mechanisms of CNS aim to find the optimal strategy to attain its desired state, which involves controllability and synchronization of networks and a control of dynamic processes on networks.

2.4.5.1 Network control

Controllability of networks is achieved using model–based or data–driven approaches [339]. Model–based approaches aim to find an optimal set of driver nodes for CNS model under the assumption of a tractable model for network dynamics, where a linear time-invariant systems are often employed to approximate the nonlinear processes that drive the directed networks [81, 340, 341]. This approach identifies the driver nodes via maximum matching approach [342], which enables the calculation of the structural controllability [340] and exact (state) controllability [341]. In this context, many studies investigate the impact of topology variations on the controllability of directed networks, ranging from degree distributions [340], connection types [81], topology switching [83] to even all possible network structures [342]. To measure the robustness of controllability, simulations of node removal and edge removal attacks are also conducted [342] and convolutional neural networks can be further utilised to improve computation efficiency [343]. Data–driven based approaches, on the contrary, learn controls from network data without knowing network dynamics [339]. Relevant studies generally focus on undirected, directed or weighted networks, where machine learning methods like reinforcement learning are used to find their optimal control parameters for desired network states [339, 344].

Synchronization of CNS, as another type of control towards the desired synchronized state, also involves intensive studies that are well summarised from the perspective of phase oscillator models, stability of synchronised state, and synchronisation in complete or spare networks [101, 345]. In addition, given the similar definitions of synchronization and consensus problem of multi-agent systems [346], they can also be studied from a unified view point by employing ideas about consensus problems across disciplinary areas to complex networks [347–349].

2.4.5.2 Process control

Controllability of dynamic processes is achieved via changing networks or introducing another dynamic process.

Control via changes of networks refers to the control of spreading dynamics via changing network topology and attributes, involving non-deep learning–based, deep learning–based and manual strategy–based approaches. Taking control of epidemic spreading processes as an example, researchers seek an optimal set of control actions including topology changes like node removal and edge removal as well as attribute changes via an antidote allocation to minimise infections [15, 26, 344]. Non-deep learning–based approaches, as is well summarised in [26], mimick spreading dynamics, with the non-deep learning models reviewed in section 2.4.3.1, and optimise the action sequence under corresponding model constraints by a mean-field approximation or geometric programming. Deep learning–based approaches use machine learning methods to seek the optimal action sequence over graphs based on the network information that is incorporated by the network embedding. For example, [350] control the epidemic processes over a temporal attributed network using reinforcement learning as a ranking module for actions of changing node attributes, where GNN is encapsulated to embed information about networks and epidemic process. Manual strategy–based approaches focus on the simulation and comparison of manual strategies that change network topology or attributes, like different social network-based distancing strategies proposed and compared to reduce infections of Covid-19 [64].

Control via interactions between processes refers to the control of spreading dynamics realised via introducing another spreading dynamics with their competitive, suppressing or supporting impact. For example, the disease containment of a single epidemic spreading dynamics A can be controlled by introducing its competitive process B, which is realised by an optimal allocation of a limited number of B spreaders to minimise the spreading of A [308, 331]. The advertising campaign A can also be controlled via introducing its collaborative spreading process B, where the best joint advertising campaign can be designed via the optimal allocation of spreader B with the aim of maximising the number of susceptible nodes [308].

2.5 How do we approach the ultimate goal?

In this section, we answer the question of how to approach the ultimate goal of modelling complex networked systems: building a Digital Twin (DT) for real world networked systems. This question is decomposed into three sub-questions: (1) What have we done so far to achieve the goal? (2) How far are we from building the DT for CNS? and (3) How can we move forward?

Existing research on modelling networked systems and their dynamics aims at representing the complex reality through networked structures that minimise information loss and ensure that the model's aims can be fulfilled. The complementary effect of the model's aims fulfilment and minimised information loss contributes to a good model of CNS and its convergence to a DT. To assess the networked systems models and narrow their gaps with DTs, we build an assessment framework from two perspectives: (i) the CNS model's aim fulfilment that diverges when a specific model's aims and an abstract model's aims each focuses on the external tasks and inner rules; and (ii) a DT faithful representation and modelling that helps to merge the requirement of both the specific model's aims about external tasks and the abstract model's aims about its inner rules.

2.5.1 What have we done so far?

Researchers have done a lot of work on modelling real world networked systems. A small number of studies have already attempted to develop Digital Twins of complex networked systems for specific application contexts, like IoT systems [150, 151] and blockchain-encapsulated systems [152, 153]. However, while many recent studies on modelling, simulation and control of complex networked systems started taking into account the necessary details to faithfully represent aspect of complex reality, none of them explicitly attempted to create a DT of CNS with all its implications which we will now discuss in more detail.

2.5.1.1 Bottom-up view: CNS-based attempts

Researchers have been trying to build a Complex Networked System (CNS) that can faithfully represent and adapt to the real world situation. The networks, as the basic representations of CNS, are approaching reality with more faithful representation of real world information and infusion of evolving dynamics. These attempts partially enable CNS model to meet the requirements of Digital Twins in terms of similarity to reality by incorporating complex inner rules and fulfilling external tasks.

Networks are approaching reality as the structural, temporal, spatial and dynamics dimensions are gradually taken into account. For networks constructed in a data-driven way based on readily observable data set, the description of the the impact of time enables the modelling of data-driven networks to capture the evolving feature of real-world systems, while spatial information represented in the network structure serves as space where dynamics take place. Spatial temporal networks, where temporal networks are modelled under the constrain of space that influences the structure of the networks, are proposed to encompass both temporal and spatial information, which are the closest to the reality data-driven network structures considered until now. For simulation-based networks and hybrid networks, the rule-based simulations have been developed to approach reality as more complex inner rules of complex networked systems are simulated to incorporate the above mentioned complexity dimensions. Networks simulated from microscopic views of agents enable the representation of either real or simulated information in a flexible and faithful manner. From another perspective, the networks have also been widely employed in different scenarios across disciplines, where a wide range of research objects can be represented and analysed with network structures. As is detailed in section 2.3.2, networks are employed to represent information in agent-based systems and graph structures, relations that are statistically or semantically extracted and constructed, complex systems infused in networks like IoTs and Blockchains.

Dynamics of and on CNSs with different possible interrelations between those two is another key aspect of building an accurate model of CNS. With the infusion of dynamics into the networks, the CNS model is getting further on the road to a DT. Referring to section 2.4, there are built-in dynamics that trigger the evolution of networks, while there are also dynamics that take place on the networks under the influence of network structures. CNS model composed of networks and dynamics are approaching reality as researchers try to understand and model the interrelations among network dynamics and dynamic processes. Until now, there is considerable literature on dynamic processes on networks that are either static or dynamic. Among different interrelations between dynamics and networks, we find that some studies consider the unilateral interrelation between dynamic processes and static networks. However, we can hardly find any literature on the interrelation of dynamic processes and network dynamics, which is actually the closest to the real world situation.

2.5.1.2 Top-down view: DT-directed methods

The digital twining tasks vary for different cases and require to be adopted and adapted in the context of a CNS model. For example, [38] model the DT of an urban-integrated hydroponic farm, where they decompose the modelling process into three crucial elements: data creation that enables an extensive monitoring system for a virtual representation of the farm through data, data analysis that helps to identity key influencing variables, data modelling that enables forecast and feedback. [351] try to shape the actual state and a possible future of the Product Data Technologies from a Closed-Loop Product Lifecycle Management (C-L PLM) perspective, where they see an intelligent product as a product system which contains sensing, memory, data processing, reasoning and communication capabilities at four intelligence levels. [352] view the physical asset and its digital twin as two coupled dynamical systems that evolve over time through their respective state spaces, where the digital twin acquires and assimilates observational data from the asset (e.g., data from sensors or manual inspections) and uses this information to continually update its internal models so that they reflect the evolving physical system. These up-to-date internal models can then be used for analysis, prediction, optimisation and control of the physical system. Referring to the previous studies, we generally decompose the modelling of a DT as tasks including (a) data processing that includes data creation and data integration, (b) data analysis with the purpose of parameter selection, (c) data modelling that enables forecast of eventualities and feedback of real system under the impact of the decision that is made by reference to the forecast.

Data processing, as the fundamental task of modelling a DT, is composed of two parts: (i) data creation enabled by an extensive and robust monitoring system that tracks the observable information, (ii) data integration that features with the record, management and retrieval of information in a real time. In the data creation stage, taking the DT of hydroponic farm built by [38] as an example, they track changing environmental conditions and crop growth through unstructured manual records and a wireless sensor network that sends data in real time to a server. For data integration function, semantic modelling that includes the application of ontology has also been employed to equip the DT with context awareness through record of data, answering queries and information retrieval [36, 353–357]. To accommodate and deal with a system's context and uncertainty when selecting an interaction mode that reconfigures the system's model in response to the model output, a fuzzy logic rule based layer has often been employed in complex systems, which enables the interaction mode selection while considering the

intermediate ranges between the modes [137, 150]. While there also emerges another popular data integration approach that employs blockchain technology. The blockchain can serve as the middleware of IoT with improved interoperability, privacy, security, reliability and scalability [358–360]. However, in some cases, it is hard to collect and process readily real-time data for a well-established DT in an efficient way and there are studies attempting to deal with problems of data integrity. For example, [361] propose a collaborative city digital twin based on federated learning, where multiple city DT can learn a shared model while keeping all the training data locally. This is a promising solution to accumulate the insights from multiple data sources efficiently and avoid violating privacy rules.

Data analysis and variable selection task is of importance to the establishment of DTs, which is presented by [38] as something that includes: (a) the influence of the environment on physical asset, (b) the influence of operable controls on the environment, and (c) the influence of manual changes on the operational controls, where, within the limitations of the data, this exercise identifies the variables which are crucial to track and forecast. Data analysis that enables variable selection is of importance to the modelling of DT. However, most studies ignore this process and just predetermine the observable variables, while the fundamental question of how to identify the minimum number of observable variables has been understudied over the years and needs a systematic research and answer. As is detailed by [352], a well-designed digital twin should be comprised of models that provide a sufficiently complex digital state space, capturing variation in the physical asset that is relevant for diagnosis, prediction, and decision-making in the application of interest. On the other hand, the digital state space should be simple enough to enable tractable estimation of the digital state, even when only partially observable. Specifically, as a rare example of selecting variables, [38] identify key influencing variables on energy use and crop yield by analysing the relationships between the broad data collected based on temperature, visible radiation, and CO2 levels.

Data modelling process for a DT is essentially a forecasting model that predicts and provides feedback on real world system to help control the DT, which includes two fundamental tasks, a forecast of extended future and a feedback of real-world system. Prediction and inference of reality that has happened before is the basic function of forecasting models, while DTs can further forecast the extended reality by predicting facts that have never happened before. For example, [362] propose a disaster city DT for

enhancing disaster response and emergency management processes, where disasters that have never happened before are simulated and real world systems are extensively forecast to enable increased visibility into network dynamics of complex disaster management and humanitarian actions. The digital twining of complex networked systems are also featured with a decision-making feedback loop with dynamically updated asset specific computational models infused [352]. Especially in cases of solving multi-objective optimisation problems for complex systems using DT, which are common in analyses of entire product lifecycle in manufacturing, researchers have proposed DT frameworks aimed at multi-objective optimisation with effective feedback from different dynamics. [363] enhance DTs of autonomous manufacturing systems through reinforcement learning of continuous data fed back from DT, where residual errors between DT and its physical counterpart are compensated and an improved autonomous system can be established. [364] propose a bi-level iterative coordination mechanism to achieve optimal design performance for AFMS, where an effective feedback of collected decision-support information from the intelligent multi-objective optimisation of the dynamic execution is presented. To deal with uncertain information in real life scenarios involving multi-objective optimization problems, a fuzzy programming method can be employed to obtain optimal solutions [365].

2.5.2 How far are the CNSs, as they are currently modelled, from DTs?

In most cases, CNS models presented in the literature, while fulfilling the relatively simple model's aims under relatively stringent assumptions, have not been developed with the goal of becoming DTs of their modelled aspects of reality. In effect they only posses partial features of DTs. Therefore, in order to bring the two areas closer together, we propose a unified assessment approach by discussing and attempting to answer the following three questions: (1) What constitutes a good DT? (2) What is a good CNS model? (3) To what extent the current CNS models approach a DT? We try to answer the first two questions with measures that aim at assessing the performance of CNS model and metrics that evaluate the quality of DT, as is shown in Fig.2.7, and try to answer question (3) in the context of a good CNS model that performs well under certain model's aims and thus has the potential of becoming a DT.

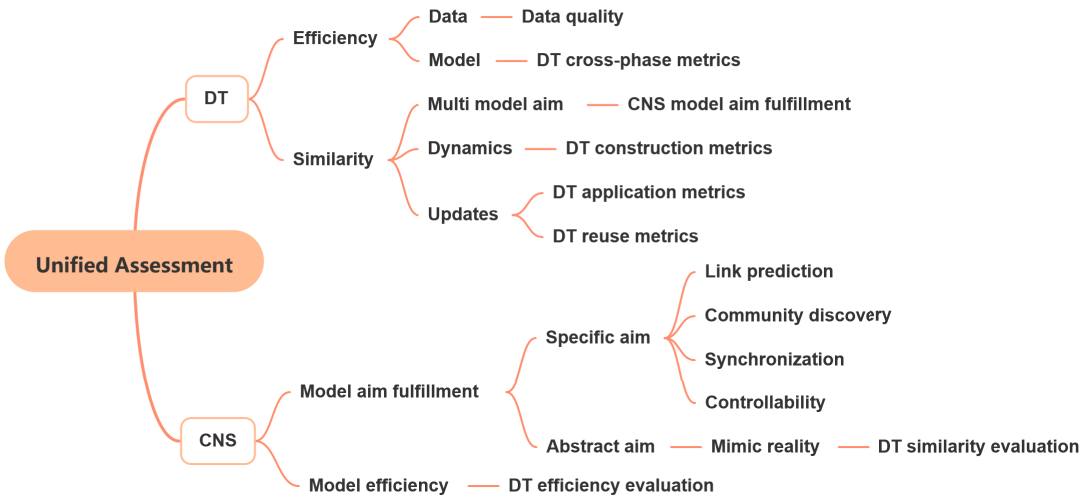


Figure 2.7: The plot of unified assessment criterion for CNS model and its distance to a DT

2.5.2.1 What constitutes a good DT and how to assess it?

DTs are featured with integrated functions like simulation, optimization and data analytics [143]. DTs use real-time processing and updates characterised by: (1) real-time connection with the physical entity, (2) self-evolution that enables a DT to learn and adapt in real-time by providing feedback to both the physical asset and the DT, (3) continuous machine learning analysis (dependent on the frequency of the synchronisation), not just a one-time output forecasting, (4) availability of time-series (or time continuous) data for monitoring, (5) level of autonomy that defines if a DT could either make changes to the physical asset itself or if it relies on a human in control who could make changes to the DT, where the property of a DT to be autonomous, not autonomous, or partly autonomous is case-dependent and (6) synchronisation which could be partly continuous or partly event-based [41]. A good DT should meet the requirement of researchers using relatively simple models while preserving the trust in the data, model and their updates [366, 367]. Based on the above features, the assessment of a good DT in the context of CNS can be categorised into two parts: (i) efficiency of data processing and modelling and (ii) similarity with reality from the perspectives of multiple model's aims, self-evolving dynamics and model updates (Fig.2.7).

The evaluation of efficiency includes data processing efficiency and modelling efficiency. Data processing efficiency involves data quality such as validity and reliability

enabled by handling the imperfection of real-time data ranging from imprecision, uncertainty, incompleteness to ambiguity during the process of information retrieval and data integration [368]. Metadata ("data about the data") captures aspects of the measurement process that may affect the reliability and future usability of the data, which partially addresses trust in data gathered by sensors [366]. The belief function theory is also utilised to estimate the reliability of the information sources [368]. However, as the efficiency of data processing is centered on the observability of experimental physical asset confined by availability of time-series and the real-time connection with the physical entity, it is hard to create quantifiable measures for various evolving application scenarios. For modelling efficiency of a DT, cost, model maturity and model adaptability summarised as a DT cross-phase metrics can be utilised in the assessment [367], where a high-quality model may cost less in maintenance and reuse. A more mature model gives the expected outcomes and meets application requirements better as time and frequency of using the model increase and a highly-adaptable model recreates the status of the real system better. The cost and model maturity are quantifiable in each application scenario [369, 370], while the adaptability is hard to be quantified but can be enhanced via parameter sensitivity analysis [371] and continuos monitoring of the model's accuracy over time.

Similarity level between modelled dynamics and reality can be evaluated from the perspectives of multiple model's aims, dynamics and the model updates of parameters in response to the real time data integration and feedback from real systems. A good DT is characterised with multiple model's aim fulfilment and well-handled trade-off between model performance and model complexity. The validation and verification of a good DT is dependent on the model-aim directed evaluation of model output for external tasks and the faithful representation of the inner rules of real systems. In case of a CNS modelled using a DT approach, the evaluation involves the comprehensive application of CNS model-aim evaluation methods (section 2.5.2.2). Similarity of the modelled network dynamics and dynamic processes with that of real world systems can be evaluated using DT construction metrics summarised by [367], including quantifiable credibility, fidelity, maturity and qualitative description of complexity and DT standardisation, while the similarity in the context of model updates with the evolving reality can also be evaluated using DT application metrics including failure rate and qualitative description of decoupling ability and parallelizability as well as DT reuse metrics including degree of reconfigurability, reconstructibility and composability. Based on the forementioned methods, the capability of forecasting events that have never happened before and the

synchronisation of nodes existing in the CNS model can be assessed and enhanced. However, there is no unified quantifiable measures across application scenarios, where the assessment and the comparison of DT modelling methods can be further studied.

2.5.2.2 Is it a good CNS?

To answer this question, both measures and standards proposed for CNS and DT can be considered towards the goal of building a good DT with high data processing efficiency and similarity of dynamics with the reality. The assessment of CNS can be divided into two parts: (i) model's aim fulfilment assessment, and (ii) model efficiency assessment (Fig.2.7).

There is considerable literature on the specific model's aims with measurable outputs such as link prediction, community discovery, synchronization, observability and controllability. The evaluation methods for community discovery can be summarised as the internal and external quality evaluation, where more detailed measures can be found in [20]. The assessment of synchronization of networks focuses on the stability of identical states [101], where evaluation of consensus of multi-agent system can also be utilised as they have similar definitions and can be studied from a unified point [347–349]. The observability, with its dual: controllability, can be categorised as structural and dynamical, each representing observability of topology [81], and variables for coupling nodes and node dynamics [82, 83]. For its evaluation, observability matrix based on the dynamic model for a linear (time-invariant) system proposed by [80] has been widely used and extended, while the observability and controllability of nonlinear networks are also studied to investigate the effect of nonlinear dynamical interdependences among variables or the connection with symmetries of networks [84, 85]. Particularly, as an evolving model can be mapped to a link prediction algorithm, performance metrics for link prediction can also assist the quantitative comparison of the accuracies of different evolving models [10]. For example, link prediction can be utilised to validate dynamic social network simulators with graph convolutional neural networks (GCN) [34]. Link prediction performance can be evaluated using precision[10], Area Under the Precision‚ÄìRecall (AUPR) curve [372], Receiver Operating Characteristic (ROC) curves, Area Under the ROC (AUC)[10], Geometric Mean of AUC and PRAUC (GMAUC) [373], Error Rate [299], SumD [298], Kendall's Tau Coefficient (KTC) [280] and Micro/Macro/Weighted Average Precision/Recall/F1 Score [374].

For abstract model's aims such as the mimics of reality using simulation-based networks, we can evaluate these models based on their similarity with the characteristics

of reality that they are required to capture, where DT construction metrics [367], as well as observability and controllability measures for CNS model can be utilised. In addition, when the simulation-based networks are encapsulated in complex networked systems with forementioned specific model's aims, the evaluation measures for specific model's aims can also reflect to what extent the network simulations reach the research goals, especially in case of link prediction where similarity of nodes, edges and their dynamics is the focus.

DT feature assessment that is applicable in CNS is mainly about the efficiency of simulation or modelling, as modelling efficiency is a good quality persued by both CNS and DT. DT application metrics and the DT cross-phase metrics utilised in the DT modelling efficiency are also applicable in the context of a CNS, even though it is not a DT. The data processing efficiency can be assessed and persued via observability and controllability measures for CNS, as well as the measures utilised to ensure the trust in the data in the evaluation of data processing efficiency for a DT.

2.5.2.3 To what extent is it approaching a DT?

The assessment of the distance of a CNS model to a DT is built on the prerequisite that the CNS model is a good CNS model, where a good CNS model has partial DT features and has the potential to approach a DT.

A DT has an appropriate level of complexity that enables it, with good model performance in terms of faithful representation of real systems, to meet the models' aim. As we have discussed in the previous sections, a CNS model can approach a DT with better model performance through appropriately increased complexity. The distance between a CNS model and a DT can also be discussed from these two perspectives: (i) complexity and (ii) model performance (Fig. 2.8).

There are no clear boundaries for the development path of a CNS model towards a DT, either in terms of model performance or complexity. CNS models with an unnecessary level of complexity, under the lower bound of development path, can be identified when there exists a less complex CNS model with equally good or even better model performance. The upper bound of model performance for a CNS model naturally exists under the limit of modelling paradigms. When a CNS model achieves better model performance through increasing complexity, while falling out of the "unlikely" scope and the "unnecessary" scope, it gets closer to a DT. The bounds of development path for DT-CNS model can also be updated with empirical findings in this space.

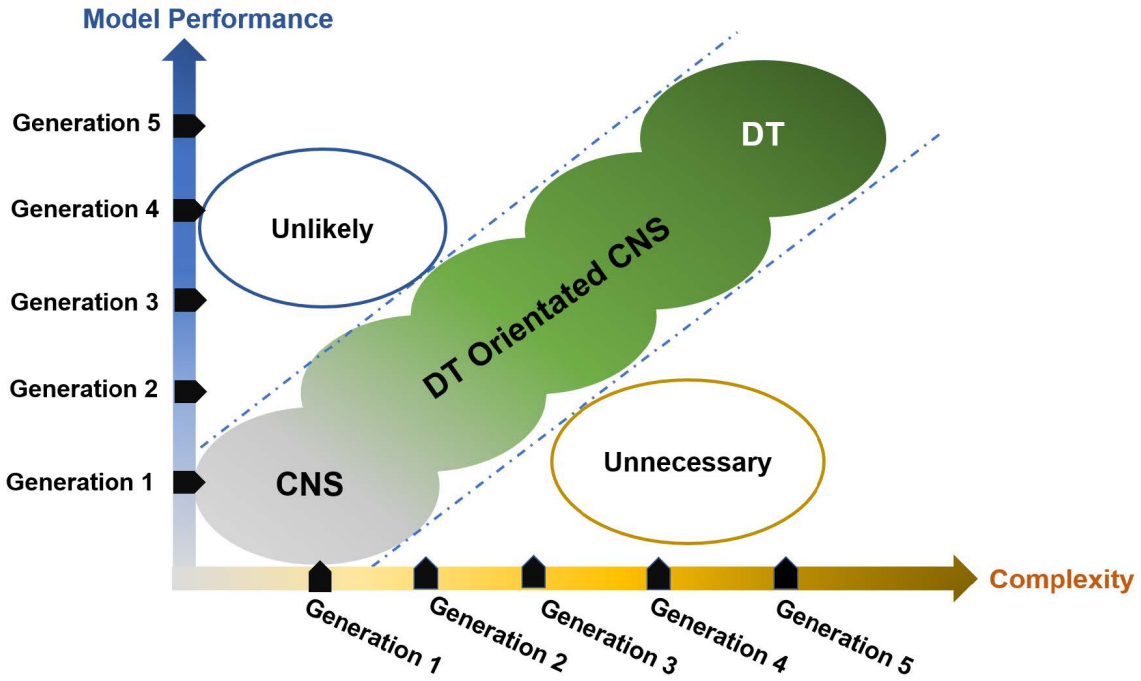


Figure 2.8: The plot of assessing the distance of a CNS model to a DT

Complexity of a CNS model is hard to measure using one, concrete measure, but we are able to rank the complexity of (i) network representation in each complexity dimension (section 2.3.1), and (ii) CNS modelling based on the 5-generation framework (section 2.4). Based on that, a complexity metric can be identified for each generation of CNS models, including their two components: (i) process and (ii) network representation (Fig. 2.9).

For generations of CNS models shown in Fig. 2.9, their components, the process dimension and the network dimension, are each represented with the coloured symbols of G_1, G_{2a}, G_{2b}, G_3, G_4 and G_5. The CNS models in each generation of models vary in dynamics complexity and temporal complexity, while for structural and spatial complexity, they can be built with any complexity level from those two dimensions. The temporal complexity of a CNS model increases as its process, network representation, or both of them, start to change over time in a manner from static (frozen in the time scale), evolving (captured in time windows) to temporal (continuous). The dynamics complexity of a CNS model also increases when the dynamics are changeable, with the process modelled based on the changing parameters or the networks evolving via the changes of inner rules. We can compare the complexity of one component of CNS models in one complexity dimension based on this complexity metric. For example, compared with the

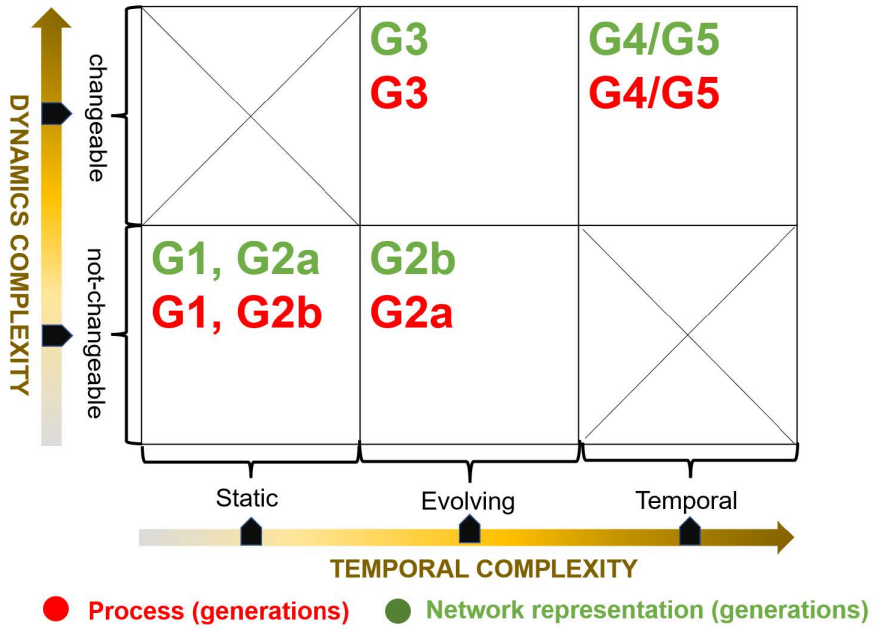


Figure 2.9: The complexity metric of a CNS model

CNS models in G_{2a}, CNS models in G_{2b} are characterised with the temporal complexity that is higher in network representation but lower in process dimension.

Model performance of a CNS model can be assessed based on the two requirements of building a good CNS model: (i) the model's aim fulfilment, and (ii) the model efficiency (section 2.5.2.2 and Fig. 2.7). There are both quantitative measures and qualitative description for a CNS model assessment, but how to combine them for a comprehensive assessment and to deal with the multi-objective optimisation problem for a good CNS model, still requires further study. To gain a rough understanding of the levels of model performance for each generation of CNS models, a model performance metric is built based on (i) the accuracy from the perspective of model's aim fulfilment and (ii) the efficiency (Fig. 2.10).

As is shown in Fig. 2.10, the efficiency is described as ex-post, delayed, real-time and ex-ante based on the CNS models' ways of data processing and modelling. CNS models and their components in the ex-post group have the lowest ranking of efficiency due to the completely post-hoc modelling, like CNS models in G_1. CNS models in the delayed group are characterised with streams of snapshots feeding into the systems across time windows with a time lag behind the real systems. CNS models in G_4 fall in the real-time group as they conduct real-time data processing and modelling. CNS

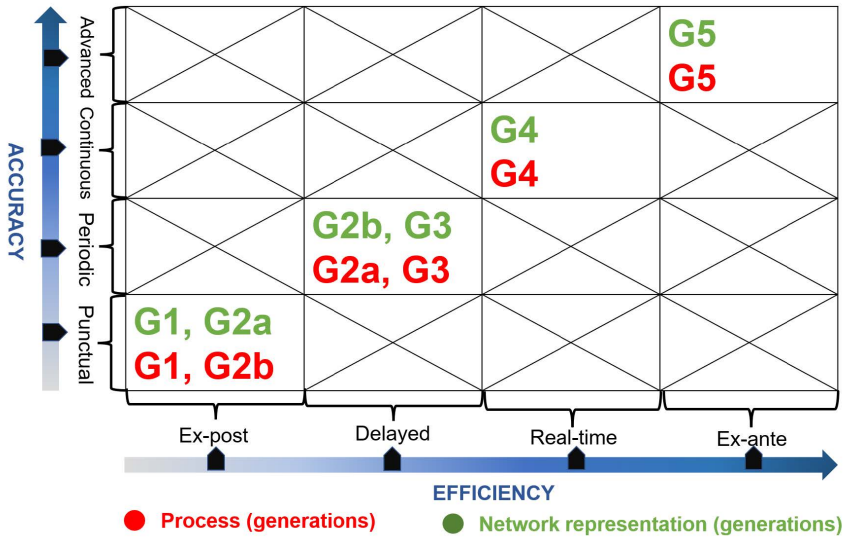


Figure 2.10: The model performance metric of a CNS model

models in G_5, also termed as DTs, are classified in the ex-ante group as they are not only reactive to the observations of real systems in a real-time manner, but also proactive to the things that have never happened before (enabled by the closed feedback loop). The other perspective, i.e. the accuracy, represents a generalised conception of model performance considering how accurately the model' aims are fulfilled. It is classified as punctual, periodic, continuous and advanced. These groups each requires a faithful representation and modelling of the information at only one static time point, within a discrete period, captured continuously or simulated in advance. The required accuracy level increases with the upgraded assessment criterion and the paradigm shift from G_1 to G_5. For example, the evaluation metrics of community discovery, like the modularity and error rate, are widely used for static networks in G_1. They can be further supplemented with a relative reconstruction error rate to analyse the temporal evolution of dynamic networks and communities in G_2 [198].

Current CNS, in terms of data efficiency, tends to rely more on complex simulation-based networks to be able to capture more realistic features or employ data-driven networks introduced with observable temporal and spatial information. However, the data quality is case-dependent in each application scenario and confined by data sparsity, data security, as well as data processing and representation techniques. It is hard to find studies on CNS built and modelled with real time information because of its observability and the difficulties of building realistic real time data simulator. Though there are studies

on CNS model built with big data [169], it is still hard to achieve data efficiency at a "real-time" level. Though in some applications of CNS model in a DT like IoT, where networked information can be gathered in a real time by sensors and integrated into a Knowledge Graph or a block chain, such a method is not applicable in all application scenarios given the requirement of equipment. Therefore, there exists a large gap when it comes to data efficiency between a good CNS model in current studies and a good DT. On the other hand, for model efficiency, some CNS models start to approach a DT with the development of modelling and computation techniques like parallel computation [375], edge computing [376, 377] and cloud computing [378], but it is hard to find such empirical research on CNS model except for CNS model encapsulated in a DT like IoT.

When it comes to the similarity of dynamism of CNS model to that of the reality, there are studies on dynamics of the spatio-temporal networks where both time and space are considered to mimic the reality [127, 195, 196]. While there are some studies that model dynamics over dynamic networks with interrelations between dynamics, including interrelations of dynamic processes over networks [18] and unilateral influence of dynamic process on static networks [379]. However, we can hardly find CNS model with mutual influence between dynamics on the networks and dynamics of the networks or dynamic processes over spatio-temporal networks with interrelations between dynamics across temporal and spatial dimensions. Therefore, as the state-of-the-art CNS models have relatively simple aims and often are developed under some strict assumptions, it is hard to find a CNS model that fulfills DT standards, unless we focus on DT systems with encapsulated networks like IoT [380] and blockchain [359], where the networked systems are built using DT-oriented and directed methods and the network features like topology and interaction of networked node dynamics are ignored.

In terms of model updates, most studies model evolving dynamics in an offline manner without model updates. There are very few attempts to enable model updates e.g. where complex systems are built on networked mobile devices utilising Federated Learning (FL) [381, 382]. FL selects random subsets of devices in an offline manner to collect the local model updates and share the updated global model with the devices. This method is also used to deploy distributed data processing and learning in wireless networks in a blockchain encapsulated in a DT [383, 384]. The divergence of model updates remain a future research gap. In addition, the assessment methods of CNS model are also in need of updates, as they are required to be more dynamic and able to evolve with model changes.

2.5.3 How can we go further?

To answer the question of how to go further to achieve the ultimate goal of building complex networked systems model that faithfully reflects a real system, we build a framework of CNS-based DT, where the modelling can be decomposed into a series of tasks that require modelling methods from both DT and CNS spaces while setting the half-way point as DT-CNS model. We start with listing the research gaps that will guide us in setting goals and tasks for future research.

2.5.3.1 Current research gaps

Based on the five generations modelling framework proposed in section 2.4 and the conducted review of the state-of-the-art, the current research can realise modelling framework of generation 1 and generation 2 and a small number of approaches can reach generation 3. There is no research on generations 4 and 5, where further studies are required to model this very complex scenarios. To build CNS models under generation 4 or 5, and in this way achieving DT-CNS model, seven research gaps need to be tackled:

1. Fulfilment of external tasks (model aim's) while faithfully mimicking the inner rules of the real system;

2. Meaningful feature extraction and model selection that enables network representation to preserve as much information as needed for the model's aim fulfilment;

3. Network simulation via models built with interpretable inner rules which are able at the same time to deal with structural observability and dynamical observability problems;

4. Dynamics of networks that not only focuses on the topology change but also incorporates attribute changes;

5. Modelling dynamically changing dynamic processes in a way that allows for continuous parameter changes under the impact of network changes;

6. Real time data acquisition and processing for CNS modelling;

7. The establishment of a feedback loop that enables continuous updates of CNS model and the changes of real systems referring to the CNS modelling.

2.5.3.2 Set the half-way point: DT-CNS

DT-CNS model emerges with the convergence of DT and CNS modelling approaches, where CNS models approach reality by introducing DT features to the modelling process while a DT incorporates networked information through blockchain, knowledge graph or IoT to assist data processing and modelling. Studies on CNS model generally focus on a single model's aim and simple research objectives with predefined set of assumptions. Also, a vast majority of them use historical rather than streaming and continuous data. DTs can encompass various functions of tools like simulation, optimization and data analytics [143] via real-time processing and updates, with research objects ranging from a single product to the society and relaxed assumptions that allow for a CNS representation and modelling in structural, temporal, spatial and dynamics complexity dimensions. When CNS model approaches DT more assumptions are being relaxed, more model's aims fulfilled and more complex features can be modelled via more efficient data processing and modelling.

There are still several challenges to overcome and trade-offs to be made on the way to DT-CNS model: (i) trade-off between model performance and multiple model's aims, (ii) trade-off between controllability and complexity, (iii) the trade-off between efficiency and accuracy. The existence of multiple model's aims poses a demanding challenge for modelling, which can only be fulfilled by compromising the model performance for a certain model's aim. As complexity of CNS model increases with richer information about network components represented and more complex inner rules modelled, it becomes more difficult to control the CNS model with limited number of features. To achieve higher accuracy, which is a measure of model performance for external tasks, also requires more complex CNS structures and dynamics, diminishing the efficiency of CNS representation and modelling.

A good DT-CNS model refers to the CNS model that simulates or models necessary reality to achieve the predetermined model's aim with a DT-level efficiency, which may not need to be a ready DT but is required to deal with the before mentioned trade-offs. Unified assessment criteria made of mathematical measures for particular model's aims of CNS and the DT evaluation metrics that reflect model efficiency can be used to assess DT-CNS model in a dynamic way, as an assistance of model selection and updates. Modelling paradigms of DT across disciplines can also be introduced to CNS model to go further on the road to a DT. Therefore, the modelling tasks of DT-CNS and the research gaps are mainly about the resolution of trade-offs between performance (output accuracy, input controllability, model efficiency) and complexity, dynamic assessment of evolving

dynamics and the introduction of DT features and paradigms to CNS.

2.5.3.3 Embarking on the modelling tasks

The modelling tasks of DT-CNS should be based on the modelling tasks of CNS and DT that are detailed in section 2.5.1 and the assessment approach from a unified view point of CNS model and DT in section 2.5.2. The network representation involves tasks of data processing, data analysis and variable selection, while the training of a node and relationships dynamics and dynamics over networks can be summarised as data modelling process, where trust in data, model and updating procedure should be considered with the requirements of observability, similarity and synchronization. Therefore, the modelling tasks of DT-CNS can be generally categorised and presented as:

1. Data processing that cope with imperfection of data;
2. Data analysis and feature selection that considers observability and controllability;
3. Network representation based on the selected variables and the similarity measures;
4. Modelling of real-time self-evolving dynamics;
5. Model updates enabled by reconfiguration and reconstruction;
6. Model evaluation over the entire process.

For the task (1) on data processing and data management, uncertainty analysis has been utilised to deal with data imperfection, while data fusion emerges as a prevalent way to capture reliable, valuable and accurate information. Also knowledge graphs as well as blockchain have been popular choices for data integration and information retrieval. However, it is challenging to ensure the efficiency of data processing and data management, especially given the "real-time" feature of a DT and the requirements of data quality, where further research is needed. There is also an issue of data sparsity which requires further research on simulation of CNS to deal with the unobservability and unavailability of data.

For the task (2), data analysis and variable selection that considers observability and controllability, has been studied in the context of CNS over the years. However, more effort is needed given the high demand for adaptability of CNS model as they approach reality. Specifically, when it comes to simulation-based networks, how to choose

the changeable variables that drive the evolution of networks while preserving the characteristics of real world situation is an interesting research gap given the problem of data scarcity resulting from the rules of keeping data security.

For the task (3), network representation based on the selected variables and the similarity measures, is thoroughly studied research area, while CNS model built with DT approaches with the emphasis on network properties has not been deeply studied. Spatio-temporal network, together with the interrelation and interconnection of dynamics within or over such networks are very interesting topics that can be further studied.

The task (4) on modelling of real-time self-evolving dynamics enabled by continuous machine learning is the core element of building a DT, where the interrelations between dynamics in CNS model remain an unexplored area especially when it comes to the mutual influence of dynamics on and of networks. More specifically, the evolving dynamics on and of a network is the most relevant to the case-dependent autonomy of DT system, which can be autonomous, not autonomous, or partly autonomous, where the research on the interventions in networked systems can be further introduced with context-awareness and autonomy.

For the task (5), model updates enabled by the reconfiguration and reconstruction, is closely related to the task (4), where the construction of feedback loop is crucial for the continuous modelling process. There is research on a DT built from the perspective of state transition in a discrete way, where how to narrow the time gap between states and extend the state transition to the continuous modelling is a research gap.

For the task (6), model evaluation over the entire modelling process of DT-CNS model needs a unified assessment framework, where concrete measures that consider features of both DT and CNS should be explored. There is already literature on measures of network analysis and DT analysis, though the principled combination of these two or proposing new integrated quality measures remain an outstanding research gap.

The above tasks show the complexity of the research that is needed to be accomplished along the way of working towards DT-Oriented modelling of Complex Networked Systems.

2.6 Conclusions

This chapter focuses on the modelling approaches of Complex Networked Systems that pave the path for its convergence to the ultimate goal: a Digital Twin of a CNS.

We review and discuss the CNS model from three perspectives: (i) model's aims that have been studied for CNS model (section 2.2), (ii) modelling paradigms that enable to represent a networked system in a way that preserves as much information as needed (section 2.3), (iii) modelling approaches for dynamics of networks and dynamics over networks that enable to meet model's aims (section 2.4). Those themes are discussed through the lenses of four, proposed by us, complexity dimensions of complex systems: (i) structural, (ii) temporal, (iii) dynamics and (iv) spatial. Discussion that considers those complexity dimensions enables to better understand current modelling challenges and quantify how far we are from achieving Digital Twin modelling capabilities when representing networked systems.

The model's aims for CNS model distinguish when they focus either on the external task to be performed on the system or modelling of the inner rules of real systems, and this division and specialisation can be eliminated when a Digital Twin is considered since it is able to undertake multiple external tasks by faithfully covering and reflecting the complexities of real systems. Models of CNS proposed over the years have been found to approach real systems with increasing complexity in structural, temporal, spatial and dynamics dimensions. To generate and preserve this heterogeneous networked information, modelling paradigms of network representation get more complex with compromised interpretability. These models either focus on inner rules of network generation at a local level or aim at a compressed network representation at a global level, but all converge to the goal for a faithful representation of real systems.

Dynamics of networks, dynamic processes on networks as well as their interrelations are three elementary sources of complexity for dynamics in the CNS model. To navigate a pathway through different levels of complexity of modelling CNS, we devise a modelling framework of CNS that considers all these three elements and consists of five generations reflecting the progress of work that has been done in this field. Each generation builds upon the previous one meaning that the next generation encompasses higher complexity levels than the previous one.

This modelling framework is agnostic to the model's aim so any of the discussed aims can be attempted using models built within each of the generations. Though, one needs to remember that models from different generations will enable to achieve selected aim to a different extent. The proposed framework also shows how Complex Networked Systems' models approach a Digital Twin with more complexity through generations, (i) generation 1: dynamic process on static networks, (ii) generation 2 with two variations: dynamic process on evolving networks and evolving dynamic process on static networks,

(iii) generation 3: evolving dynamic processes on evolving networks with interrelations between them, (iv) generation 4: temporal dynamic processes on temporal networks with interrelations between them and the acquisition of real time information and finally (v) generation 5 that further introduces modelling framework of generation 4 with an information feedback from CNS's model to the real system. From generation 1 to generation 5, the real system can be represented more faithfully with richer information captured and finally a CNS-based DT can be created in generation 5. Current studies have made good progress under modelling framework of generation 1 and generation 2. Only small number of approaches reach generation 3. For generation 4 and 5, there is no research in this space and further studies are required to model this very complex scenarios to achieve better performance in the context of any of the presented model's aims.

To be more aware of how to approach a Digital Twin with CNS model, we propose an assessment framework (section 2.5) that aims at quantifying the distance of CNS model to DT from the perspective of CNS model's aim fulfilment and the perspective of a DT's faithful representation of reality. A half-way point referred to as a DT-CNS model is proposed to bridge the gap between the current approaches to modelling of CNS and the ultimate goal of a DT (generation 5 models) for future study.

The goal of the future research in the space of complex networked systems and network science more broadly is to develop a DT-CNS model that will enable to address research gaps presented in section 2.5.3. Integrating dynamic networks with dynamic processes and allowing for mutual influence between them (allowing at the same time for continuous adaptation of the system using streaming data as an input) will make it possible to create the DT-CNS models. This will be a major breakthrough in the space of modelling CNS.

Overall, this chapter answers **RQ1** by a comprehensive survey on the modelling approaches of Complex Networked Systems as they approach the ultimate goal: a Digital Twin of a CNS. The research contributions, regarding each research objective, can be summarised as follows:

- **RQ1**: How to compare modelling paradigms of CNS and measure their distances to a DT?

    1. **RO1.1**: Conduct a comprehensive critical review and comparative analysis of CNS model.

RC1.1: A comprehensive and comparative review of existing CNS models modelling paradigms.

2. **RO1.2**: Propose a conceptual modelling framework for DT-CNS.

   RC1.2: New conceptual modelling framework for DT-CNSs.

3. **RO1.3**: Propose an assessment approach to conceptually compare and evaluate CNS models.

   RC1.3: Novel assessment framework for evaluation of models developed within DT-CNS modelling framework.

Moreover, the proposed generations of modelling framework and four complexity dimensions in this chapter also serve as the theoretical basis for the DT-CNS modelling in the studies presented in the following chapters. Based on this conceptual framework, we launch our research journey towards a DT by paving a development path from a SNS to a generation 1 DT-SNS based on heterogeneous node features in chapter 3. In chapter 4, we propose an extendable modelling framework for the generation 1 DT-CNS models in social network space based on heterogeneous node features and rules of interactions and epidemic transmission. This is followed by chapter 5, where the generation 1 DT-CNS models for social networked systems are extended with feature representation principles for more expressive power of node features and more accurate representation of real systems. In chapter 6, the DT-CNSs modelling framework for social networked systems progresses to generation 2 and generation 3 by introducing changes in interaction rules and their interrelations with the epidemic outbreak. Overall, the generations of modelling framework work as a road map for developing DT-CNS models in this research project. In this process, DT-CNS models get increasingly complex in the structural and dynamics dimensions by introducing realistic information related to nodes' features, interaction rules and representation principles.

CHAPTER 3

REVIEW AND ASSESSMENT OF SOCIAL NETWORK SIMULATORS

3.1 Introduction

Social network simulators (SNSs) aim at faithfully representing real networked systems and attempt to model the inner rules of network dynamics. They generate simulation-based networks (built with simulated data) or hybrid networks (built with both real and simulated data) to deal with the unobservability problems resulting from data sparsity, privacy concerns and lack of ground-truth [159, 385].

As reviewed in chapter 2, the complexity of the social network simulations results from the heterogeneity of the network components (nodes, edges and their attributes) as well as dynamically changing behaviour of those elements. Current studies have made good progress to attain a desired complexity level. They start from relatively simple simulations which are based on predetermined network statistics or connection principles about network topology, including such classical examples as Barabasi-Albert model for the scale-free network [45] or preferential attachment of nodes based on similarity, popularity or both of them [46]. Complexity of these networks increases in the structural dimension when node attributes (e.g. age, gender, geo-space, etc.) or edge attributes (e.g. direction, weight, relationship, etc.) are introduced. To incorporate the impact of structural complexity on network simulation, social network simulators built on interaction rules have been employed due to their flexible definitions based on

topology, attributes or both of them, like homophily [47, 56], triadic structure [47, 386] and geographic proximity [64].

The social network simulations get even more complex when they enable the network structure to evolve using inner rules that define how components can change over time. Studies on SNSs focus mainly on the topology change (e.g. the instantaneous social contacts [33, 35]). Only a small number of studies account for both attribute change and topology change. For example, the SNS proposed by [34] generates social networks with node attributes including features (e.g. age, gender, etc.) and preference for each feature when interacting with others, termed as social DNA (sDNA). The network topology is formed based on sDNA and evolves as sDNA changes randomly over time. However, as an example of one its existing drawbacks, the mechanisms for attributes and preferences changes need to be further studied to allow more realistic modelling.

Studies on SNSs typically focus on the social intervention analysis and conduct an optimisation of inner rules for achieving the highest possible similarity with the target network [24, 47]. The consistent network growth enabled by the iterative application of SNSs unfolds the desired network complexity at the cost of time and energy [387]. The time step in these studies is often assumed to have the same meaning with the iteration, where the networks grow linearly with time (e.g. predetermined number of added nodes or rewired edges within one iteration between time step t and $t+1$) [24, 47]. However, the time truly spent on each iteration of SNS varies with the complexity of networks and computation resources [34] and the time that measures the duration of each SNS iteration can serve as an indicator of model efficiency. Current studies on SNSs take the similarity between simulated networks and the target networks as the objective of running an SNS [24, 34, 47], while none of them considers model efficiency.

Therefore, we propose **RQ2** and, in this chapter, answer this question by a topical review on the current state-of-the-art SNSs and the extension of SNS towards DT-SNSs considering heterogeneous node features (**RO2.1**). We aim to conceptually evaluate the current existing SNSs and the proposed DT-SNS (**RO2.2**) under the proposed evaluation protocol (chapter 2.5 and and in the meantime, provide a possible pathway of extending SNSs towards a Digital Twin Oriented SNS under the generations of modelling framework (chapter 2.9).

- **RQ2**: How to build and evaluate DT-SNS that accommodates heterogeneous features of nodes?

    1. **RO2.1**: Develop a DT-Oriented Social Network Simulator that enables social

network simulation based on heterogeneous node features.

2. **RO2.2**: Evaluate proposed DT-SNS.

In this chapter, we review and assess the current state-of-the-art SNSs under (i) the Digital Twin Oriented Modelling framework encompassing different complexity dimensions and (ii) the assessment framework based on similarity and efficiency, as proposed in chapter 2. To provide a possible pathway of extending SNSs towards a Digital Twin Oriented SNS, we build an inner rule-based SNS that enables the social network simulation to take into account structural and dynamic dimensions of complexity that builds on [34]. This extension brings us one step closer to a Digital Twin Oriented SNS (DT-SNS) by introducing heterogeneous features of nodes. We also propose an approach to assess the similarity between real and simulated networks with the composite performance indexes based on both global and local structural measures, while taking runtime of the simulator as an indicator of its efficiency. We use Karate Club network as an example to primarily illustrate but also examine to what extent the proposed SNS can model the structural complexity of a real social network. The diverse network patterns simulated by SNSs and their respective similarity levels to the real network assessed by the composite performance index, given different structural complexity levels, reveal the challenges of SNS performance evaluation for the future DT-SNSs.

Overall, the main contributions of this chapter are:

- reviewing current state-of-the-art SNSs under the five genertaions of DT-CNS modelling framework (**RC2.1**);
- proposing a DT-SNS based on heterogeneous node features to faithfully represent static network topology (**RC2.1**);
- a case study on the modelling and assessment of DT-SNSs with different levels of structural complexity (**RC2.2**).

The rest of this chapter is structured as follows. Section 3.2 reviews the current state-of-the-art of social network simulations and offers comparison between different SNSs in the context of modelling framework proposed in [1]. Section 3.3 presents the methodology of building an SNS to generate desired networks. Following this, section 3.4 introduces the real data employed in the analysis and presents the results of the experiments. Finally, conclusions and future work are given in section 3.5.

3.2 Current state-of-the-art of Social Network Simulators

We review here and discuss the current developments in the context of SNSs from the following four perspectives: (i) modelling prerequisite – observability – that determines the observable information to be modelled and the unobservable information to be simulated; (ii) modelling generations connected with the complexity of modelling SNSs; (iii) complexity dimensions that describe the complexity of social network simulations; and (iv) assessment of SNSs concerned with the distance between the social network simulations and the target networks.

3.2.1 Modelling prerequiste

Modelling prerequisite – observability, in the context of SNSs, is concerned with the ability of reconstructing the desired complexity of networks from a limited set of observed network components in finite time with an understanding of their dynamically changing behaviour [1]. The network components include nodes, edges and attributes of nodes and edges. To generate networks with desired topology and attributes, the existing SNSs simulate the unobservable (represented with a ○) and partially observable network components (represented with a ◑) based on the observable network components (represented with a ●) and an inner rule that directs the network growth (Tab. 3.1).

As is shown in Tab. 3.1, the simulated networks are built with purely simulated topology and attributes, dealing with the unobservability problem of all network components for the analysis of the topological features [23, 45, 48, 61, 192, 390–398, 400–403] and mimicking of real networked systems [24, 49, 199, 388, 399].

The hybrid networks are built with the partially observable and partially simulated network components.The observable network components refer to the available information related to nodes, edges and attributes, which contributes to generating realistic networks. When only partial information is observable, we simulate the unobservable network components needed for a faithful network representation. In this space, such networks, called hybrid networks, are composed of partially observable and partially simulated information related to nodes, edges and attributes. Some SNSs incorporate the observable nodes and node attributes while simulating the unobservable edges for a desired states of networks [34, 47, 64, 65, 194, 202, 289]. Some SNSs consider the partial observable edges and conduct link prediction for the unobservable edges [9, 46, 62, 181],

Table 3.1: Current SNSs and the observability of network components in their simulations

Type	Topology		Attributes		Existing SNS
	Nodes	Edges	Nodes	Edges	
Simulated	○	○	○	○	[48], [61], [23], [388], [192], [63], [389], [342], [390], [391], [392], [393], [394], [395], [396], [397], [398], [399], [400], [401], [402], [403]
Hybrid	●	○	●		[47], [64], [34], [289], [202], [194], [66], [97], [65]
	●	○	●	○	[404], [164], [278], [405]
	●	◑			[46], [62], [181]
	●	◑	●		[9], [297]
	●	◑		◑	[302], [287]

while some SNSs incorporate edge attributes in this process [287, 297, 302]. In general, there is limited research that, in the simulation process, takes into account real-world attributes of both nodes and edges. There are some works that account for limited number of real node attributes [9, 34, 47, 64–66, 97, 164, 194, 202, 278, 289, 297, 404, 405] and when it comes to edges only two simulators partially consider real-world edges [287, 302]. As a result, most of the hybrid simulators take into account some of the topology information but not many consider attribute information and this points to the need for further work on SNSs, especially in the context of learning from continuously streaming data. Such advancement would bring current SNSs closer to the Digital Twin paradigm.

3.2.2 Modelling generations

Modelling generations, as proposed in chapter 2, consider three key elements of modelling complex network dynamics, the simulation of (i) networks, (ii) processes over networks, and (iii) the interrelation between the two. There are five generations of modelling paradigms (generation 1, generation 2, generation 3, generation 4 and generation 5), where SNSs can be built with increasing dynamics complexity through generations and finally reach the goal of a DT in generation 5.

We review and discuss the above mentioned three elements in the context of the

modelling generations while considering two types of heterogeneity (i) the existence of a given element (represented with a △) and in addition to its existence, (ii) the capability of an element to change over time (represented with a ▲) (Tab. 3.2).

Table 3.2: Current SNSs and their generations

Stage	Network	Process	Interrelation	Existing SNS
Generation 1	△			[45], [46], [47], [24], [48], [61], [23], [388], [192][342], [390], [391], [392], [404], [394], [395], [396], [397], [398], [399], [400], [401], [402], [403], [289], [202], [194], [97], [164], [278], [287]
	△	△		[49], [50], [51], [52], [31], [32]
Generation 2a	△	▲		[406], [407]
Generation 2b	▲			[64], [34], [63], [389], [66], [65], [405], [62], [181], [9], [297], [302], [408], [250], [112]
	▲	△		[67], [68]
Generation 3	▲	△	△	[330]
	△	▲	△	[335], [334], [336]

Generation 1 of models focuses on dynamic process on static networks (Tab. 3.2 and Fig. 2.5(a)). They simulate networks that are "frozen" in time, with a dynamic process taking place on the networks where parameters of this process do not change during the simulation (e.g. epidemic spreading process on static social networks with a fixed infection rate [14, 32]). Most studies focus on SNSs in generation 1 [18, 23, 24, 31, 32, 45–52, 61, 97, 164, 192, 194, 202, 278, 287, 289, 342, 388, 390–404]. Also, many of the simulators generate static social networks without any consideration of dynamic processes [23, 24, 45–48, 61, 97, 164, 192, 194, 202, 278, 287, 289, 342, 388, 399, 404].

SNSs in generation 2 have two variations, one termed as generation 2a and the other as generation 2b. SNSs in generation 2a focus on evolving dynamic process on static networks (Tab. 3.2 and Fig. 2.5(b)). They simulate a static network where dynamic process changes its parameters over time and gets captured in snapshots (e.g. epidemic spreading processes on static social networks with a changeable infection rate [406]). There are only few studies on SNSs in generation 2a [275, 406, 407].

SNSs in generation 2b focus on dynamic process on evolving networks (Tab. 3.2 and Fig. 2.5(c)). They simulate network snapshots that describe the network topology changes over time, where the dynamic process takes place without changing its parameters (e.g. epidemic spreading process with a non-changeable infection rate on social network that evolves over time [68]). There are many studies on SNSs that model only the dynamics of network structure in this space [9, 34, 62–68, 112, 181, 194, 203, 250, 297, 302, 389, 405, 408] while few of them additionally consider the dynamic processes [67, 68].

SNSs in generation 3 focus on evolving dynamic processes on evolving networks with interrelations between them (Fig. 2.5(d)). They simulate networks, dynamic processes and their changes using snapshot approach (data are processed in batches), while incorporating the interactions between a network and a process. For example, the epidemic spreading across the network can leave some nodes dead and get them removed [330], while the infection rate of epidemic spreading process can also vary depending on nodes' groups (structural patterns) [334, 335]. Only few SNSs consider the interrelations between the networks and the dynamic processes, where they just study the dynamic processes with non-changeable parameters on evolving networks [330] or evolving dynamic processes on static networks [334–336].

SNSs in generation 4 focus on temporal dynamic processes on temporal networks with interrelations between them and the continuous acquisition of real time information (Fig. 2.5(e). They simulate networks and the dynamic processes with instantaneous changes of network topology and parameters, while incorporating the interactions between a process and a network. SNSs in generation 5 further extend the SNSs in generation 4 by closing the feedback loop between the SNSs and the real system and can be identified as an idealised state that can be named as a Digital Twin. In generation 5, the temporal dynamic processes occur on temporal networks with interrelations and the real-time acquisition and feedback of information between the SNS and reality (Fig. 2.6). More specifically, compared with generation 4, generation 5 is equipped with another real-time information flow from the reality to the model. This enables the two-way real-time information feedback between the SNS and reality. Currently, there are no studies on SNS in generation 4 and generation 5 and further studies are required to model such high complexity scenarios to approach a DT.

3.2.3 Complexity dimensions

Complexity dimensions describe the complexity of social networks resulting from the heterogeneity of their network components (topology and attributes) and temporal

dynamics of those components and their attributes. As proposed in [1], there are four complexity dimensions (i.e. structural, spatial, temporal and dynamics) that need to be considered when representing and modelling a network. The structural and the spatial complexities are concerned with the existence of network attributes and the space where the topology can be embedded respectively. The temporal and the dynamics complexities are connected with the changeable network components and their dynamics, respectively.

We review and discuss the network components and network dynamics of the social networks generated by the existing SNSs while considering two types of heterogeneity: (i) the existence of the components in a static time scale (represented with a △) and, in addition to its existence, (ii) the capability of the component to change over time (represented with a ▲) (Tab. 3.3).

Table 3.3: Current SNSs and the components of their simulations

Stage	Topology		Attributes		Network dynamics	Existing SNS
	Nodes	Edges	Nodes'	Edges'		
Generation 1 and 2a	△	△			△	[45], [48], [61], [23], [388], [24], [199], [97], [399], [192], [49], [50], [46][51], [52], [398], [406], [32], [31], [275], [393], [390], [401], [391], [392], [394], [400], [395], [402], [403], [396], [397]
	△	△	△		△	[289], [202], [47], [194]
	△	△		△	△	[287], [342]
	△	△	△	△	△	[404], [164], [278], [409]
Generation 2b and 3	△	▲			△	[62], [63], [410], [67]
	△	▲	△		△	[34], [64], [65], [66], [9], [297]
	△	▲	▲		△	[34], [408]
	△	▲		▲	△	[302]
	△	▲	△	▲	△	[250], [68]
	△	▲	▲	▲	△	[405]
	△	▲			▲	[181], [112], [389]

As is shown in Tab. 3.3, most SNSs focus on static networks with no attributes, which are categorised as generation 1/generation 2a models and characterised with the lowest level of complexity in each dimension. They generate networks based on predetermined network statistics and connection principles about topology [23, 24, 45, 46, 48–52, 61, 97, 192, 199, 388, 390–403]. Some other SNSs, with higher level of structural complexity, incorporate node attributes [47, 194, 202, 289], edge attributes [287, 342]

or both of them [164, 278, 404] into the generation process of static networks. The SNSs in generation 2b/generation 3 have higher level of temporal complexity as network components are allowed to change over time, and they generally just consider the topology change [9, 34, 62–66, 297]. Few SNSs focus on changeable network dynamics, where the existing ones just consider topology and its changes over time, with a higher level of dynamics complexity but the lowest level of structural complexity [112, 181, 389]. To the best of our knowledge, there are no other SNSs in generation 4/generation 5 that would enable to model the desired level of complexity of network structure and its dynamics.

3.2.4 Assessment of SNSs

Assessment of SNSs is concerned with investigating how close the generated networks are to the real systems that they attempt to model.

For SNSs, which generate simulated or hybrid networks to deal with the unobservable information, their assessment involves network statistics like density, degree distribution, shortest path, assortativity, modularity, clustering coefficient and betweenness [274]. Most SNSs employ network statistics concerned with the global network structures like degree centrality, betweenness centrality, and PageRank coefficient [24] to measure the similarity between the social network simulations and the target network. As an example, SNS proposed by [24] simulates the unobservable edges to achieve the desired betweenness centrality, PageRank, local clustering coefficient and degree distribution of the networks. There are also SNSs that simulate the missing edges for networks with partially observable edges, which can be treated as a link prediction task and evaluated considering the prediction performance with precision [9, 10], Micro Precision [268], Area Under the ROC (AUC) [10], Error Rate [299], etc. However, the above mentioned measures just focus on the states of static networks, whereas further studies are required for the measures considering network changes over time.

As an indicator of the SNSs' efficiency, the runtime of the simulation is considered by few studies [34, 408], which is proved to be influenced by the network components such as the number of simulated features and the network size.

To summarise, SNSs in the current studies can not capture enough complexity from the real world for an idealised DT modelling due to observability reasons. Though some SNSs consider real information, such as node attributes, there remains an undeveloped research space concerning how to gradually craft the increase of complexity for an improved SNS performance in one or several aspects. By increasing structural complexity, we aim to provide one of the possible pathways for extending the existing SNSs towards

a DT-SNS discussed in section 3.3. Through experiments in section 3.4, we also reveal the challenges of DT-oriented modelling by presenting the complex network patterns and, thus, the difficulties they pose to the SNS performance evaluation and SNS extension towards a DT.

3.3 Towards Digital Twin Oriented Social Network Simulator

After reviewing and evaluating current state-of-the-art SNSs in section 3.2, we select one of the most promising SNSs from the evaluated ones to facilitate the generations of social networks at different complexity levels based on real and simulated data. We build an inner rule-based SNS referring to and extending work in [34]. The inner rule-based SNS generate social networks based on the interaction rules that characterise nodes' preferences for connecting with others. We decide to extend this SNS various complexity levels through changes in interaction rules. We also provide an example of developing an extensible SNS towards a Digital Twin-Oriented SNS for our future study. The two steps of building such an SNS that mimics the target network include: (i) the proposal of an inner rule-based SNS with extensible complexity in section 3.3.1 and (ii) an optimised iterative application of this SNS for similarity and efficiency in section 3.3.2 considering the observability and complexity. We have included some SNSs in the Appendix A.1, while employing the extensible SNS initially proposed by [34] as an example in the following part of this chapter.

3.3.1 An Extensible Social Network Simulator

The inner rule-based SNS framework proposed by [34] models people as nodes, with the node attributes including features (i.e. such as age, gender, etc.) and individual preferences towards particular features termed as social-DNA (sDNA). The edges between these nodes represent relationships between people, which are connected according to preferential attachment rule based on topological and non-topological (attributes) characteristics. The edges can be directed or undirected and they are time-stamped with the iterative application of the SNS.

Based on this framework, we treat each node attribute as a tuple defined with a feature and sDNA (preference for this feature and weight of preference), while allowing their variations among individuals and changes over time. This enables the social

network simulations with an extensible complexity in structural, spatial, temporal and dynamics dimensions. To implement this complexity, the detailed settings about network components and network dynamics are proposed and presented below.

3.3.1.1 Network components

Network components include: (i) nodes (e.g. people), (ii) edges (e.g. social contact or relations), (iii) attributes including node attributes (e.g. age, gender, geo-space, etc.) and edge attributes (e.g. direction, weight, relationship, etc.). They vary in diversity and their capability to change over time.

We assume that the SNS simulates an undirected network $\hat{G}$ with N attributed nodes within M iterations, the network formed in the k_{th} iteration is represented as

$$G_k = (V, E_k, A_k) \quad k \in [1, M] \tag{3.1}$$

which is composed of a fixed node set $V = \{v_1, \cdots, v_N\}$, an edge set that grows through iterations: $E_k = \{(v_i, v_j)_k | , v_i, v_j \in V, i \neq j\}$ and a node attribute set $A_k = \{a_k(v_1), \cdots, a_k(v_N)\}$ that can vary with nodes and change for each iteration.

Especially, referring to [34], the set of attributes $a_k(v_i)$ for node v_i is defined as

$$a_k(v_i) = \{f_k(v_i), p_k(v_i), w_k(v_i)\} \quad v_i \in V \tag{3.2}$$

which includes a feature vector $f_k(v_i)$ and the social-DNA (sDNA) defined with two vectors of the same length as $f(v_i)$. Any attribute, in this context, can be represented with a three-value tuple that is composed of feature, preference for this feature and weight of preference.

$p_k(v_i)$ determines whether to prefer similar feature with a binary value of 1 (prefer) or -1 (not prefer), and $w_k(v_i)$ represents the weight of preference for the feature with the range of $[0, 1]$.

The sDNA can be set at the level of individual nodes, groups or the whole populations in a static manner or can mutate over time. This variability of node attributes forms another source of structural complexity and requires further study.

3.3.1.2 Network dynamics

Network dynamics drives the network growth based on the current network topology and attributes. Referring to the previous study by [34], the edges of the undirected network

are created without removal according to the ranking scores $s_k(v_i, v_j)$ of each node pair based on the two-way evaluation between node v_i and v_j:

$$s_k(v_i, v_j) = q_k \Phi_k(v_i, v_j) + r_k \Delta_k(v_i, v_j) + c_k^{\tau} \Pi_k(v_i, v_j) \quad v_i, v_j \in V, i \neq j \tag{3.3}$$

where $\Phi_k(v_i, v_j)$, $\Delta_k(v_i, v_j)$ and $\Pi_k(v_i, v_j)$ each represents the feature-based, popularity-based and shortest-path score with the weights of q_k, r_k and c_k respectively. τ denotes the transposition of vector c_k in the equation.

The feature-based score is based on the nodes' preferences $sDNA(v_i)$ and $sDNA(v_j)$ for their features $f(v_i)$ and $f(v_j)$. $\odot$ is the Hadamard product, a binary operation that multiplies the corresponding elements of two matrices with the same dimension.

$$\Phi_k(v_i, v_j) = |f(v_i) - f(v_j)|^{\tau}(w_k(v_i) \odot l_k(v_i)) + |f(v_i) - f(v_j)|^{\tau}(w_k(v_j) \odot l_k(v_j)) \tag{3.4}$$

The popularity-based score incorporates the preference for nodes with higher degrees, with m_k representing the preferential attachment parameter and d_k represents the calculation of node degrees in the k_{th} iteration:

$$\Delta_k(v_i, v_j) = m_k d_k(v_i) + m_k d_k(v_j) \tag{3.5}$$

The shortest path-based score is calculated based on the preference for the shortest path-length l within the range $[1, q]$ between the node pair:

$$\Pi_k(v_i, v_j) = \sum_{l=1}^{q} \gamma(v_i)^l I^l(v_i, v_j) + \sum_{l=1}^{q} \gamma(v_j)^l I^l(v_i, v_j) \quad l \in [1, q] \tag{3.6}$$

where $\gamma(v_i)^l$ represents node v_i's weight of preference for the path-length l and $I^l(v_i, v_j)$ represents the existence of the path-length l between node pair v_i and v_j with a binary value of 0 or 1.

$$s_k(v_i, v_j) = q \Phi_k(v_i, v_j) + r \Delta_k(v_i, v_j) + c \Pi_k(v_i, v_j) \tag{3.7}$$

where $\Phi_k(v_i, v_j)$, $\Delta_k(v_i, v_j)$ and $\Pi_k(v_i, v_j)$ each represents the feature-based, popularity-based and shortest-path score with the weights of q, r and c.

3.3.2 Optimisation towards a Digital Twin

A high quality SNS, in order to approach an idealised Digital Twin status, needs the optimised simulation of network components and network dynamics considering the three elements reviewed in section 3.2: (i) observability, (ii) complexity, and (iii) assessment of social network simulators.

3.3.2.1 Development path

Development path of the SNS towards a DT can be divided into small steps, where with each step the complexity of a simulated social network gradually increases. Under the assumption of a fixed node set and non-attributed edges, we define a development of SNS towards a DT with increasing structural complexity and temporal complexity (Tab. 3.4).

Table 3.4: Small steps of building up the complexity.

Stage	Topology		Attributes		Dynamics
	Nodes	Edges	Nodes'	Edges'	
Generation 1	△	△			△
	△	△	△		△
Generation 2	△	▲	△		△
	△	▲	▲		△
Generation 3	△	▲	▲		▲

In generation 1, the structural complexity increases when more features are considered with a longer feature vector $f_k(v_i)$ as part of a node attribute $a_k(v_i)$. The sDNA, p_k and w_k, represent the corresponding preference and weight of preference for respective similar features in the feature vector, composing another parts of node attribute $a_k(v_i)$. The sDNA also increases the structural complexity when more variability of the nodes' preferences (ranging from population-based, group-based to individual-based) is incorporated. Feature selection and the sDNA simulation is the necessary step required for an appropriate structural complexity level.

In G2, the temporal complexity increases when edges E_k and the node attributes A_k start to change with every simulation step, creating snapshots of evolving networks (with overall number of iterations k). This requires further study on what changes, how it changes, which network state the change leads to. In generation 3, the network dynamics can change with the changeable weighting factor q_k, r_k and c_k for the ranking scores of node pairs, as well as the changeable mechanism that generates sDNA through various stochastic distributions.

We can simulate the unobservable networked information and remove the unnecessary steps if the increased complexity depreciates the performance of an SNS measured using criteria of similarity and efficiency. For each step ahead, we optimise the unobservable sDNA, p_k and w_k and the other parameters with the SNS proposed in section 3.3.1 to

achieve the maximised similarity (between the G_k and the desired $\hat{G}$) and the minimised runtime for its simulation with efficiency (equation 3.8).

$$
\begin{aligned}
&\min||G_k - \hat{G}|| \\
&\min \sum_{z=1}^{k} t_z \\
&s.t. p_k \in \{-1, 1\} \\
&\quad q_k, r_k, c_k \in (0, 1] \\
&\quad w_k, m_k, \gamma_k^l \in (0, 1]
\end{aligned}
\tag{3.8}
$$

The efficiency of SNS is measured with the runtime $\sum_{z=1}^{k} t_z$ spent on the k iterations of running the SNS. t_z, denoted with the index z, represents the z_{th} iteration of social network simulation. We propose the composite performance indexes to measure the similarity of social network simulations and the target network.

3.3.2.2 Composite performance indexes

Composite performance indexes measure the distance $||G_k - \hat{G}||$ between the simulated social network G_k and the target network $\hat{G}$ based on their similarity from both the global and the local perspectives [274]. As shown in equation 3.9, the composite performance index is calculated as the average distance between the simulated and the real networks in terms of each network measure η_i.

$$
||G_k - \hat{G}|| = \frac{1}{n} \sum_{i=1}^{n} ||\eta_i(G_k) - \eta_i(\hat{G})|| \tag{3.9}
$$

The network measures (node degree distribution, shortest path length distribution, modularity, assortativity, clustering coefficient distribution and the triad significance profile), each denoted by η_i, evaluate the networks from different perspectives concerning the topological characteristics. The global perspective, concerned with the high level outcomes of interactions between the number of actors, includes measures such as node degree distribution, shortest path length distribution and the values of modularity and assortativity [274]. The local perspective, connected with interpersonal relations within subgraphs, includes measures such as the clustering coefficient distribution and the triad significance profile [135, 136, 274, 411].

We employ the Jensen‚ÄìShannon divergence (JS divergence) [412] to quantify their differences in distribution and the Manhattan distance to measure the differences

of normalised values. Their missing values are replaced with 1 to indicate the large gaps between the network simulations and the target network. Overall, the composite performance index of each network simulation, with values ranging from 0 to 1, is calculated as the weighted average of these values. A lower value of this quality indicator means a more similar simulation compared with the target network.

3.3.2.3 Multi-objective optimisation

Multi-objective optimisation, towards an optimal similarity and efficiency, is required for the iterative application of DT-SNS. To figure out an optimised sDNA in this process, we employ the fast elitist Non-dominated Sorting Genetic Algorithm (NSGA-2) proposed by [413] for three reasons.

First, NSGA-2 is a multi-objective evolutionary algorithm, which enables a Pareto-optimal solution set with alternative solutions in or on Pareto-optimal front [413]. Second, NSGA-2 is characterised with low computational requirements, elitist approach, and parameter-less sharing approach [413]. Third and also the most important, NSGA-2, as a type of genetic algorithm, works best in the search of multiple parameters [414] while enabling the optimisation of sDNA based on genetics selection principles. Finally, we employ the Augmented Scalarization Function (ASF) method [415] to find the best solution from the solution set obtained by NSGA-2 algorithm. It requires the definition of weights to characterise the importance of respective performance concerning similarity and efficiency, which enables the generation of an optimal solution.

3.4 Results and Assessment

The social network simulations, which are generated using the extended SNS, vary with the node attributes. These node attributes include the feature $f_k(v_i)$ and the feature's preference: $p_k(v_i)$ and $w_k(v_i)$. Specifically, we assume a population-based preference, where all nodes share the same preference vector $p_k(v_i) \equiv p$ and $w_k(v_i) \equiv w$, without any change across iterations, which are calculated for an optimal level of similarity and efficiency based on a target network. Therefore, the extended SNSs in our experiment fall in generation 1 and generate static networks that are composed of fixed nodes and edges (Tab. 3.5). The Zachary's Karate Club network [416], composed of 34 nodes with one binary attribute, is used as an illustrative example in this experiment. The nodes and edges each represent club members and their ties. Each member has a binary attribute, denoting whether they belong to the "Mr. Hi" club or the "Officer" club. For

each iteration, 4% of edges are added without removal. Ten network statistics are used in the assessment stage, including the density, assortativity, modularity, shortest path length and degree distribution in the global perspective, and the clustering coefficient and the significance profile of four types of triadic closures considering the binary node attributes.

Table 3.5: The complexity of the extended SNSs.

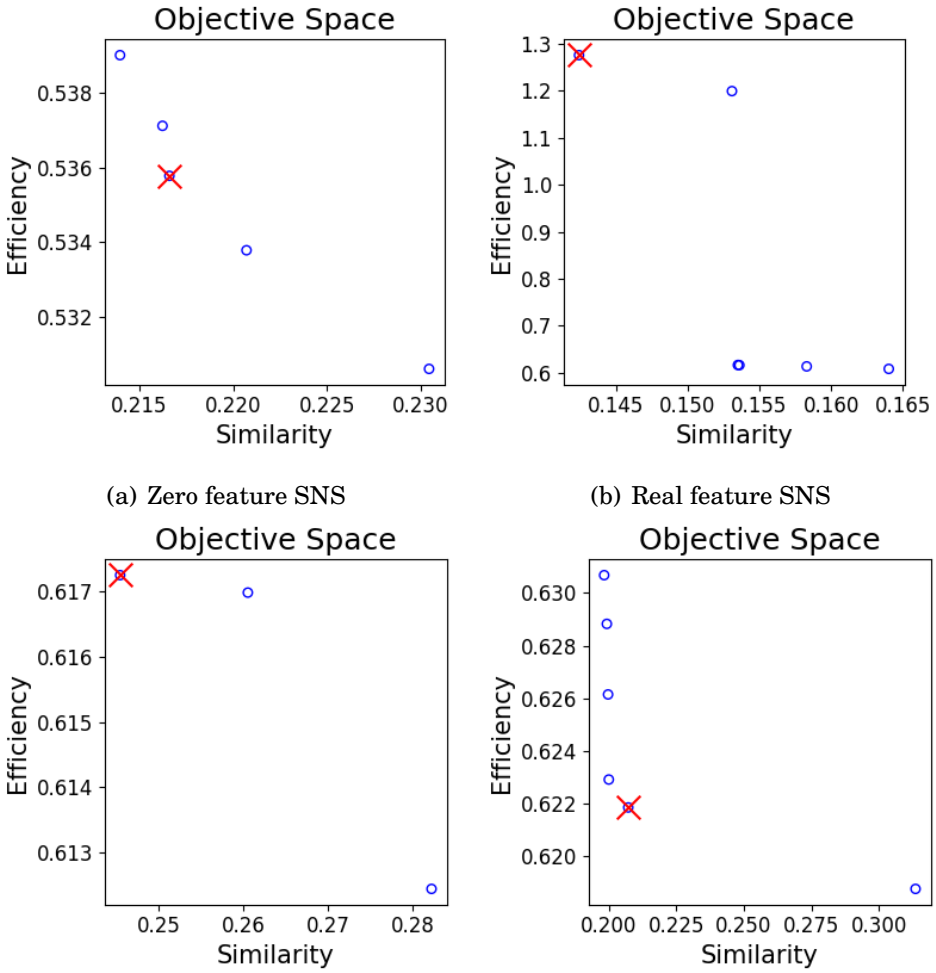
Extended SNSs	Stage	Topology		Attributes		Dynamics
		Nodes	Edges	Nodes'	Edges'	
Zero feature–based SNS	Generation 1	△	△			△
Real feature–based SNS	Generation 1	△	△	△		△
Simulated feature–based SNS	Generation 1	△	△	△		△
hybrid feature–based SNS	Generation 1	△	△	△		△

Under the above assumptions and settings, the impact of feature selection for feature vector $f(v_i)$, concerned with the structural complexity, over the iterations is studied. This involves the benchmark SNS for networks without attributes (zero feature–based SNS), SNS for networks with one real node attribute (real feature–based SNS), SNS for networks with one unobservable node attribute that is randomly simulated as 0 or 1 (simulated feature–based SNS) and SNS for networks that are built with both the real and the simulated attributes (hybrid feature–based SNS). The complexity of the generated social network and the performance of the social network simulator are each presented and analysed in section 3.4.2 and section 3.4.1 respectively. We also characterise the complexity of the extended SNSs in Tab. 3.5. Please note that all of these analyses are intended to provide insights into a discussion of how feasible and challenging the possible pathways to the development of more realistic DT-SNSs are likely to be and what one can expect along such a journey.

3.4.1 Performance of the proposed social network simulator

We measure the similarity between the simulated social networks and the target network with the composite performance indexes based on a variety of measures from both the global and the local perspectives and take the runtime of each simulation as an indicator for efficiency. Rather than proposing a performance measure or reaching final conclusions to be universally employed for SNSs, we aim to reveal the challenges of SNS performance evaluation for the future DT-SNSs by discussing the diverse network patterns and the respective varying similarity levels that contribute to the composite performance

index. The performance of the optimised SNSs considering multiple objectives, including similarity and efficiency, are shown in Fig. 3.1.



(a) Zero feature SNS

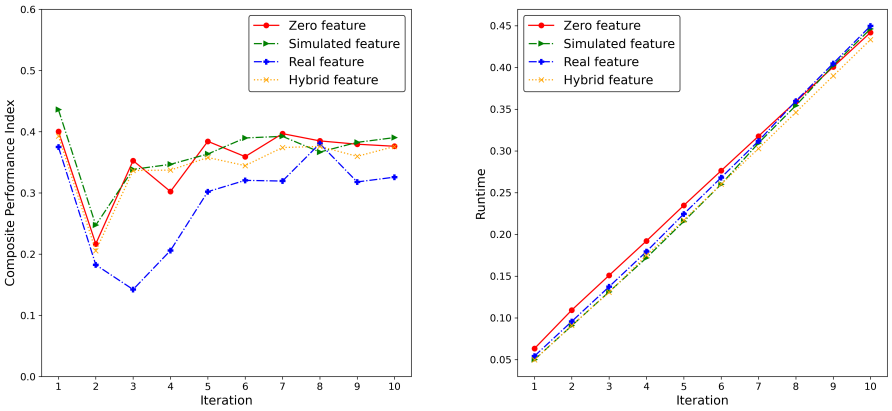
(b) Real feature SNS

(c) Simulated feature SNS

(d) Hybrid feature SNS

Figure 3.1: The performance of optimised SNSs based on similarity and efficiency, including zero feature–based SNS (a), real feature–based SNS (b), simulate feature–based SNS (c) and hybrid feature–based SNS (d)

Fig. 3.1 presents the model performance related to similarity and efficiency for the non-discriminated solutions of SNS optimisation. For each solution, the similarity level decreases with the increase in efficiency. To find the best solution set, we employ the Augmented Scalarization Function (ASF) and attribute higher importance to the similarity objective than the efficiency objective by setting their respective weights as 0.8 (for similarity) and 0.2 (for efficiency). The best solutions, composed of the optimised parameters for the SNSs, are presented in Appendix A.2. Given the best solution, the performance obtained based on the similarity and efficiency is shown in Fig. 3.2.



(a) The comprehensive composite performance index (b) The runtime of the network simulation

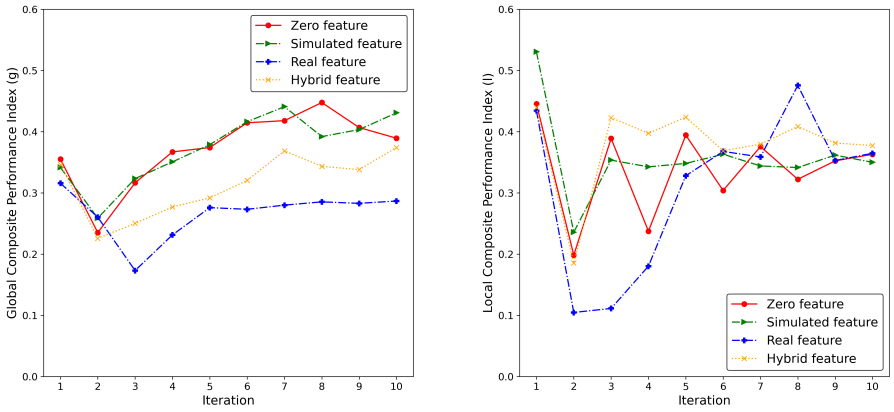
Figure 3.2: The performance of SNS based on the comprehensive composite performance index (a) and the efficiency (b)

As we can see in Fig. 3.2(a), the distance between the simulated social networks and the target network topology fluctuates over the iterations. All generated networks share a similar trend of composite performance index, which firstly decreases in the first two or three iterations and then gradually increases. All the SNSs achieve the lower values of composite performance index around the second iteration. After that, the increase in composite performance index is caused by the increasing number of edges simulated in each iteration that deviates from number of edges in target network. The composite performance index first decreases and reach the lower value when network simulations approach the target with a similar number of edges. It increases when more edges are simulated, leading to deviated network patterns from the target. The real feature–based SNS, with the introduction of one feature, achieves the lowest composite performance index among all the SNSs since the second iteration, within the shortest time. The simulated feature-based SNS introduces a simulated feature to the zero feature-based SNS, which, compared with the zero feature-based and the hybrid feature-based SNSs, produces a generally higher composite performance index over the iterations.

For each SNS, the time spent on each iteration increases as more edges are added and more topological information is considered in the network growth, which, given the similar computation load of feature-based score with the non-changeable feature set, can be attributed to the calculation of popularity-based and shortest path score. There is little difference in runtime among the benchmark: zero feature–based SNS and other SNSs considering the small feature set that varies little.

Overall speaking, the introduction of the real feature to the SNS, with an increased structural complexity, improves the model performance in terms of similarity. In addition to the real feature, the consideration of the unobservable feature and its random simulation with the hybrid feature–based SNS do not bridge its distance with the target. Based on the SNS modelled with one real feature and an optimised population–based preference (the real feature-based SNS), we are able to approach the state of the target network while having the network growth captured in snapshots.

To better understand the contributions of the global measures and the local measures to the composite performance index, we respectively calculate the global composite performance index (Fig. 3.3(a)) and the local composite performance index (Fig. 3.3(b)), which each covers the five global measures and the five local measures of networks that respectively contribute to the 50% of the composite performance index employed in the optimisation process.



(a) The composite performance index from a global perspective

(b) The composite performance index from a local perspective

Figure 3.3: The performance of SNS respectively based on the global and the local composite performance indexes.

Fig. 3.3(a) shows the values of composite performance index from the global perspective (composite performance index (g)) based on the average differences between simulations and target networks in density, modularity, assortativity, shortest path length distribution and degree distribution. For value-based measures, including density, modularity and assortativity, we use Manhattan distance to calculate their respective difference from the target. In contrast, the differences of distribution-based measures including the shortest path length and the degree distributions are measured with the JS divergence. The values of the global composite performance index (g) range from 0.17 (for

real feature–based SNS) to 0.34 (for simulated feature–based SNS), and share a similar decreasing trend in the first two or three iterations with the composite performance index calculated from both the global and the local perspective (Fig. 3.2). Generally all the SNS reach the lowest global composite performance index (g) around the second iteration and then deviate from the target with an increasing global composite performance index (g). The real feature-based SNS, compared with the other SNSs, achieves a lower value starting from the third iteration, similar to the trend of composite performance index calculated from both the global and the local perspective (Fig. 3.2 and Fig. 3.3(a)).

Fig. 3.3(b) shows the composite performance index from the local perspective (composite performance index (l)) based on the differences between simulations and target network in the significance profiles and the clustering coefficient. Similarly, the differences in the value-based measures including the significance profiles, and the distribution-based measures including the clustering coefficient, are respectively measured with the Manhattan distance and the JS divergence. The trend of local composite performance index (l) for each SNS is similar, with a larger value range from 0.10 (for real feature–based SNS) to 0.53 (for simulated feature–based SNS) , which decreases sharply in the first two iterations and then fluctuates around 0.35. Generally all the SNSs reach their lowest composite performance index (l) around 0.2, except for the real feature-based SNS, which achieves the lowest value of 0.10. The real feature-based SNS keeps a lower composite performance index (l) than the other SNSs over the second, third, fourth and fifth iterations.

For each SNS, with its iterative application, the global composite performance index (g) is smaller than the local composite performance index (l) in the first iteration. This indicates that the similarity between real and simulated networks, with small number of edges added, is higher when comparing networks from the global perspective than from the local perspective. Additionally, the real feature-based SNS achieves the smallest composite performance index (best performance) with the highest level of efficiency and similarity when looking at the local and global perspectives separately as well as when they are combined.

3.4.2 Complexity of the social network simulation

We conduct a comparative analysis of the structural complexity built up through the iterative application of zero feature–based SNS, real feature–based SNS, simulated feature–based SNS and the hybrid feature–based SNS. The network states of social network simulations and their changes across iterations are assessed based on the global

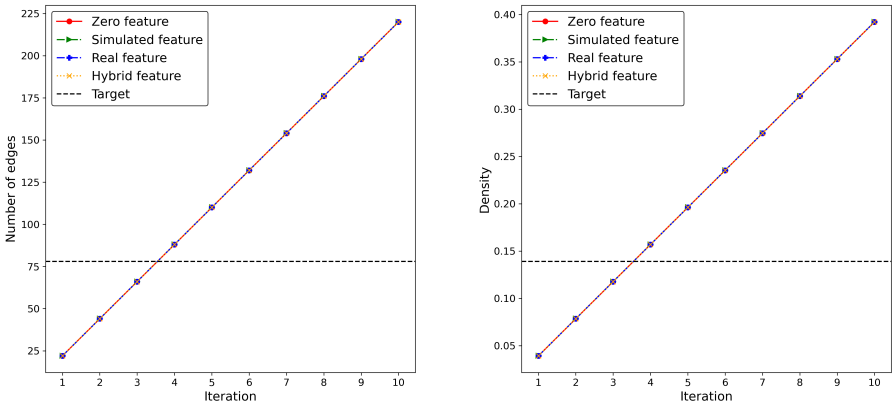
and the local measures involved in the composite performance index of the optimisation process in section 3.3.2.

3.4.2.1 Global perspective

To have a better understanding of the network states and their changes across iterations from the global perspective connected with interactions over large number of actors, we conduct a comparative analysis of social network simulations from each SNS based on density, modularity, assortativity, degree distribution and shortest path length distribution.

3.4.2.2 Density

Density of networks is calculated as a fraction of existing edges to all possible connections between all node pairs [274]. The density of networks increases with more edges added over the iterations. As is shown in Tab. 3.6, there are 78 edges in the target network, with a network density of 0.1390. In Fig. 3.4, all the simulated social networks share the same values of the edge number and density within each iteration under the assumed 4% edges (22 edges) to be added. The number of edges and the density of networks approach the values of the target network in the 4_{th} iteration.



(a) The number of edges (b) Density of networks

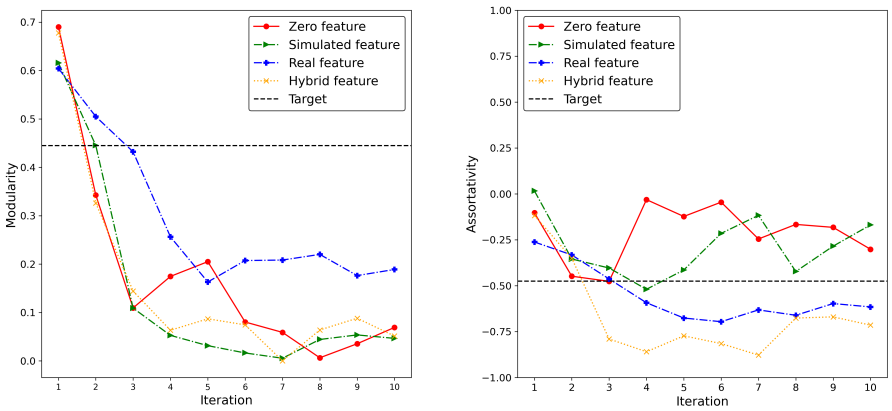
Figure 3.4: The number of edges and the density of social network simulations over the iterations

Table 3.6: The number of nodes and edges, as well as the density of the target network.

	No. nodes	No. edges	Density	Modularity	Assortativity
Target	34	78	0.14	0.42	-0.48

3.4.2.3 Modularity

Modularity measures the extent to which the network is clustered and how strong those clusters are. Its value is between 0 and 1, where a larger value represents a strong community structure. As is shown in Fig. 3.5(a), the modularity of each simulated network decreases over iterations, approaching the target modularity 0.4156 (Tab. 3.6) in the second or third iteration. The real feature–based SNS generally has the largest modularity over the iterations, indicating the strongest community structure for all the simulated networks. This can be caused by the binary feature, as nodes with the same feature and the same preference tend to have similar behaviours, increasing the number of connections within a community. The social network simulated by the simulated feature-based SNS gets nearest to the target modularity in the second iteration.



(a) Modularity of networks (b) Assortivity of networks

Figure 3.5: The modularity and the assortivity of social network simulations

3.4.2.4 Assortativity

Assortativity measures whether the linked nodes have a similar degree (number of connections). Its value ranges between -1 and 1. With a larger positive value of assortativity, nodes tend to link to other nodes with the same or similar degree, indicating a stronger

assortative mixing pattern [274]. As is shown in Fig. 3.5(b), throughout the iterations, the assortativity of simulated networks is negative and fluctuates around −0.5. This can be caused by the weak preference for higher degrees and the strong preference for dissimilar features and shortest path lengths. The zero feature-based SNS is closest to the target assortativity of −0.4762 (Tab. 3.6) in the third iteration, while the other simulated networks approach the target in the second, third or fourth iteration.

3.4.2.5 Degree distribution

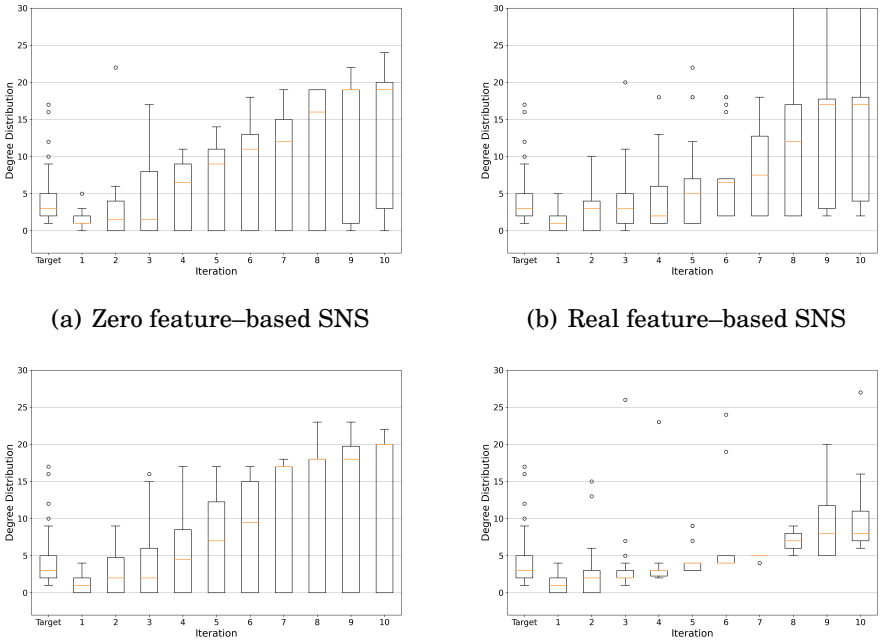
Degree distribution is a distribution of node degrees in a given network [274]. As shown in Tab. 3.7, the node degree of the target network fluctuates around the average value of 4.59 with a standard deviation of 3.82, ranging from 1.00 to 17.00. Fig. 3.6 shows node degree distributions of the target network and the simulated social networks over iterations. For each SNS, the range of the degree distribution generally gets larger with more edges added over the iterations. This indicates that the node degrees vary within the network and change over time, implying the diversity of network topology. The average values of the degree distribution fluctuates around 5 with a small increase over time, which firstly approach the average node degree of the network and then deviates through iterations. The degree distributions of zero feature-based SNS (Fig. 3.6(a)) and the real feature-based SNS (Fig. 3.6(b)) have patterns more similar to the target network than the simulated feature-based SNS (Fig. 3.6(c)) and the hybrid feature-based SNS (Fig. 3.6(d)). And to quantify the level of similarity, we calculate the JS divergence based on the target degree distribution (Fig. 3.7(a)).

Table 3.7: The degree distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minimum
Target	4.59	3.82	5.00	2.00	17.00	1.00

As is shown in Fig. 3.7(a), all the simulated social networks have the same average node degree value, which increases systematically from 1 to 13 with the same number of edges added over the iterations. The average degree approaches the average degree of the target network in the third iteration and then deviates to higher values.

In Fig. 3.7(b), all the simulated networks have similar trend of JS divergence in terms of degree distribution, which firstly decreases to the lowest point at around third iteration and then increases. The hybrid feature-based SNS, in the third iteration, reaches the lowest JS divergence among all the SNS and across all iterations. Correspondingly, the

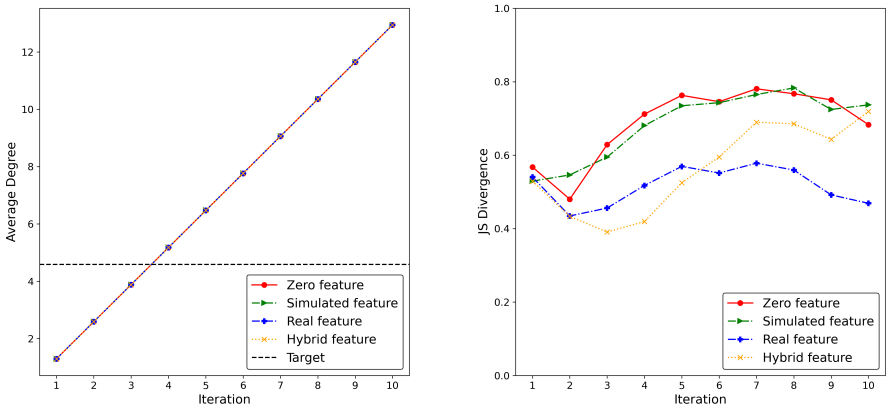


(a) Zero feature–based SNS (b) Real feature–based SNS

(c) Simulated feature–based SNS (d) Hybrid feature–based SNS

Figure 3.6: The degree distribution of networks

degree distribution of the simulated social network obtained using the hybrid feature-based SNS in the fourth iteration (Fig. 3.6(c)) has similar average values and quantiles as the target network, with few nodes having larger degree than nodes in the target network.



(a) Average value (b) JS divergence

Figure 3.7: The average value and the JS divergence of degree distribution.

3.4.2.6 Shortest path length

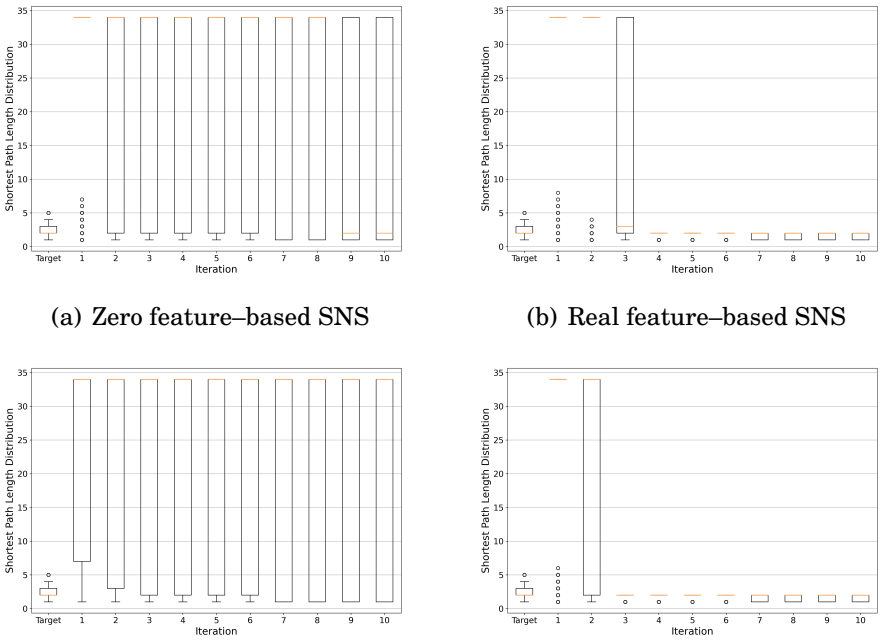
Shortest path length between two nodes is defined as the number of edges along the shortest path between a pair of nodes [274]. As shown in Tab. 3.8, the shortest path length in the target network fluctuates around an average value of 2.41 with a standard deviation of 1, ranging from 1.00 to 5.00. Fig. 3.8 shows the shortest path length distributions of the target network and the simulated social networks over iterations when different SNSs were used. The shortest path lengths of simulated networks value between 1 and 33 as we assume no self-links exist. The shortest path length between connected nodes can not approach an upper limit at 34 given 33 nodes. In contrast, the shortest path length between the unconnected nodes is infinite, which is hard to be measured and visualised in a distribution. Therefore, to describe the unavailable paths between unconnected node pairs, we additionally assign these node pairs with the upper limit for shortest path length 34. For real feature–based SNS and hybrid feature–based SNS, the shortest path length is distributed around the average value over 20 in the first iteration and then converges to 2. This can be caused by the unavailability of paths between node pairs given a small number of edges in the first iteration. For zero feature–based SNS and simulated feature–based SNS, the shortest path length is generally distributed around the average value over 20 over the iterations. This indicates the unavailability of paths between node pairs in the simulation process.

Table 3.8: The shortest path length distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minimum
Target	2.41	0.93	3.00	2.00	5.00	1.00

As is shown in Fig. 3.8(a), the average values of the shortest path length for all for all the social network simulations decrease with addition of edges, approach the target 2.4 and then fluctuate around 2, indicating small-world properties [274]. The hybrid feature–based SNS hits the target in the third iteration within the shortest time.

In Fig. 3.9(b), the JS divergence of shortest path length distribution firstly decreases to the lowest values in the third iteration and then fluctuates around 0.40 for real feature–based SNS and hybrid feature–based SNS. The real feature–based SNS and hybrid feature–based SNS gradually get all the nodes connected and finally result in the same shortest path length distributions, between the value 1 and 2, given the same number of edges (Fig. 3.8). In contrast, the JS divergence of shortest path length distribution for zero feature–based SNS and simulated feature–based SNS fluctuates



(a) Zero feature–based SNS (b) Real feature–based SNS

(c) Simulated feature–based SNS (d) Hybrid feature–based SNS

Figure 3.8: The shortest path length distribution of networks, where the path lengths value between 1 and 34 as we assume no self-links and replace the unavailable (infinite) paths with the lengths of 34.

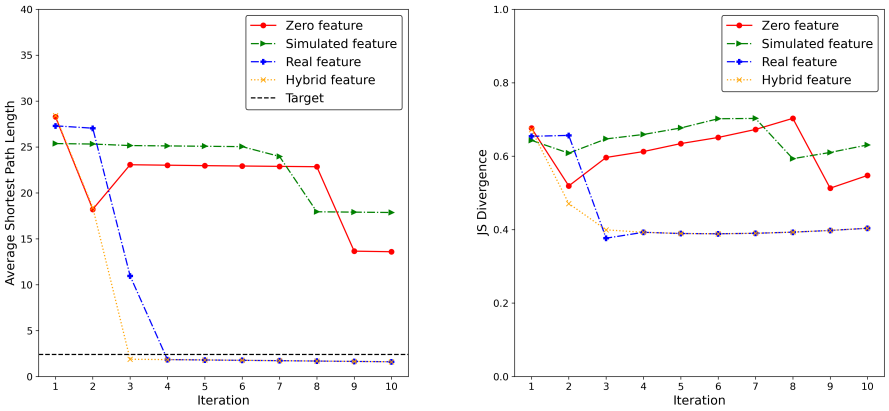
around 0.60. There are unavailable paths between the node paris over iterations for zero feature–based SNS and simulated feature–based SNS. The hybrid feature-based SNS achieves the lowest JS divergence in the third iteration, compared with the other SNSs across iterations.

3.4.2.7 Local perspective

To have a better understanding of the network states and their changes in between, we conduct a comparative analysis of the social networks obtained from running each SNS focusing on the local perspective. We focus here on the clustering coefficient distribution and the significance profiles of the four types of triads.

3.4.2.8 Clustering coefficient

Clustering coefficient describes the probability of a node's neighbours to be connected. Its value is between 0 and 1 [274]. As shown in Tab. 3.9, in the target network, clustering coefficient fluctuates around an average value of 0.57 with a standard deviation of 0.34, ranging from 0.00 to 1.00. Fig. 3.10 shows, for each SNS and each iteration, the clustering



(a) Average value (b) JS divergence

Figure 3.9: The average value and the JS divergence of shortest path length distribution.

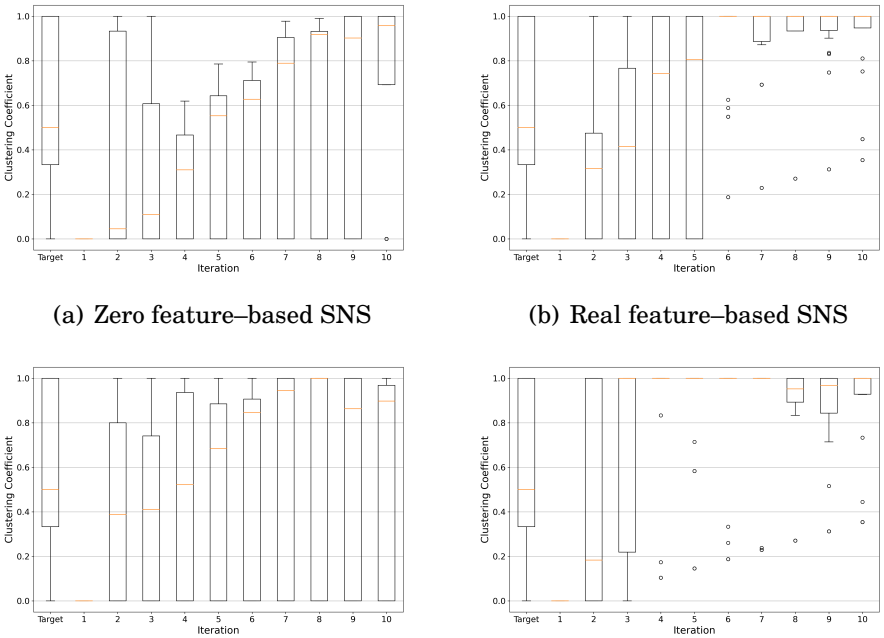
coefficient distribution of the simulated social networks. For all the networks, the values of the clustering coefficient start with 0 in the first iteration and then increase over iterations, with its distribution getting closer to that of the target and then converging to the value 1. The clustering coefficient values of the feature-based SNSs converge faster than that of the zero feature-based SNS. The real feature–based SNS gets closest to the target in the second iteration within the shortest time.

Table 3.9: The clustering coefficient distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minmum
Target	0.57	0.34	1.00	0.33	1.00	0.00

As is shown in Fig. 3.11(a), the average values of the clustering coefficient increase over iterations, indicating a larger probability of neighbours of one node to be connected [274]. The average clustering coefficient of the hybrid feature-based SNS converges to the higher bound 1 faster than in other SNSs and keeps steady at around 0.9 starting from the fourth iteration.

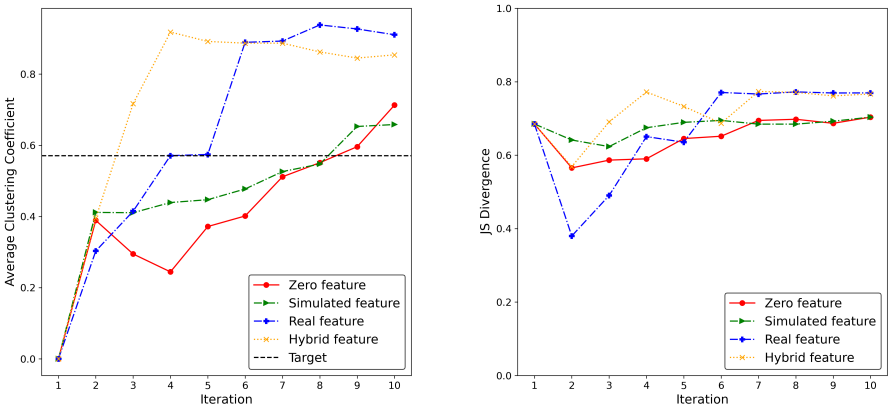
In Fig. 3.11(b), the JS divergences of the real feature-based SNSs firstly decrease to the lowest values at around 0.38 and then gradually increase to around 0.70. Compared with the other SNSs, the real feature-based SNS achieves the lowest JS divergence in the second generation, getting closest to the target network within the shortest time.



(a) Zero feature–based SNS (b) Real feature–based SNS

(c) Simulated feature–based SNS (d) Hybrid feature–based SNS

Figure 3.10: The clustering coefficient distribution of network simulations



(a) Average value (b) JS divergence

Figure 3.11: The average value and the JS divergence of clustering coefficient distribution.

3.4.2.9 Subgraphs significance profile

Subgraphs significance profile describes the similarity of subgraphs and their numbers in a given network when compared to random networks of the same size and number of edges [411]. Assuming M types of subgraphs, the statistical significance of subgraph i ($i \in \{1, \cdots, M\}$) is defined by its Z-score [134, 411]:

$$Z_i = \frac{n_i - < n_i^{rand} >}{\sigma_i^{rand}} \quad i \in \{1, \cdots, M\} \tag{3.10}$$

where n_i, $< n_i^{rand} >$ and σ_i^{rand} represent the number of subgraph i in the studied network, the mean number of its occurrences in the random network ensemble and the corresponding standard deviation respectively. The significance profile SP_i after normalisation is defined in equation 3.11, where a positive value indicates a more often occurrence of subgraph i in a network than that in a set of random network ensembles.

$$SP_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}} \tag{3.11}$$

We focus on the significance profile of a social structure, triadic closure, which is an interconnected three-node subgraph and can be categorised into four types considering the binary node attributes of the target network (Fig. 3.12 and Tab. 3.10). The binary attribute determines whether a node (individual) belongs to the "Mr. Hi" club or the "Officer" club, while the edges represent the relations between the nodes. We respectively identify the triadic closures as triadic closure 1, triadic closure 2, triadic closure 3 and triadic closure 4 based on their node diversity considering the binary attribute (Fig. 3.12).

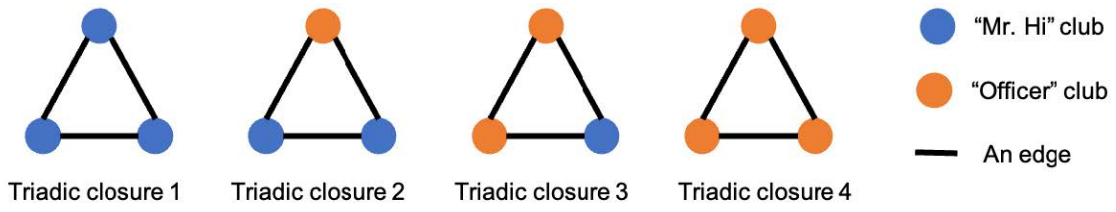


Figure 3.12: The four types of triadic closure in the network simulations

In Tab. 3.10, four types of triadic closures are identified with $i \in \{1, 2, 3, 4\}$ based on a binary node attribute that decides the number of nodes attributed with the "Mr Hi" club or the "Officer" club. There are 15 observations for tragic closure 1 and 26 observations for triadic closure 4, more than the numbers of the other two triadic closures, which means that nodes within the same club are more likely to be connected and form a triadic

Table 3.10: The definition, numbers and significance profiles of triadic closure of the target network.

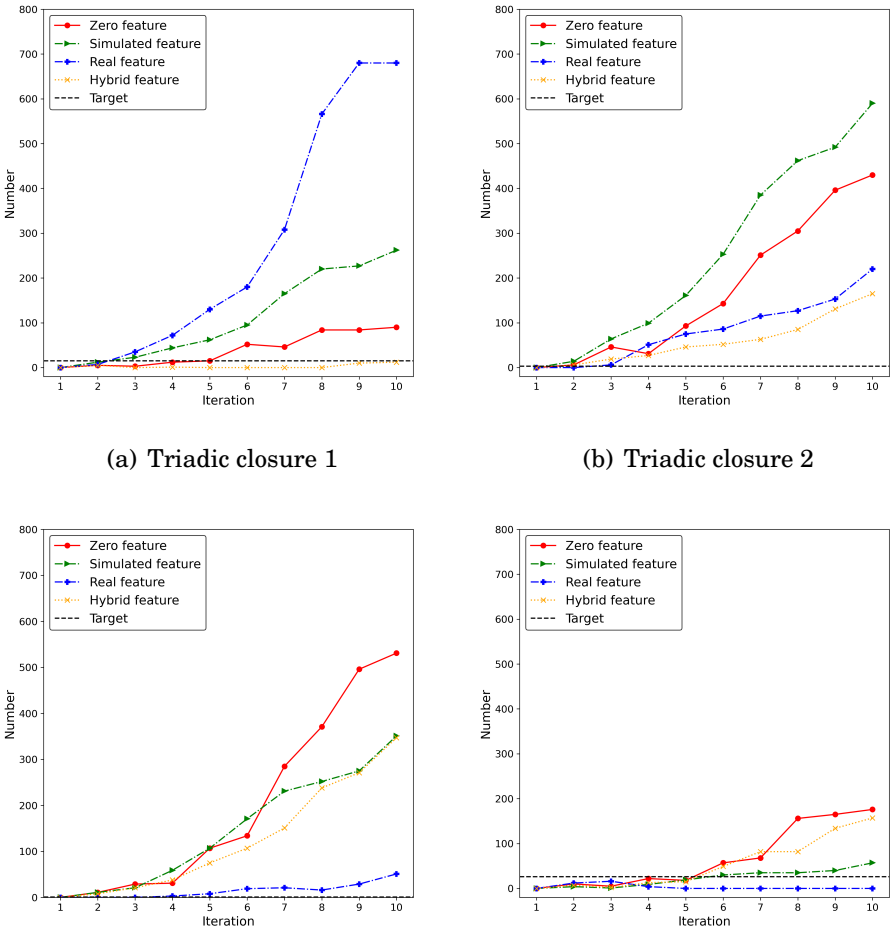
Type	No. nodes		No. subgraphs	Z–score
	"Mr. Hi" club	"Officer" club		
Triadic closure 1	0	3	15	0.52
Triadic closure 2	1	2	3	-0.07
Triadic closure 3	2	1	1	-0.14
Triadic closure 4	3	0	26	0.84

closure. The significance profile of triadic closure 1 and triadic closure 4 is positive, contrasting with the negative values of triadic closure 2 and triadic closure 3. The triadic closure 1 and triadic 4 appear in the target network more often than in a set of random network ensembles.

Fig. 3.13 shows the number of the four triadic closures in the simulated networks, which increases with addition of edges. There are more observations of triadic closure 1, 2 and 3 than triadic closure 4. All the SNSs share a similar increasing trend for triadic closure 1 and 2 over iterations, with the target number reached in around the second iteration (Fig. 3.13(a) and Fig. 3.13(b)). The zero feature–based, the simulated feature–based and the hybrid feature–based SNSs firstly reach the target occurrences of triadic closure 3 in the first iteration and then simulate more of them, compared with the other SNSs (Fig. 3.13(c)). The zero feature-based SNS tends to have more triadic closure 4 simulated and hits the target number in the third iteration within the shortest time (Fig. 3.13(d)). To better understand the significance of these triadic closures, we calculate their Z–scores (Fig. 3.14).

As shown in Fig. 3.14, the positive/negative Z–score indicates a more/less frequent occurrence of triadic closures in comparison to random networks. There are missing Z–scores for the networks simulated by the hybrid feature-based SNS in the first two iterations. This is caused by the unavailability of triadic closures when only small number of edges are added.

The zero feature–based, the real feature–based and the simulated feature-based SNSs achieve a generally positive Z–score for triadic closure 1 over the iterations, while the hybrid feature–based SNS produces negative ones (close to zero) (Fig. 3.14(a)). The SNSs achieve a positive Z–score for triadic closure 2 starting from the second iteration and afterwards deviate from the negative target (Fig. 3.14(b)). All the SNSs generally enable



(a) Triadic closure 1

(b) Triadic closure 2

(c) Triadic closure 3

(d) Triadic closure 4

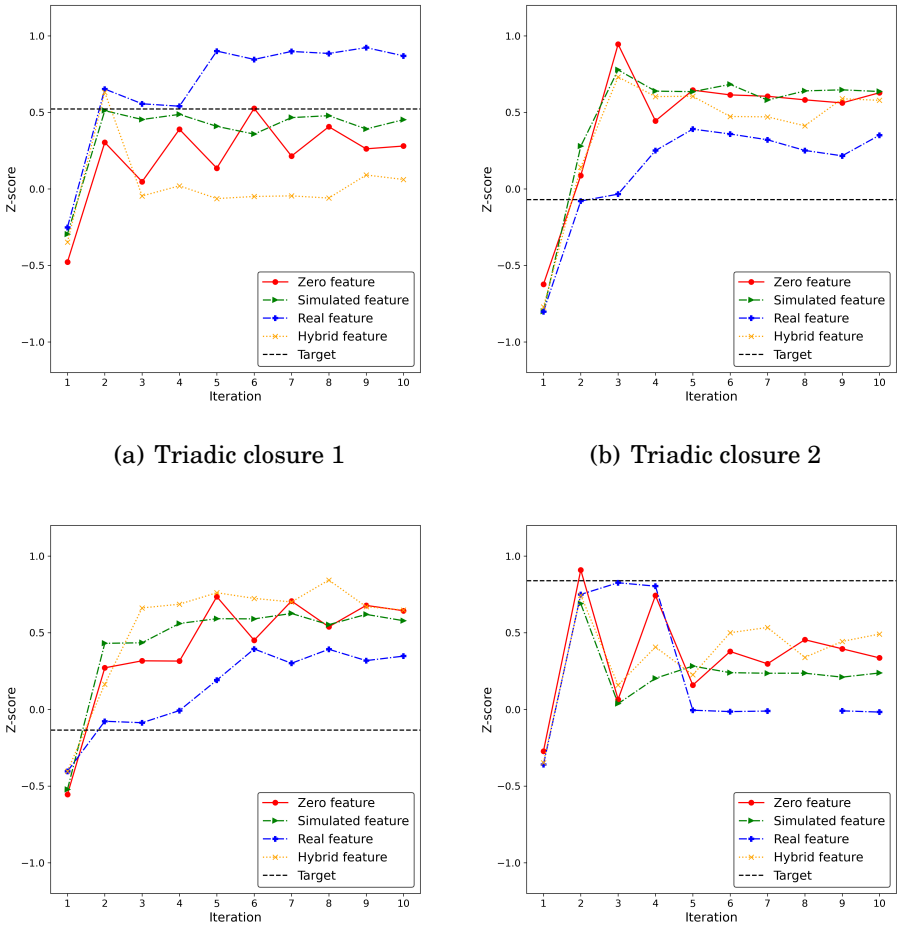
Figure 3.13: The number of the four types of triadic closure in the network simulations

a positive Z–score for triadic closures over iterations, except for the first iteration, where SNSs produce negative values ((Fig. 3.14(a), Fig. 3.14(b), Fig. 3.14(c) and Fig. 3.14(d)).

The real feature–based SNS, compared with the other SNSs, attains a higher value of Z–score for the triadic closure 1, and a generally lower Z–score for the triadic closure 2, the triadic closure 3 and the triadic closure 4. The real feature–based SNS generally gets closest to the target Z–scores, which, for triadic closure 1, 2, 3 and 4, is reached in the fourth, second, second and third iteration respectively (Fig. 3.14).

3.4.2.10 A brief summary

To conclude the analysis, from a global perspective, the density, the modularity and the average node degrees increase with more edges added over the iterations for each



(a) Triadic closure 1

(b) Triadic closure 2

(c) Triadic closure 3

(d) Triadic closure 4

Figure 3.14: Z–score values of triads in the social network simulations

SNS. This indicates a denser network and a stronger community structure and it is the expected result as the number of nodes remains static and only number of edges increases over the iterations. The average shortest path lengths decrease when more nodes get connected, and stabilise at a value that is smaller than 3, showing small-world properties. From a local perspective, the average clustering coefficient and number of triadic closures increase with edge addition over iterations, indicating more interpersonal connections within the subgraphs of the simulated networks.

Among all the SNSs and across iterations, the real feature–based SNS gets closest to the target when considering shortest path length distribution, clustering coefficient distribution and subgraph significance profiles. The simulated feature–based SNS gets closest to the target modularity. The zero feature–based SNS gets closest to the target

assortativity. The hybrid feature–baseed SNS gets closest to the degree distribution. Generally all the network measures for all the social network simulators approach their target values and then deviate from them as the number of edges exceeds the number of edges of the target network. The SNS that achieves the closest distance to the target network measure (as measured through JS divergence) varies from case to case and requires different numbers of iterations. Real feature-based SNS can generally achieve a higher level of similarity with the target value of each measure through a smaller number of iterations within a shorter time and this indicates that including different features in the simulation process leads to different network patterns, which may respectively deviate from that of the real networks to a different extent. The trade-offs between the similarity levels considering multiple network measures pose a challenge to the performance evaluation of SNSs and their extension towards a DT. Current studies on network simulators generally employ a single network measure or combine a few to evaluate the similarity between the simulated and real social networks. The selection and combination of various features in the SNS extension and the selection and combination of multiple network measures for SNS performance evaluation in a systemic way remains a research gap and requires further study.

3.5 Conclusions

In this chapter, we firstly review the current state-of-the-art of the Social Network Simulators under the modelling and assessment framework proposed in the chapter 2 and identify research gaps in the space of social network modelling. To progress the field further and achieve a simulator that can better model real networks, we extend one of the promising SNS proposed by [34] towards a Digital Twin Oriented SNS by proposing one possible pathway of increasing structural complexity, which is to include informative features that are simulated or observed in the real world. With this SNS, each node attribute is composed of a node feature and a social DNA (the node's preference for a similar feature and the corresponding weight of preference).

We test different settings of sDNA in DT-SNS to see which one results in social network that is closest to the target network. We propose to assess the similarity between target and simulated networks using metrics calculated at both global and local levels while taking the runtime of SNS as an indicator of efficiency. This DT-SNS also serves as a tool to analyse the complexity of social network simulations built up over iterations. As an illustrative example, we conduct experiments and assessments of this DT-SNS

with the Karate Club network, which results show the possibility of optimisation and assessment of DT-SNS towards a Digital Twin.

Literature review revealed that the majority of existing SNSs focus on generating purely simulated networks, while only a small proportion of SNSs aim to model real nodes and node attributes with/without observable information about edges. Under this constraint, we review and discuss the modelling social networks through generations towards the ultimate goal of a DT. Current SNSs generally focus on static networks or dynamic networks. Few SNSs consider networks, the dynamic process on the network and their interrelations. The interrelated social networks and dynamic processes, with the continuous acquisition of real-time information and feedback, also remain a research gap and require further study. We also discuss the complexity of social network simulations from the perspective of four dimensions and review the current ways of assessing SNSs. Most SNSs simulate social networks composed of fixed nodes and edges, while some SNSs incorporate network attributes and slow changes in network topologies. Current SNSs consider the global level similarity between the social network simulation and the target network as an indicator of its performance, where neither the local level similarity nor the efficiency is incorporated.

In the experiment, we extend an SNS towards a DT-SNS to help illustrate the possible pathways and the challenges of developing a DT-SNS. The extension of an SNS towards a DT-SNS incorporates different levels of structural complexity and involves experiments on the benchmark, zero feature-based SNS, and its extensions, including simulated feature-based SNS, real feature-based SNS and the hybrid feature-based SNS concerned about both the real and the simulated features. We calculate the composite performance index of different networks by combining the results of various network measures. For a deeper understanding of the composite performance index that changes over iterations, we also conduct a comparative analysis of these SNSs to see the complexity of their social network simulations over iterations. The experimental results show that among all the SNSs, the real feature-based SNS, with an appropriately increased structural complexity, has the best performance in the efficiency and similarity from both the local and the global perspective. To be more specific with each measure, real feature-based SNS can also get closer to the target value through a smaller number of iterations. However, this conclusion holds for the experiments presented in this chapter, considering the pathway of SNS extension, the information employed in network simulation and the way we calculate the composite performance index. We need to develop a more systematic approach to developing DT-SNS and propose referable SNS evaluation metrics

for further study. And as SNSs add more edges over the iterations, the runtime of SNSs increases, and a stronger community structure and an assortative mixing pattern emerge, with more triadic closures as more neighbouring nodes get connected. Overall, all the measures involved in the composite performance index approach their target values and then deviate from them as the simulated network exceeds the target network's density. However, different similarity levels achieved by SNSs considering different network measures reveal the challenge of an accurate SNS performance evaluation given the specific requirements of social network simulations. This is a research gap to be addressed in our future study.

Overall, this chapter answers **RQ2:** by a topical review on the current state-of-the-art Social Network Simulators (SNSs) and an experiment on developing a current-existing SNS towards a DT-SNS. The research contributions can be summarised as follows:

- **RQ2**: How to build and evaluate DT-SNS that accommodates heterogeneous features of nodes?

    1. **RO2.1**: Develop a DT-Oriented Social Network Simulator that enables social network simulation based on heterogeneous node features.

       RC2.1: DT-SNS that enables to faithfully represent static network topology with a given complexity level.

    2. **RO2.2**: Evaluate proposed DT-SNS.

       RC2.2: A case study on the modelling and assessment of DT-SNS with different levels of structural complexity.

The DT-SNSs can be extended with structural variations and temporal changes to approach a DT of the real systems. There is a requirement for a future study on the sDNA that varies across groups or individuals. Further research is also required on the network topology that changes over time and the process dimension that simultaneously interacts with the network dimension. Specific SNS performance evaluation criteria, considering various SNS complexity levels, also require further study. More generally, this chapter serves as a starting point of our future work on exploring the complexities of the real systems, and in the next chapter (chapter 4), we first progress the DT-SNS with higher complexity levels by proposing a generation 1 DT-SNS modelling framework for static networks and the dynamic processes on these networks.

CHAPTER 4

HETEROGENEOUS NODE FEATURES AND INTERACTION RULES FOR DT-CNS MODEL

4.1 Introduction

Over the years, Complex Networked Systems (CNSs) have been modelled with increasing structural complexity levels that resulted in models with higher and higher accuracy. The ultimate goal of these research efforts is to build models towards accurate reflection and, sometimes, extension of reality. Family of models that aim to perfectly reflect real world networked systems can be seen as Digital Twin Oriented Complex Networked Systems (DT-CNSs) [1]. Our previous discussion conceptualises ideas of DT-CNS models and proposes the related generations of modelling framework that progress with increasing complexity levels (chapter 2). Based on this theoretical basis, we propose to build and extend a DT-SNS considering the increasing structural complexity levels related to the heterogeneous node features (chapter 2). There are no current realisations of DT-CNS models and the goal of the research presented in this chapter is to bring us closer to the ultimate goal of DT-CNS models with an example of social networked system.

As reviewed and analysed in chapter 2, the complexity of CNS models results from the heterogeneity of CNS models' components and interactions between them as well as dynamic processes (e.g. epidemic spread) on those systems. Current studies on CNS models focus on the least complex scenario where a single dynamic process takes place on a static network without changes of the network components (nodes and edges) or

the process parameters [31, 32, 49–52]. To increase complexity of modelling, and in this way to approach Digital Twin, a natural way forward is to increase the heterogeneity of nodes' features (characteristics of each specific node) and their preferences to create relationships as well as allow the system to evolve over time, both from the perspective of structure and process.

The complexity of modelling CNS is constrained by the system's observability. The observability is concerned with the ability to faithfully reconstruct the state of a system from a limited set of measured variables in finite time [1, 79]. It not only determines the available ground-truth information to be represented and modelled but also the information to be simulated towards achieving the ultimate goal of a Digital Twin (DT). With different levels of observability, we categorise the CNS components as data-driven, simulation-based and hybrid [1], each composed of real data [19, 203, 250], purely simulated data [34] and both real and simulated data respectively. Current studies generally build CNS models with hybrid components (e.g. real networks and simulated dynamic processes) due to the partial observability of a real-world scenario [17, 31, 32, 49–52, 406, 407].

Therefore, we propose **RQ3** and answer this question by addressing three research objectives.

- **RQ3**: How to build and evaluate DT-CNS model considering the feature preferences?

    1. **RO3.1**: Develop an extensible DT-CNS model based on heterogeneous node features and rules of feature preference.
    2. **RO3.2**: Develop evaluation protocol for DT-CNS model.
    3. **RO3.3**: Evaluate the proposed DT-CNS model.

In this chapter, we propose an extendable modelling framework for DT-CNSs in social network space (**RO3.1**), composed of (i) a flexible social network model that allows for static network formation based on heterogeneous node features and interaction rules related to connection preferences, and (ii) a process model for the Susceptible-Infected (SI) epidemic spread with a predetermined seed selection strategy and an infection rate. We propose a set of evaluation metrics to validate the faithfulness of DT-CNS models, considering their similarity with real system, efficiency and reproducibility of the network patterns and the infection occurrence on such networks (**RO3.2**). We evaluate the proposed DT-CNS model and present a case study on the disaster resilience of social

networks considering their infection occurrence within specific time and social distance given an epidemic outbreak (**RQ3.3**). Our experiments focus on various simulated networks, where nodes have heterogeneous attributes and the interactions between the nodes are driven by various rules (e.g. social networks driven by age features and an interaction rule – preference for a similar age when nodes interact). The results show that the heterogeneous features and interaction rules (related to preference for features) influence the network structure and can enhance epidemic outbreak. Nodes with preferred by others features have more connections and, as a result, get exposed to higher infection risk within the same proximity to the initial infection node.

The main contributions of this chapter are:

- Constructing an extendable DT-CNS modelling framework composed of a network model based on heterogeneous nodes' features and rules driving the interactions between nodes and a process model spreading over the network (**RC3.1**);
- Creating an evaluation protocol for faithfulness, concerned with similarity, efficiency and reproducibility (**RC3.2**);
- Validating the influence of heterogeneous nodes' features and interaction rules on network growth and level of epidemic spread (**RC3.3**);
- Suggesting disaster mitigation policies dependent on age diversity and social preference (**RC3.3**).

The rest of this chapter is structured as follows. Section 4.2 provides background in the space of dynamics of and on the networks. Section 4.3 presents the methodology of building and driving a CNS model towards a DT. Following this, section 4.4 builds and evaluates the simulation-based CNS models. Finally, in section 4.5, we conclude the study and outline future directions.

4.2 Background

This section presents the network and process models developed and employed in the current studies. It discusses (i) interaction rules that drive the network growth through preference for node features and (ii) transmission rules governing the node adaptability to an epidemic spread in different conditions.

Network model aims at providing a faithful representation of a network that minimises information loss [1]. There are two widely employed principles for network formation and growth: preferential attachment [45, 53–55] and homophily [56–59]. By incorporating these two principles, this chapter starts to build DT-CNS models based on observable networked information

Preferential attachment describes the "rich get richer" principle [55], where highly connected nodes increase their connectivity faster than their less-connected peers [53]. A scale-free network model based on the preferential attachment to higher node degrees was first proposed in [45]. With time, more and more studies in the network science space started looking into other topological features, such as clustering coefficient, closeness centrality and eigenvector centrality [54, 55]. In our study, we allow preferential attachment to the nodes with specific feature values concerned with topology and attributes.

Tendency to associate ourselves with other similar to us is called homophily [56]. It can either be choice homophily or induced homophily. The first one is a preference for similar attributes, e.g. people tend to interact with those similar interests. Induced homophily is a result of the interaction opportunities due to structural proximity [57–59] (e.g. people who interact because they work in the same building). In the network embedding paradigms, the induced homophily generally suggests that highly connected nodes should be embedded closely in the latent representation space [417]. In contrast, the choice homophily effect leads to the consideration of attribute proximity [418]. In our study, we account for both choice homophily and induced homophily based on the proximity of structures and attributes.

Process model for the spreading phenomenon on networks has two components: (i) seed selection strategy for the first contagious nodes, and (ii) the transmissibility that describes the spread conditional on node features [14, 18, 419, 420].

Seed selection strategies identify single/multiple seeds (source nodes of the spreading process) in the initial/subsequent steps (single-stage seeding/sequential seeding) from which the spread starts. Current studies generally select seeds based on the centrality measures, such as degree centrality, betweenness centrality, closeness centrality and eigenvector centrality [421, 422] or identify driver nodes and using them as seeds [419].

The transmissibility involved in the contagion models depends on many, different elements. For example, node adoptability, concerning the probability to adopt a spread, rises when there are multiple exposures for multiple contagions/simplicial contagion, compared with a simple contagion [333, 406]. In addition, the transmissibility can also

vary with specific interactions (e.g. physical contact, airborne spread, etc.) among node pairs or groups [67]. Particularly, [333] explains the current contagion models with conditional transmissibility, such as the threshold model, stochastic contagion model, diffusion percolation model, independent cascade model and Reed–Frost model.

Overall, current network models are built on interaction rules, either related to preference for topological features (preferential attachment) or preference for similar node attributes (homophily). Current rule-based process models focus on seed selection strategies and transmission rules. However, none of the studies combines heterogeneous features and interaction rules and investigates their influence on the network growth and the corresponding epidemic spread. Therefore, this chapter proposes an extendable modelling framework for DT-CNSs based on heterogeneous node features and diverse interaction rules.

4.3 Digital Twin Orientated Complex Networked System

In this part, we provide a description of an extendable CNS simulator in section 4.3.1 and propose an evaluation protocol for the faithfulness of CNS models in section 4.3.2. In section 4.3.3 we describe the optimisation process to obtain CNS models that faithfully represent reality and approach DT.

4.3.1 An extendable DT-CNS modelling framework

The extendable CNS simulator enables to model the dynamics of and on the social networks. It is devised with (i) a network model based on the interaction rules of homophily and preferential attachment, as well as (ii) a process model (transmission mechanism) that uses different seed selection strategies and transmission rules dependent on varying conditions.

4.3.1.1 Network and Network Dynamics

In this chapter, we focus on social networks composed of nodes (people), edges (people's interactions) and the attributes of nodes and edges. We model the interactions between nodes with network dynamics governed by the interaction rules.

Network at time t can be represented as $G_t = \{V_t, E_t, A_t, Z_t\}$ based on the observable and simulated information of the changeable network components: nodes V_t =

$\{v_{1,t}, \cdots, v_{n_t,t}\}$, edges $E_t = \{e(v_{i,t}, v_{j,t}) | v_{i,t}, v_{j,t} \in V, i \neq j\}$, node attributes $A_t = \{a(v_{i,t}) | v_{i,t} \in V\}$ and the edge attributes $Z_t = \{z(v_{i,t}, v_{j,t}) | v_{i,t}, v_{i,t} \in V, i \neq j\}$. Network dynamics drives network formation and growth over time.

The node attribute vector $a(v_{i,t})$ for node v_i, referring to [34], is defined as

$$a(v_{i,t}) = \{\mathbf{f}(v_{i,t}), \mathbf{p}(v_{t,i}), \mathbf{w}^p(v_{t,i}), \mathbf{h}(v_{i,t}), \mathbf{w}^h(v_{t,i})\} \quad v_i \in V \tag{4.1}$$

which includes an l-length feature vector $\mathbf{f}(v_{i,t})$ (l is the number of features) and a social-DNA (sDNA) defined with another four l-length vectors. Features presented in $\mathbf{f}(v_{i,t})$ are characteristics that describe each individual $v_{i,t}$, such as age, gender and interests. $\mathbf{p}(v_{t,i})$ determines the preference for features with a trinary value: 0, 1, or -1, which each, respectively, indicates: 0 – no preferential attachment to the corresponding node feature in $\mathbf{f}(v_{i,t})$, 1 – preferential attachment to a higher feature value or -1 – to a lower feature value. For example, the scale-free network model, where nodes connect with other popular nodes, uses pure preferential attachment (a favourable preference) to node degrees. Similarly, $\mathbf{h}(v_{t,i})$ determines the preference for similar features with a trinary value: 0, 1, -1, each, respectively, representing no homophily (0), heterophily (1) and homophily (-1). For example, in a social network where people like to interact with people of similar ages, we would use pure homophily to describe the negative preference for age differences. Both of these preferences are followed with a weighting factor $\mathbf{w}^p(v_{t,i})$ and $\mathbf{w}^h(v_{t,i})$ within the range of $(0, 1]$.

The feature $\mathbf{f}(v_{i,t})$ and the sDNA $\{\mathbf{p}(v_{t,i}), \mathbf{w}^p(v_{t,i}), \mathbf{h}(v_{i,t}), \mathbf{w}^h(v_{t,i})\}$ can vary with nodes in a frozen time scale and/or mutate over time. They can also co-evolve with the network topology based on the rules of network growth:

$$\pi(v_{i,t}, v_{j,t}) = [\frac{1}{2}\pi^p(v_{i,t}, v_{j,t}) + \frac{1}{2}\pi^h(v_{i,t}, v_{j,t}) + \epsilon_{ij,t}] * I_{ij,t} \quad v_i, v_j \in V, i \neq j \tag{4.2}$$

where the network growth is driven by ranks of score $\pi(v_{i,t}, v_{j,t})$ of any node pair through mutual evaluation concerned about their preferences. It is calculated as the sum of preferential attachment score $\pi^p(v_{i,t}, v_{j,t})$, the homophily score $\pi^h(v_{i,t}, v_{j,t})$ and the random interference $\epsilon_{ij,t} \sim \mathcal{N}(0, \sigma^2)$ that follows a random normal distribution. $I_{ij,t}(\eta)$ is a binary vector that denotes the an encounters or zero encounters of this node pair with 1 or 0 dependent on the encounter rate $\eta \in [0, 1]$.

The preferential attachment score $\pi^p(v_{i,t}, v_{j,t})$ incorporates the preference $\mathbf{p}(v_{i,t})$ for the other nodes with higher/lower feature values for $\mathbf{f}(v_{i,t})$:

$$\pi^p(v_{i,t}, v_{j,t}) = \frac{1}{2l}\mathbf{f}(v_{j,t})^\tau(\mathbf{p}(v_{i,t}) \odot \mathbf{w}^p(v_{i,t})) + \frac{1}{2l}\mathbf{f}(v_{i,t})^\tau(\mathbf{p}(v_{j,t}) \odot \mathbf{w}^p(v_{j,t})) + 1 \tag{4.3}$$

The homophily score $\pi^h(v_{i,t}, v_{j,t})$ incorporates the preference $p(v_{i,t})$ for the other nodes with similar/dissimilar feature values $f(v_{i,t})$:

$$\begin{aligned}\pi^h(v_{i,t}, v_{j,t}) = & \quad \frac{1}{2l}|\mathbf{f}(v_{i,t}) - \mathbf{f}(v_{j,t})|^{\tau}(\mathbf{h}(v_{i,t}) \odot \mathbf{w}^h(v_{i,t})) \\ + & \quad \frac{1}{2l}|\mathbf{f}(v_{i,t}) - \mathbf{f}(v_{j,t})|^{\tau}(\mathbf{h}(v_{j,t}) \odot \mathbf{w}^h(v_{j,t})) + 1\end{aligned} \tag{4.4}$$

The score list $\Pi_t = \{\pi(v_{i,t}, v_(j,t))|v_{i,t}, v_{i,t} \in V, i \neq j\}$ stays frozen with fixed nodes and node attributes (features and sDNA). The CNS simulator connects $\lambda_{e,t}$ pairs of nodes based on the score ranks and generates a static network to represent discrete interactions at this frozen time point, with an edge intensity of

$$\gamma(v_{i,t}, v_{j,t}) = \begin{cases} \frac{\pi(v_{i,t}, v_j, t)+2l}{4l} & \text{if } \lambda_{e,t} \text{ is reached;} \\ 0 & \text{if else.} \end{cases} \tag{4.5}$$

The score list Π_t changes with any node feature change/preference mutation. This may result in changes of node pair ranks and their connections, finally leading to network evolution.

4.3.1.2 Process and Process Dynamics

In this chapter, we also investigate the epidemic spreading process on the social networks. Process Dynamics drives the spreading process $\mathrm{P} = \{\Omega, Pr_t, \mathbf{B}_t, \mathbf{R}_t\}$ on the networks and is defined by the following elements: seed selection strategy Ω, transmission probability Pr_t given others' adoptability $\mathbf{B}_t = \{\mathbf{b}(v_{i,t})|v_{i,t} \in V_t\}$ and the resulting infection status $\mathbf{R}_t = \{\mathbf{r}(v_{i,t})|v_{i,t} \in V_t\}$.

The seed selection strategy identifies the first contagious nodes at the beginning of epidemic spread. The seed selection strategy $\Omega = \{\omega_1, \omega_2, \cdots\}$ for single/multiple seeds in the initial/sequential stages is defined with a set of rules ω_t:

$$\omega_t(\mathbf{s}_t, \mathbf{w}_t^s, k_t) \tag{4.6}$$

where k_t represents the number of seeds selected at time t. $\mathbf{s}_t$ is a l-length vector that determines the seed preference for node features $\mathbf{f}(v_{i,t})$ with values: 0, 1 or -1, each representing no preference, preference for higher/lower feature values, respectively. This is accompanied by a l-length weighting vector $\mathbf{w}_t^s$, including weights of preference within

range $(0,1]$. The first k_t seeds are selected at time t, dependent on the descending ranks of node score $\epsilon(v_{i,t})$:

$$\epsilon(v_{i,t}) = \mathbf{f}(v_{i,t})^{\tau}(\mathbf{s} \odot \mathbf{w}^{s}) \tag{4.7}$$

For example, we can build a seed selection strategy based on node popularity preference. In this context, popular nodes, featured with significant node degrees, are preferred and assigned high score ranks, finally selected as the contagious seeds.

The seed selection results in the nodes' infection status $\mathbf{r}(v_{i,t})$, a m-length vector that determines whether the node is infected or not with binary values of 0 or 1. Its transition from the previous infection status $\mathbf{r}(v_{i,t-1})$ to the current status $\mathbf{r}(v_{i,t})$ is dependent on the node adoptability $\alpha(v_{i,t})$, also termed as the Process DNA (pDNA):

$$\mathbf{b}(v_{i,t}) = \{\mathbf{c}(v_{i,t}), \Theta(v_{i,t}), \mathbf{w}^{c}(v_{i,t})\} \tag{4.8}$$

The pDNA is represented through three q-length vectors: $\mathbf{c}(v_{i,t}) = [c_1(v_{i,t}), \cdots, c_q(v_{i,t})]$ specifies the q independent conditions of node $v_{i,t}$ considering the node infection status $\mathbf{r}(v_{i,t})$ and the node features $\mathbf{f}(v_{i,t})$. The independent conditions presented in $\mathbf{c}(v_{i,t})$ identify various individual characteristics that pose each individual at infection risk and determine how the epidemic spreads. As an example, there is an epidemic spread between males with physical contact. In this context, the independent conditions include gender (male/female) and exposure (with/without physical contact).

$\Theta(v_{i,t}) = [\theta_1(v_{i,t}), \cdots, \theta_q(v_{i,t})]$ represents the adoptability-thresholds to be considered when evaluating each node condition; and $\mathbf{w}^{c}(v_{i,t}) = [w_1^{c}(v_{i,t}), \cdots, w_q^{c}(v_{i,t})]$ includes the multiplier effect of node adoptability on state transition, within the value range $(0,1]$.

With the pDNA $\mathbf{b}(v_{i,t})$, the transition probability $Pr(\mathbf{r}(v_{i,t})|\mathbf{r}(v_{i,t+\Delta t}))$ can be calculated as:

$$log(Pr(\mathbf{r}(v_{i,t})|\mathbf{r}(v_{i,t-\Delta t}))) = \sum_{j=1}^{q} \delta_{c_j(v_{i,t}),\theta_j(v_{i,t})} * log(w_j^{c}(v_{i,t})) \tag{4.9}$$

where δ is a Kronecker function that determines, with binary values between 0 and 1, whether a specific condition $c_j(v_{i,t})$ for node $v_{i,t}$ meets the threshold $\theta_j(v_{i,t})$. To be more specific, $\delta_{c_j(v_{i,t}),\theta_j(v_{i,t})} = 0$ indicates the elimination of the multiplier effect $w_j^{c}(v_{i,t})$, while given $\delta_{c_j(v_{i,t}),\theta_j(v_{i,t})} = 1$, $w_j^{c}(v_{i,t}) \longrightarrow 0$ indicates full resilience of the node.

4.3.2 An Evaluation Protocol on Faithfulness

Below, we propose an evaluation protocol to formally capture different interpretations of faithfulness under the constraints of observability that vary with the no/complete/partial ground-truth scenarios.

4.3.2.1 Complexity

Emergent patterns observed in CNS models and their complexity are the starting point of evaluation. The evaluation metrics of the emergent patterns existing in CNS models in the Network dimension and the Process dimension (Tab. 4.1) respectively focus on the characteristics of network topology, network attributes, the transmissibility of the spread and the infection occurrence.

Table 4.1: Measures of complex patterns emerging from CNS representation.

CNS components		Pinpoint	Local	Global
Network	Topology	Node[112], Edge [46]	Triad [386], Quadrangle [136], Community [87, 160, 198], Clustering coefficient [135], Closure coefficient [135]	Degree distribution [274], Shortest path length [274], Modularity [232, 274], Assortativity, Betweenness, Closeness.
	Attribute	Node attribute [289], Edge attribute [164, 278]		Feature distribution, Hill numbers [423]
Process	Spread	Transmissibility to nodes	Transmissibility to groups	Transmissibility to the population
	Infection	Infected nodes	Infected groups	Infected population

These measures range from (i) the pinpoint level connected to the individual component of the network/process (e.g. a node/edge added/removed over time; transmissibility to a node,together with its resulting infection status); (ii) the local level concerned with the local interactions among grouped individuals (e.g. clustering coefficient employed to assess local structure); and (iii) the global level concerned with the emergent global characteristics resulting from the interactions within the population (e.g. degree distribution based on the number of edges connected to each node; Hill numbers that describe the diversity of the population (nodes) based on node attributes).

4.3.2.2 Performance

Performance of modelling CNS dynamics, given the quantified complexity of CNS model, can be evaluated in respect to different requirements in terms of the CNS model patterns' reproducibility, the corresponding similarity between the CNS model patterns with a target state (e.g. the target network topology required in link prediction tasks; etc.) as well as efficiency.

Table 4.2: Measures used to assess performance of simulation, prediction and/or control models of CNS dynamics.

Requirement	Level	Measures
Similarity	Pinpoint	Precision [9, 46], Recall [181], Area Under the Precision Recall (AUPR) curve [372], Receiver Operating Characteristic (ROC) curves, Area Under the ROC (AUC) [10], Geometric Mean of AUC and PRAUC (GMAUC) [373], Error Rate [299], SumD [298], Kendall's Tau Coefficient (KTC) [280], Micro/Macro/Weighted Average Precision/Recall/F1 Score [268, 374]
	Local	2-sample Kullback-Leibler divergence [412], Manhattan distance [424], Canberra Distance [424], Euclidean distance [425], Matusita distance [425]
	Global	2-sample Kullback-Leibler divergence [412], Manhattan distance [424], Canberra Distance [424], Euclidean distance [425], Matusita distance [425], Earth Mover,Äôs Distance [426], Similarity metrics respectively based on entropy distance, spectral distance, modality distance, cosine of the angle between two graphs [427].
Efficiency	Data processing	Time delays compared with the real time data flow [428]
	Simulation/Modelling	Runtime of the simulation/modelling [34, 408]
Reproducibility	Same data	Yes/No
	Same statistics	2-sample Kullback-Leibler divergence [412]
	Same phenomena	ω-index [429], City organization index [430]

Based on the measures describing complex emerging patterns in a CNS model (Tab. 4.2), similarity between the created and the target CNS model patterns also ranges from the pinpoint level, the local level, to the global level.

The pinpoint level similarity is generally considered in prediction tasks, like the link prediction and its evaluation with precision [10] where the similarity is interpreted

in the context of the links predicted vs links actually appearing. In contrast, the local and the global similarity are generally employed in the simulation optimisation of CNS models toward the target network topological features.

Efficiency required for the CNS dynamics involves the discussion from two perspectives: the efficiency of data processing, concerned with the time delays compared with the real-time data flow [428], as well as the efficiency of modelling, connected to the runtime of CNS models [34, 408].

Reproducibility requires the equivalent results of the same tasks [431]. It varies with the increasingly demanding levels of equivalence in the recreated data, ranging from the same phenomena and statistic distribution to the same data. Current CNS models generally fall into the category of the same statistics. For example, the scale-free network model generates networks that share the same node degree distributions based on a given parameter set.

4.3.2.3 Faithfulness

Faithfulness of CNS representation and modelling is defined and evaluated based on complexity and model performance. Simulation-based, data-driven, and hybrid CNS models each poses different requirements of efficiency, reproducibility and similarity, which results in different meanings of faithfulness (Tab. 4.3).

Table 4.3: Requirement of a faithful CNS representation and modelling under observability constraints.

Requirement	Level	Simulation-based	Data-driven	Hybrid
Similarity	Pinpoint		●	○
	Local	●		○
	Global	●		○
Efficiency	Data processing		●	○
	Simulation/Modelling	●	●	●
Reproducibility	Same data		●	
	Same statistics	●		●
	Same phenomena	●		●

As is shown in Tab. 4.3, the solid and the hollow circle each represents the fulfilment or partial fulfilment of the corresponding requirement for each type of modelling. The

data-driven CNS models, as well as the data-driven components of hybrid CNS models, generally require pinpoint similarity with the real data and efficiency in data processing and modelling. The simulation-based CNS models and the simulation components of Hybrid CNS models need global and local similarity, as well as efficiency of a simulation process. In terms of reproducibility, data-driven CNS models, in the ideal case, are required to generate the same data, mimicking the reality. In contrast, for the simulation-based CNS models and the hybrid CNS models, the same statistics and phenomena are needed and sufficient. For example, some data-driven CNS models employ network embedding methods in node classification tasks [34, 285, 296]. They require a pinpoint level of similarity with ground-truth information about network topology and attributes, data processing and modelling efficiency, and reproducible classification results for validation and comparative analysis. The simulation-based CNS models, such as the scale-free network models, simulate realistic networks based on the assumptions of interaction rules – the preferential attachment to popular nodes [23, 109]. They require a global level similarity and the reproducibility of the statistics that measure the required network characteristics, such as the degree distributions reproduced by the same scale-free network model. Some hybrid CNS models employ real social networks and simulate the spreading processes on these networks [432]. They first represents the real networks with pinpoint level similarity and then, based on these networks, simulate spreading processes with reproducible spreading results.

4.3.3 Optimisation towards a Digital Twin

Under observability constraints, we optimise the DT-CNS models for an appropriate level of complexity, which enables a minimised information loss with a limited number of variables. We also create a set of modelling procedures to drive the optimisation of DT-CNS models considering different CNS model types and their required faithfulness given the level of observability (Fig. 4.1).

As shown in Figure 4.1, the devise and optimisation of the CNS model towards a DT go through the DT-CNS model preparation, DT-CNS representation and modelling, and the DT-CNS model evaluation procedures. We prepare a DT-CNS model by targeting a specific state of real systems and, under the observability constraints, deciding the CNS model type (section 4.3.1). After that, we determine the faithfulness needed for a DT-CNS model under the evaluation protocol proposed in section 4.3.2. In addition, we devise and optimise an extendable CNS model towards a DT through an appropriate selection of features (Tab. 4.4) and rules (Tab. 4.5).

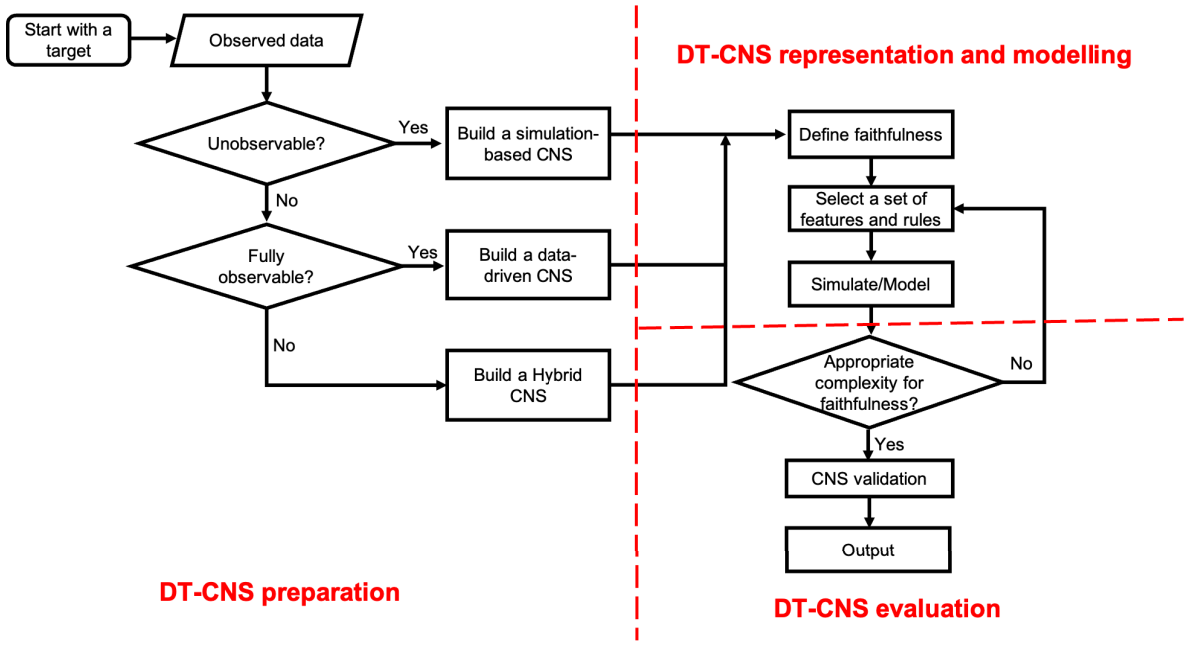


Figure 4.1: The procedures of building Digital Twin-Oriented Complex Networked Systems.

4.3.3.1 Features

Features in the DT-CNS model context represent the attributes and the topological characteristics of nodes and edges in CNS models. As is shown in Tab. 4.4, the network dimension has the ascribed and topological features, each resulting in and from the network growth. The process dimension focuses on the ascribed and the resulting conditions that change transmissibility and influence the spreading process. They are ascribed or resulting conditions dependent on whether they exist before or after the spread. Nodes with specific ascribed features can have different susceptibilities given an epidemic exposure, such as gender and age. The epidemic spread can also result in conditions such as the states of infected, immune and susceptible. These resulting conditions also change the susceptibility of nodes in the next round of epidemic spread.

4.3.3.2 Rules

Rules are used to govern the behaviours of nodes in the DT-CNS models and they can be prescribed or learned from data. Those rules drive the network growth and the propagation of the spreading process based on various CNS features. Rules that are considered in the framework are presented in Tab. 4.5. We parameterise the network growth with the social DNA vector based on preferential attachment and homophily

Table 4.4: Examples related to features and the conditions in the DT-CNS representation and modelling.

Features&Conditions				Data-driven	Simulation-based
Network	Node	Ascribed features	Continuous	Height, weight, etc.	Normal distribution, uniform distribution, etc.
			Discrete	Age, etc.	Poisson distribution, etc.
			Categorical	Gender, nationality, group, etc	Bernoulli distribution
		Topological features	Continuous	Node clustering coefficient, centrality, etc.	
			Discrete	Node degree, shortest path lengths, etc.	
	Edge	Ascribed features	Continuous	Intensity of contact/relationship, etc.	Normal distribution, uniform distribution, etc.
			Discrete		Poisson distribution, etc.
			Categorical	Relations, Types of contact, etc.	Bernoulli distribution
		Topological features	Continuous	Duration of contact/relationships, etc.	
			Discrete	Frequency of contact/Rating of relationships, etc.	
Process	Seed	Resulting conditions	Categorical	Infection status.	
	Spread	Ascribed conditions	Categorical	Immunity, susceptibility, etc.	
		Resulting conditions	Categorical	Infection status, an exposure, multiple exposure, etc.	

rules. In addition, we model the spreading process through seed selection and the process DNA vector for node adoptability. The experiments on DT-CNS models in this chapter investigate the network growth based on the interaction rules presented in Tab. 4.5. The seed selection strategy and the transmission rules in Tab. 4.5 are part of the design but will be developed in the future research.

Table 4.5: The examples of possible rules governing the interactions, seed selection and epidemic transmission in DT-CNS representation and modelling.

Rule		Parameters	Value range	Description	Terminology
Network	Preferential attachment	$\mathbf{p}(v_{i,t})$	$\{-1,0,1\}$	Preference for larger/smaller feature values	social DNA
		$\mathbf{w}^p(v_{i,t})$	$(0,1]$	Corresponding weight pf preference	
	Homophily	$\mathbf{h}(v_{i,t})$	$\{-1,0,1\}$	Preference for similar/dissimilar features	
		$\mathbf{w}^h(v_{i,t})$	$(0,1]$	Corresponding weight of preference	
Process	Node adoptability	$\Theta(v_{i,t})$	$\{0,1\}$	Adoptability threshold for each condition.	process DNA
		$\mathbf{w}^c(v_{i,t})$	$(0,1]$	The multiplier effect on transmissibility when meeting the threshold	
	Seed selection	$\mathbf{s}_t$	$\{-1,0,1\}$	seed preference for node features	seed strategy
		$\mathbf{w}_t^s$	$(0,1]$	Corresponding weight of preference	
		k_t	$(0,1]$	Number of seeds	

4.3.3.3 Trade-offs

Trade-offs between different requirements occur over the entire modelling and evaluation process to achieve highest possible faithfulness. These requirements involve the evaluation of CNS models with different measures, respectively, considering the three aspects: similarity, efficiency and reproducibility. However, constrained by the available

CNS measures and modelling techniques for multi-objective optimisation, we make trade-offs in the modelling and evaluation process (section 4.3.2) including:

- single-objective optimisation that combines different measures of one of these aspects: similarity, efficiency or reproducibility
- multi-objective optimisation considering all three aspects: similarity, efficiency and reproducibility
- minimised information loss for faithfulness and minimum number of variables under the constraint of observability

Current studies combine different similarity measures to calculate the loss function for an optimised CNS model [417]. Exceptionally, [54] proposes a comparative directed graph to compare the similarity of CNS models based on the sum of different ranks considering different similarity measures.

The integration and trade-offs between similarity, efficiency and reproducibility emerge as a research gap under the proposed evaluation criteria. The data-driven scenarios have focused on minimised information loss and a minimum number of variables for years, which involves the discussion on feature selection/extraction and model validation. However, the features in the context of DT-CNS model, simulation-based, data-driven or hybrid, differ between those scenarios due to the obseravability challenges and the simulation of informative and realistic features for DT-CNS models call for in-depth study in the future.

Finally, to validate the performance of a DT-CNS model under the evaluation criteria, data-driven CNS models, with more concerns related to the minimised information loss and the reproducible representation accuracy, can employ the cross-validation approaches through data split and repeated experiments. In contrast, for the simulation-based and hybrid CNS models with partial observability, CNS ensembles that create a series of CNS representations with quantifiable, reproducible and realistic characteristicsare expected to prove the validity of DT-CNS models.

4.4 Results

In this section, we conduct experiments on simulation-based DT-CNS models to investigate the disaster resilience of different social networks considering the infection occurrences on these networks in an epidemic outbreak. We conduct a comparative

analysis of simulation-based DT-CNS models driven by selected rules (preferential attachment and homophily) and distributions of the ascribed feature – age.We take the degree distribution of a scale-free network (composed of 90 nodes and 1400 edges, built with Barabasi-Albert network model [45]) as the target for the simulation of the static networks. To simulate the social networks, we assume 90 nodes, 1400 edges and an encounter rate of 0.8 for the respective node pairs. A simulation-based dynamic process takes place on the networks without changing its parameters over the temporal scale. We evaluate these models in terms of complexity and their faithfulness, considering the global level similarity (measured by the difference between degree distributions), simulation efficiency, and reproducibility of the same statistics.

We respectively term the involved simulation-based DT-CNS models as $DT-CNS_U$, $DT-CNS_B$, $DT-CNS_I$, $DT-CNS_L$, $DT-CNS_R$ and $DT-CNS^{H+}$, $DT-CNS^{H-}$ $DT-CNS^{P+}$, $DT-CNS^{P-}$ and $DT-CNS^{HP}$, to characterise the DT-CNS modelling paradigms with uniform (U), bell (B), inverse bell (I), left-skewed (L) and right-skewed (R) age distributions, respectively driven by preferential attachment to larger feature values ($P+$), preferential attachment to smaller feature values ($P-$), heterophily ($H+$), homophily ($H-$), and rules that optimally incorporate the above principles for a target degree distribution.

4.4.1 Simulation-based Networks

In this section, we build the simulation-based networks with 90 nodes, 1400 edges and a single feature – age, ascribed to each node. We presume that the age involved in disaster resilience range between $[0,90]$ while categorising the age groups every ten years [433].

4.4.1.1 Feature

We simulate the uniform, the bell, the inverse bell, the left-skewed and the right-skewed shape distributions for a fixed number of nodes allocated in each age group. We also introduce uncertainty through a randomly generated age value within the range required by each age group (Fig. 4.2).

In Fig. 4.2, the numbers of nodes in each age group vary depending on distribution. There are 10 nodes evenly allocated for each age group in a uniform distribution. In contrast, more/fewer nodes are in their 40s in a bell/inverse bell distribution than in other age groups. In a left-skewed distribution, most nodes are over 60, and vice versa for a right-skewed distribution. We evaluate the age diversity with the Hill number,

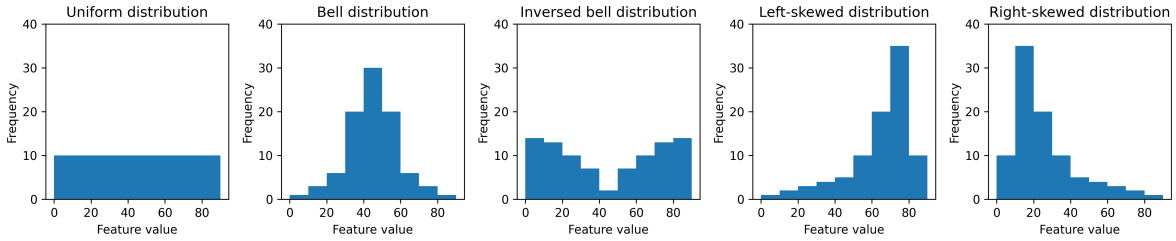


Figure 4.2: The age distributions used in the experiments.

which measures the effective number of equally abundant species (age groups) [434, 435]. The varying diversity levels differentiate the respective age distributions, indicating the population's age differences and resulting in different connection patterns.

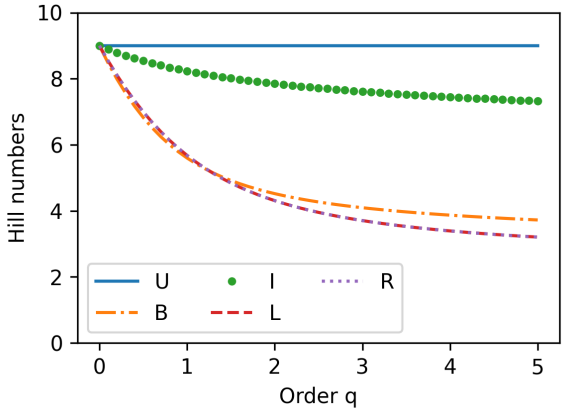


Figure 4.3: The Hill number for each age distribution.

In Fig. 4.3, we calculate the Hill numbers with an order value that determines the sensitivity to the relative frequencies of the species (age groups), valuing between 0 and 9. A higher order value indicates a higher sensitivity level to the relative number of nodes allocated in each age group. With an order value of 0, the Hill numbers get insensitive to the abundance of the respective species (age groups) and keep at 9 the number of age groups for all the age distributions [432]. With the same order value, the higher Hill number indicates higher diversity. As the age distributions transform from the uniform, inverse bell, bell to skewed age distributions, their Hill numbers decrease, indicating less age diversity.

4.4.1.2 Preference

Given the age feature and the different rules of network formation, including the preferential attachment to larger feature values (P+), the preferential attachment to smaller feature values (P-), the heterophily (H+), the homophily (H-) as well as the optimised combination of both the preferential attachment and the homophily principles (PH), the scores of node pairs vary a lot and indicate different strengths of relationships. The parameter set-ups of preferences and weights of preferences is included in Tab. B.1 in the appendix B.1.

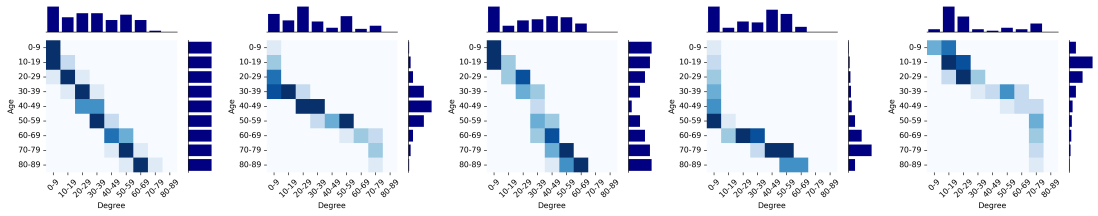
4.4.1.3 Network Topology

Given the same threshold factor $\lambda_e = 1400$, that determines the number of edges, the simulated networks are correspondingly characterised with varying network topology (Fig. 4.4, Fig. 4.5 and Fig. 4.6).

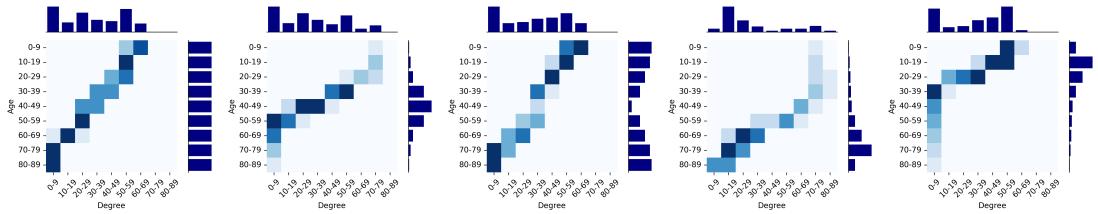
Degree Distribution is a distribution of node degrees in a given network [274]. As shown in Tab. 4.6, the node degree of the target network fluctuates around the average value of 31.11 with a standard deviation of 12.90, ranging from 6.00 to 70.00.

Fig. 4.4 shows age and degree distributions with a heatmap and two histogram plots for age (Fig. 4.2) and node degree (Tab. 4.6). We find that the node degrees of each age group varies with the interaction rules related to preferences for age. For example, the $DT{-}CNS^{P+}$ models, built on positive preferential attachment to age, tend to have more older nodes connected with each other and those older nodes have higher node degrees than in case of other models. Similarly, the $DT{-}CNS^{H-}$ models built on homophily when it comes to age wire more nodes within the same age groups, enabling the nodes in age groups with more members to have higher node degrees.

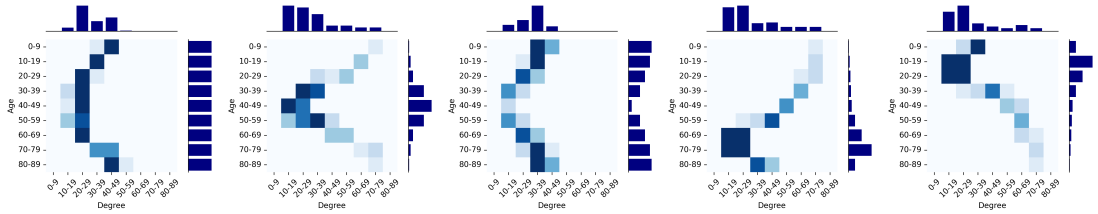
Given the same interaction rules, the node degrees of respective age groups also vary with the shapes of age distributions. The symmetric and asymmetric shape distributions can lead to a different trend of node degree changes with age change. For example, for $DT{-}CNS_B^{H+}$ based on the $H+$ (heterophily) rule and the bell-shape age distribution, the node degree first decreases from $[70-79]$ to $[10-19]$ when the node age transits from $[0-9]$ to $[40-49]$, which then increases from $[10-19]$ to $[70-79]$ as the node age transits from $[50-59]$ to $[60-89]$. This contrasts with the case of $DT{-}CNS_L^{H+}$, where the node degree first decreases from $[70-79]$ to $[10-19]$ when the node age transits from $[0-9]$ to $[60-69]$, which then increases from $[10-19]$ to $[40-49]$ as the node age transits from $[70-79]$ to $[80-89]$. This phenomenon can be caused by the number of nodes in each age



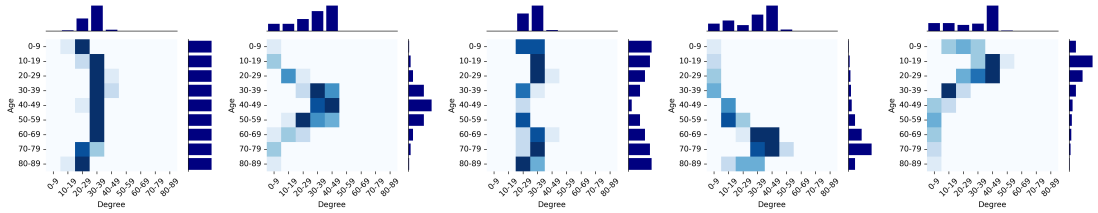
(a) $DT-CNS_U^{P+}$ (b) $DT-CNS_B^{P+}$ (c) $DT-CNS_I^{P+}$ (d) $DT-CNS_L^{P+}$ (e) $DT-CNS_R^{P+}$



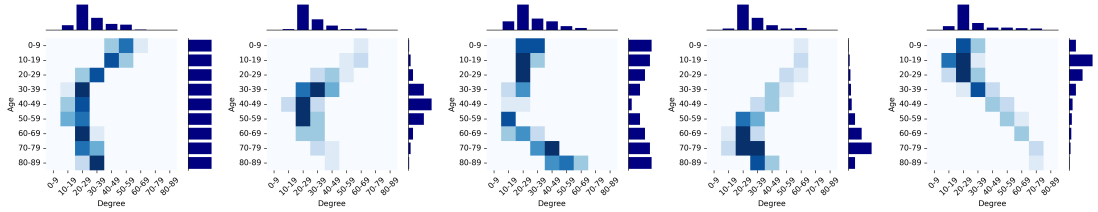
(f) $DT-CNS_U^{P-}$ (g) $DT-CNS_B^{P-}$ (h) $DT-CNS_I^{P-}$ (i) $DT-CNS_L^{P-}$ (j) $DT-CNS_R^{P-}$



(k) $DT-CNS_U^{H+}$ (l) $DT-CNS_B^{H+}$ (m) $DT-CNS_I^{H+}$ (n) $DT-CNS_L^{H+}$ (o) $DT-CNS_R^{H+}$



(p) $DT-CNS_U^{H-}$ (q) $DT-CNS_B^{H-}$ (r) $DT-CNS_I^{H-}$ (s) $DT-CNS_L^{H-}$ (t) $DT-CNS_R^{H-}$



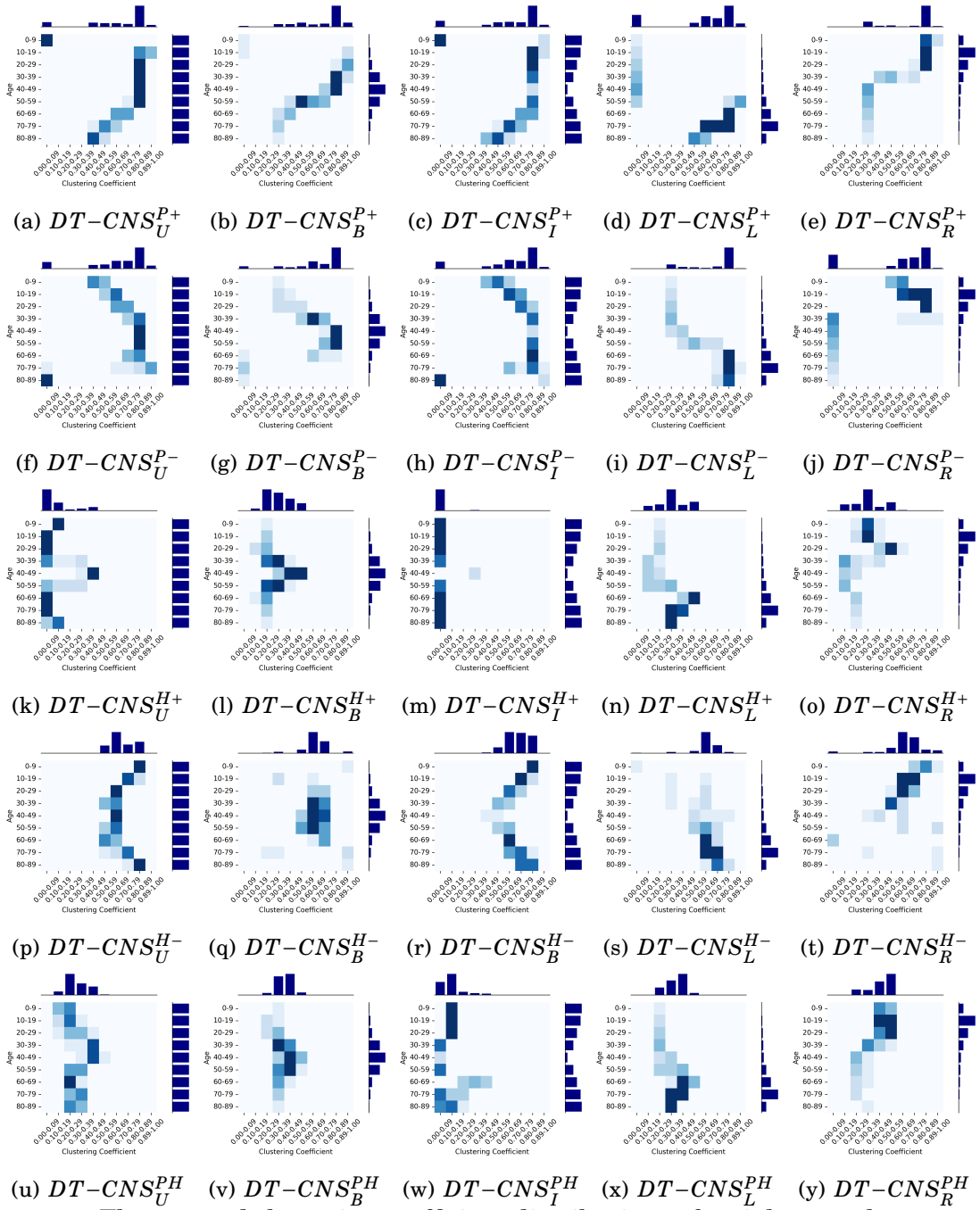
(u) $DT-CNS_U^{PH}$ (v) $DT-CNS_B^{PH}$ (w) $DT-CNS_I^{PH}$ (x) $DT-CNS_L^{PH}$ (y) $DT-CNS_R^{PH}$

Figure 4.4: The age and degree distributions of social networks generated by DT-CNS models based different features and rules.

group, as nodes in sparse age groups are preferred and connected with those dissimilar others in dense age groups. This inevitably leads to higher node degrees for these sparse age groups. In contrast, the dense age groups have lower node degrees as they prefer the sparse age groups and have limit connections with these preferred nodes.In addition, when we optimise the combined preferences considering both preferential attachment and homophily ($DT-CNS^{PH}$ models), the shapes of degree distributions approach the shape of a power-law distribution. This indicates that the introduction of features and the optimisation of corresponding preferences enable better achievement of the target states of networks.

Clustering Coefficient describes the probability of a node's neighbours to be connected. Its value is between 0 and 1 [274]. As shown in Tab. 4.6, in the target network, clustering coefficient fluctuates around an average value of 0.47 with a standard deviation of 0.08, ranging from 0.79 to 0.33.

Fig. 4.5 shows age and clustering coefficient distributions with a heatmap and two histogram plots for age (Fig. 4.2) and clustering coefficient (Tab. 4.6). We find that nodes in age groups with bigger number of people with connection preferences for a similar group of nodes tend to cluster with a higher clustering coefficient. For example, the young and middle-aged nodes in the $DT-CNS^{P+}$ models, built on positive preferential attachment to age, prefer to be connected with old nodes, where these nodes cluster around the popular old nodes.Similarly, the nodes in $DT-CNS^{H-}$ models, where interactions are based on homophily phenomenon, like to be connected with similar others. This generally leads to a higher clustering coefficient than in the case of other DT-CNS models. However, nodes in $DT-CNS^{H+}$ models, driven by heterophily phenomenon, prefer to be connected with dissimilar others. They generally have lower clustering coefficients than in the case of other DT-CNS models because they select different nodes to connect and do not cluster around the same node. When we optimise the combined preferences considering both preferential attachment and homophily, the shapes of clustering coefficient distributions vary with age distributions. As nodes tend to connect with similar and old nodes (referring to the optimised preferences for both preferential attachment and homophily in Tab. B.1 in the appendix B.1), the old and dense age groups tend to have higher clustering coefficient. Among them, the $DT-CNS_U^{PH}$ and the $DT-CNS_I^{PH}$ models characterised with more even age group allocations generate a clustering coefficient with a power-law shape because less nodes fulfil one of the requirements for larger age values and denser age groups, which induces a lower clustering coefficient in the respective cases.



(a) $DT-CNS_U^{P+}$ (b) $DT-CNS_B^{P+}$ (c) $DT-CNS_I^{P+}$ (d) $DT-CNS_L^{P+}$ (e) $DT-CNS_R^{P+}$

(f) $DT-CNS_U^{P-}$ (g) $DT-CNS_B^{P-}$ (h) $DT-CNS_I^{P-}$ (i) $DT-CNS_L^{P-}$ (j) $DT-CNS_R^{P-}$

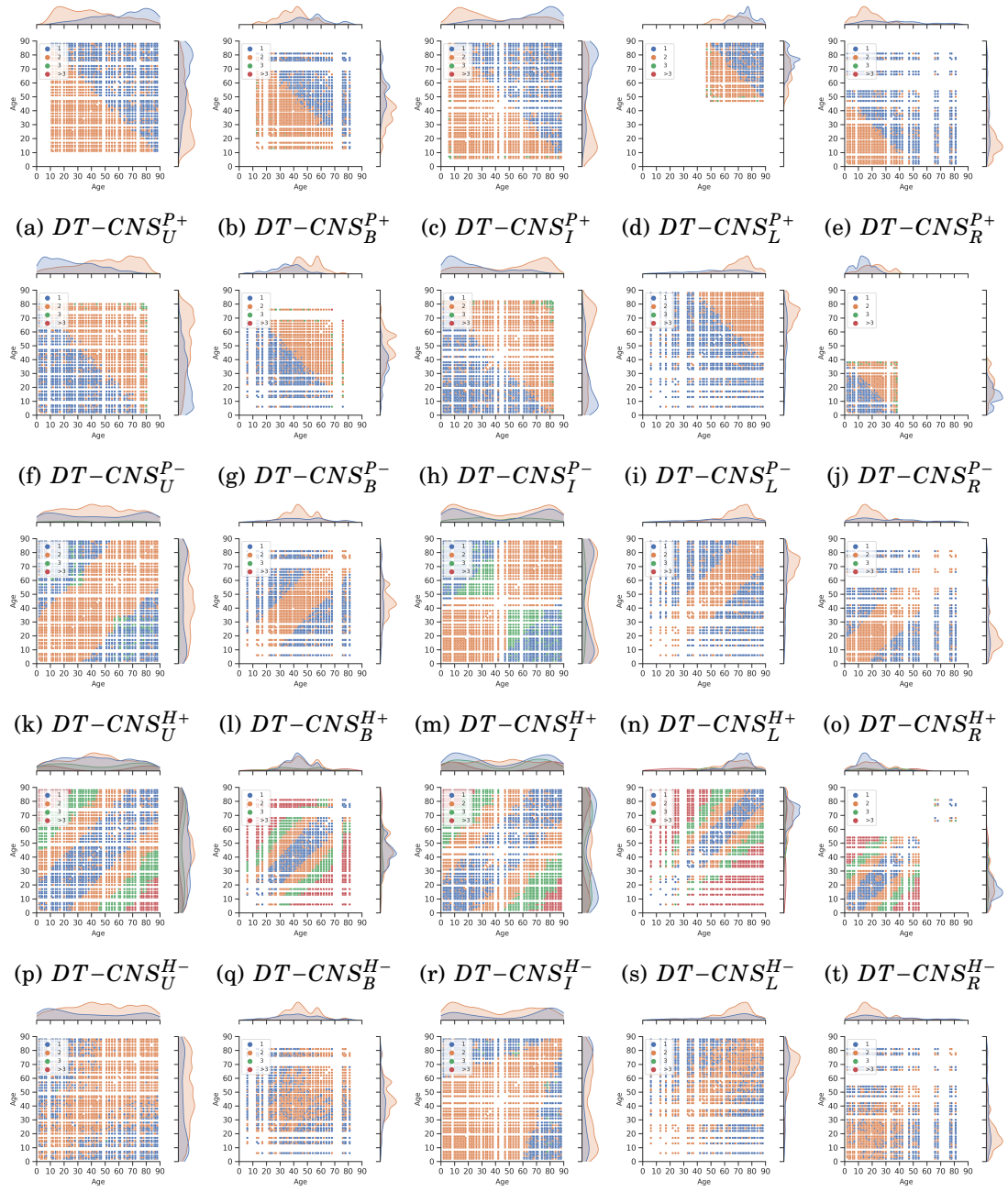
(k) $DT-CNS_U^{H+}$ (l) $DT-CNS_B^{H+}$ (m) $DT-CNS_I^{H+}$ (n) $DT-CNS_L^{H+}$ (o) $DT-CNS_R^{H+}$

(p) $DT-CNS_U^{H-}$ (q) $DT-CNS_B^{H-}$ (r) $DT-CNS_B^{H-}$ (s) $DT-CNS_L^{H-}$ (t) $DT-CNS_R^{H-}$

(u) $DT-CNS_U^{PH}$ (v) $DT-CNS_B^{PH}$ (w) $DT-CNS_I^{PH}$ (x) $DT-CNS_L^{PH}$ (y) $DT-CNS_R^{PH}$

Figure 4.5: The age and clustering coefficient distributions of social networks generated by DT-CNS models based different features and rules

Shortest Path Length between two nodes describes the number of edges along the shortest path between a pair of nodes [274]. As shown in Tab. 4.6, the shortest path length in the target network fluctuates around an average value of 1,65 with a standard deviation of 0.48, ranging from 1.00 to 3.00.



(a) $DT{-}CNS_U^{P+}$ (b) $DT{-}CNS_B^{P+}$ (c) $DT{-}CNS_I^{P+}$ (d) $DT{-}CNS_L^{P+}$ (e) $DT{-}CNS_R^{P+}$

(f) $DT{-}CNS_U^{P-}$ (g) $DT{-}CNS_B^{P-}$ (h) $DT{-}CNS_I^{P-}$ (i) $DT{-}CNS_L^{P-}$ (j) $DT{-}CNS_R^{P-}$

(k) $DT{-}CNS_U^{H+}$ (l) $DT{-}CNS_B^{H+}$ (m) $DT{-}CNS_I^{H+}$ (n) $DT{-}CNS_L^{H+}$ (o) $DT{-}CNS_R^{H+}$

(p) $DT{-}CNS_U^{H-}$ (q) $DT{-}CNS_B^{H-}$ (r) $DT{-}CNS_I^{H-}$ (s) $DT{-}CNS_L^{H-}$ (t) $DT{-}CNS_R^{H-}$

(u) $DT{-}CNS_U^{PH}$ (v) $DT{-}CNS_B^{PH}$ (w) $DT{-}CNS_I^{PH}$ (x) $DT{-}CNS_L^{PH}$ (y) $DT{-}CNS_R^{PH}$

Figure 4.6: The age and shortest path length distributions between connected nodes of social networks based on different features and rules

Fig. 4.6 shows age and shortest path length distributions with a scatter plot and two same Kernel Density Estimation (KDE) plots for shortest path length referring to the age values (Fig. 4.2 for age distributions and Tab. 4.6 for information related to the shortest path length). Specifically, in Fig. 4.6, we present the shortest path length at 1, 2, 3 and over 3 between the connected nodes with blue, orange, green and red dots. The x-axis and y-axis represent age values and the scattered dots represent the shortest path lengths between given pair of nodes of given ages. We find that the shortest path lengths between the nodes are generated by all the $DT-CNS$ models range between 1 and 3 except for the $DT-CNS^{H-}$ models built on the $H-$ (homophily) rule. This is because in $DT-CNS^{H-}$ models, similar nodes cluster within the similar age groups, leading to longer paths to connect the dissimilar nodes. For example, in the $DT-CNS_L^{H-}$ model, which is built with a left-skewed age distribution and the homophily principle, a large number of paths between $0-9$ and $80-89$ age groups have value over 3 as a high proportion of young nodes in the left-skewed age distribution cluster together without many direct connections with the nodes for which the age range is $89-90$. In addition, the distribution of the shortest path lengths is also highly related to the preference principles. The age groups with preferred features have shorter paths to other nodes. For example, the nodes which are old in $DT-CNS^{P+}$ models have more direct connections with others and the shortest paths of length of 1 due to the fact that they have preferred value of feacture age. When we optimise the combined preferences considering both preferential attachment and homophily, due to the complex interaction rules, the shapes of shortest path lengths distributions vary greatly with age distributions and become very hard to interpret. This indicates that the increasing dynamics level depreciates the network patterns' interpretability but enables the creation of more complex network and more faithful scenarios.

Summary Statistics, related to the target network and the simulated networks, are presented in Tab. 4.6. It incorporates information related to the degree distributions, clustering coefficient distributions and shortest path length distributions. In addition, we use JS divergence to evaluate the distances between these network patterns with that of the target network (Tab. B.2 in the appendix B.2). More specifically, the most similar network patterns are identified with green colour. Compared with the target network, the network simulated by the $DT-CNS_U^{PH}$ model generates the most similar degree distribution and shortest path length distribution. $DT-CNS_L^{PH}$ model generates the most similar clustering coefficient distribution. This indicates that network models,

driven by different features and the optimised interaction rules, approach different characteristics of the target network. These differences challenge the evaluation and development of DT-CNS models, remaining a research gap to be addressed in future studies. However, in our experiments, we only focus on the degree distribution of target network and investigate how heterogeneous features and rules influence the performance of DTCNS models in recreating similar degree distributions.

Table 4.6: Topological information of the networks generated with the DT-CNS models.

Features	Rules	Nodes		Node Degree				Clustering coefficient				Shortest path length				
		Connected	Unconnected	Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
Target Network		90	0	31.11	12.90	70	6	0.47	0.08	0.79	0.33	0	1.65	0.48	3	1
Uniform	$P+$	80	10	31.11	21.20	70	0	0.66	0.27	1.00	0.00	845	20.22	36.09	90	1
	$P-$	81	9	31.11	21.48	68	0	0.67	0.28	1.00	0.00	765	18.47	34.76	90	1
	$H+$	90	0	31.11	9.64	52	16	0.12	0.14	0.45	0.00	0	1.69	0.54	3	1
	$H-$	90	0	31.11	6.16	43	19	0.69	0.09	0.89	0.57	0	2.05	0.98	5	1
	PH	90	0	31.11	11.56	63	13	0.31	0.09	0.52	0.16	0	1.65	0.48	3	1
Bell	$P+$	89	1	31.11	21.34	77	0	0.73	0.21	1.00	0.00	89	3.61	13.03	90	1
	$P-$	87	3	31.11	21.50	75	0	0.70	0.22	0.89	0.00	264	7.46	21.93	90	1
	$H+$	90	0	31.11	15.70	73	13	0.36	0.11	0.57	0.16	0.00	1.65	0.48	2	1
	$H-$	90	0	31.11	12.85	47	2	0.68	0.10	1.00	0.30	0	2.12	1.16	8	1
	PH	90	0	31.11	10.32	68	17	0.40	0.06	0.52	0.28	0	1.65	0.48	2	1
Inverse bell	$P+$	79	11	31.11	21.24	66	0	0.66	0.28	1.00	0.00	924	21.96	37.26	90	1
	$P-$	80	10	31.11	21.54	67	0	0.66	0.29	1.00	0.00	845	20.23	36.08	90	1
	$H+$	90	0	31.11	8.03	42	12	0.01	0.05	0.36	0.00	0	1.79	0.67	3	1
	$H-$	90	0	31.11	4.71	40	22	0.72	0.10	0.88	0.48	0	2.18	1.07	5	1
	PH	90	0	31.11	12.63	63	10	0.14	0.10	0.47	0.00	0	1.66	0.5	3	1
Left skewed	$P+$	75	15	31.11	21.34	64	0	0.61	0.32	1.00	0.00	1230	28.69	40.82	90	1
	$P-$	90	0	31.11	22.47	81	7	0.74	0.19	0.90	0.30	0	1.65	0.48	2	1
	$H+$	90	0	31.11	16.80	73	14	0.37	0.11	0.54	0.15	0	1.65	0.48	2	1
	$H-$	90	0	31.11	14.50	50	1	0.66	0.11	0.83	0.00	0	2.42	1.67	9	1
	PH	90	0	31.11	11.25	67	15	0.39	0.08	0.53	0.23	0	1.65	0.48	2	1
Right skewed	$P+$	90	0	31.11	22.41	79	7	0.73	0.2	0.93	0.31	0	1.65	0.48	2	1
	$P-$	73	17	31.11	21.69	62	0	0.59	0.34	1.00	0.00	1377	31.92	42.04	90	1
	$H+$	90	0	31.11	17.07	74	15	0.36	0.12	0.6	0.14	0	1.65	0.48	2	1
	$H-$	90	0	31.11	15.48	51	1	0.67	0.17	1.00	0.00	504	13.06	29.21	90	1
	PH	90	0	31.11	14.24	74	14	0.46	0.09	0.58	0.25	0	1.65	0.48	2	1

Tab. 4.6 shows the number of connected and unconnected nodes and the information about the degree distributions, clustering coefficient distributions and the shortest path length distributions that vary depending on the values of the features and rules employed. Generally, all the nodes get connected with the $DT-CNS^{H+}$, $DT-CNS^{H-}$ and $DT-CNS^{PH}$ paradigms. $DT-CNS^{P+}$ or $DT-CNS^{P-}$ paradigms, driven by preferential attachment to old/young ages, result in some unconnected old/young nodes given any age distribution due to the strong preference for young/old ages. In contrast, $DT-CNS^{H+}$, $DT-CNS^{H-}$ and $DT-CNS^{PH}$ paradigms consider the effect of similar or dissimilar

ages, where all nodes can find similar/dissimilar others within/across age groups and connect with them. The node degree distributions for each DT-CNS model paradigm share the same mean values but have different standard deviation, maximum and minimum values. Degree distributions, generated by $DT{-}CNS^{H^-}$ and $DT{-}CNS^{PH}$ paradigms, generally have smaller standard deviations and maximum values. In contrast, $DT{-}CNS^{P^+}$ and $DT{-}CNS^{P^-}$ paradigms have larger ones. The maximum degree of the respective paradigms are similar with the number of connected nodes, which indicates that the most popular node tends to directly connect with most of the connected nodes (resulting in power law node degree distribution). The node clustering coefficient distributions for $DT{-}CNS^{P^+}$ and $DT{-}CNS^{P^-}$ models based on preferential attachment principles and $DT{-}CNS^{H^-}$ models built on homophily generally have larger average values, standard deviations and maximum values than other models, which results from the dense clusters created by preferred nodes (Fig. 4.5). The shortest path length distributions are especially influenced by the number of unconnected nodes and the corresponding non-existing ('fake') paths, which are assumed as the maximum path length plus one (i.e. 90). For the fully connected networks, the average shortest path length fluctuates around 2, given different age distributions and preferences. It increases significantly when there are fake paths for unconnected node pairs.

4.4.2 Simulation-based Processes

In this section, we build simulation-based processes to analyse disaster resilience given different network structures under the influence of various age distributions. The susceptible nodes get infected given exposure with fixed transmissibility rate, while infected nodes get infectious over time without recovery. The epidemic spread on the networks propagates one step (edge) away for each time step.

4.4.2.1 Seed Selection

We investigate the most severe case of epidemic explosion by assuming a single seed selection in the initial stage based on the largest node degree (Tab. 4.7) and allow the epidemic propagation for 5 time steps within 6 steps (edges) away from the seed (first infection). Within this time and distance range, most of the connected nodes finally get infected as they are directly/indirectly connected with the seed. As shown in Tab. 4.6 and Tab. 4.7, almost all the connected nodes are connected with the seed node, resulting in a higher infection risk and the full infection.

Table 4.7: The single seed selection based on the largest node degree.

Features	Rules	Seed	
		Node degree	Age group
Uniform	$P+$	70	80-89
	$P-$	68	0-9
	$H+$	52	80-89
	$H-$	43	20-29
	PH	63	0-9
Bell	$P+$	77	70-79
	$P-$	75	10-19
	$H+$	73	70-79
	$H-$	47	40-49
	PH	68	0-9
Inverse bell	$P+$	66	80-89
	$P-$	67	0-9
	$H+$	42	80-89
	$H-$	40	20-29
	PH	63	80-89
Left skewed	$P+$	64	80-89
	$P-$	81	20-29
	$H+$	73	20-29
	$H-$	50	70-79
	PH	67	0-9
Right skewed	$P+$	79	50-59
	$P-$	62	0-9
	$H+$	74	70-79
	$H-$	51	10-19
	PH	74	70-79

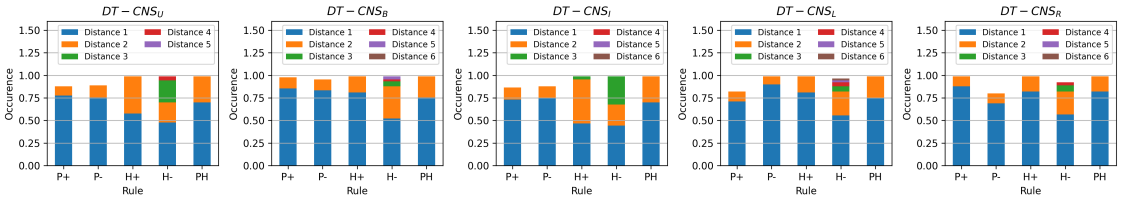
As shown in Tab. 4.7, the seeds selected for all the modelling paradigms generally fall in an age group which are preferred by much denser age groups. For example, the old (young) nodes are preferred by all other nodes in $DT-CNS^{P+}$ paradigms ($DT-CNS^{P-}$ paradigms) and thus have the largest number of connections, which almost directly connect all the connected nodes in the respective model (Tab. 4.6).

4.4.2.2 Infection Occurrence

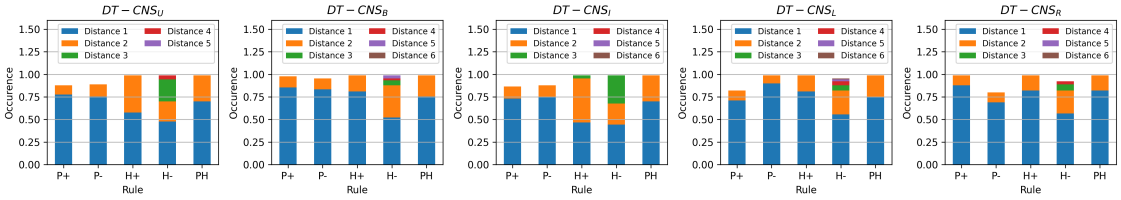
The infection occurrence refers to the proportion of infected nodes to the entire population within a specific distance from the seed. Given the infectious seed, we simulate an epidemic spreading process based on different transmissibilities (ranging in $[0.2, 0.4, 0.6, 0.8]$) and investigate their impact on the infection occurrence.

In Fig. 4.7, we represent the number of infection occurrences within specific number of steps (edges away) from the first infection. The vertical axis and the horizontal axis each represents the rules of network formation and the infection occurrence among the population (90 nodes in total). As we simulate the epidemic spread for 6 time steps and allow the epidemic propagation by one step (edge) for each time step, nodes within 6 steps away from the first infection are likely to be infected and thus have infection risks in the epidemic simulations. Nodes out of this range will not be exposed to the infection risks and thus some of the nodes stay uninfected despite the increase of transmissibility. The infection occurrence generally stays the same given any transmissibility between 0.4 and 1.0. Given different rules, the $DT-CNS^{P+}$ and the $DT-CNS^{P-}$ paradigms can achieve the biggest number of infections within a distance of 1 (one edge away) from the seed; $DT-CNS^{H+}$ and $DT-CNS^{PH}$ reach the upper limit of 90 within a distance of 2; $DT-CNS^{H-}$ gets the biggest number of infections within a distance of around 6. This indicates that the networks, driven by $P+$ or $P-$ rule (preferential attachment to old/young ages), reach the maximum infection occurrence within the first time step as most nodes are directly connected with the seed and can get infected. Given various age distributions, $DT-CNS_I$ has the fewest infections within the distance of 1, indicating the less explosive effect of the epidemic outbreak from the seed. The above mentioned phenomenon suggests that the isolation policies imposed on nodes 1 or 2 steps away from the seed can efficiently reduce infection occurrence especially in networks driven by the $H-$ rule, which describes the nodes' preferences for similar features.

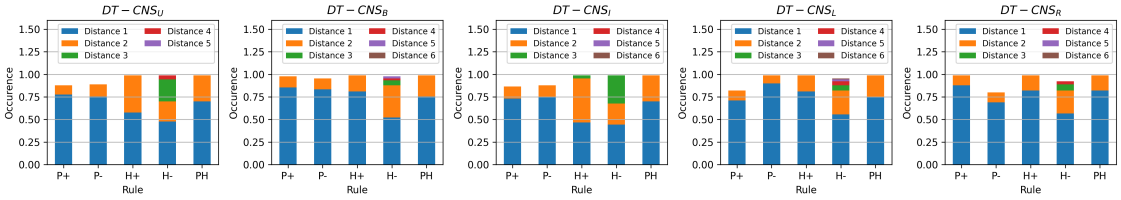
With decreasing transmissibility of the epidemic spread given the exposure, it takes more time steps to achieve the infection status shown in Fig. 4.7. In Fig. 4.8, the number of infected nodes within the first time step (represented as Period 1) gets smaller when



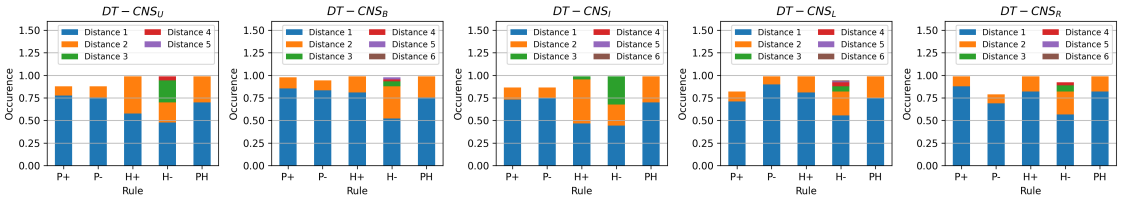
(a) 1.0 transmissibility given an exposure



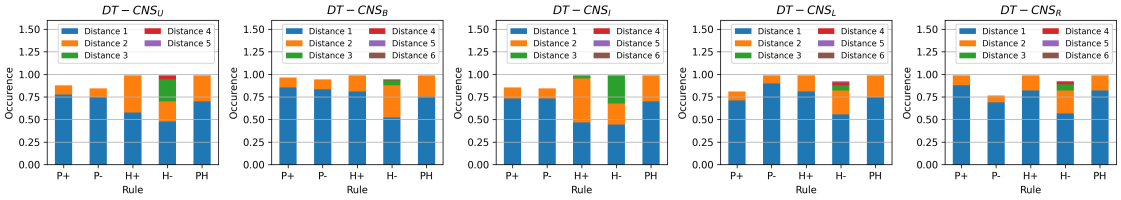
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure

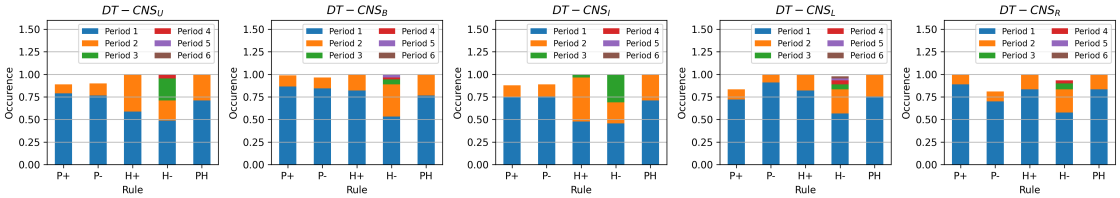


(d) 0.4 transmissibility given an exposure

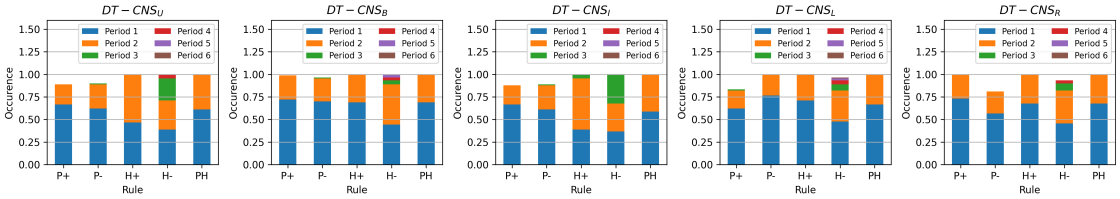


(e) 0.2 transmissibility given an exposure

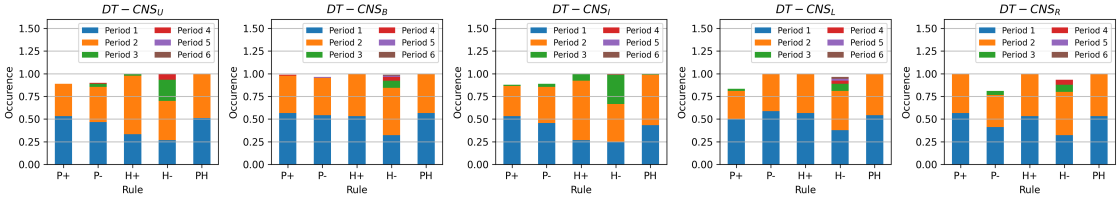
Figure 4.7: The infection occurrence within certain distance to the seed given different transmissibilities.



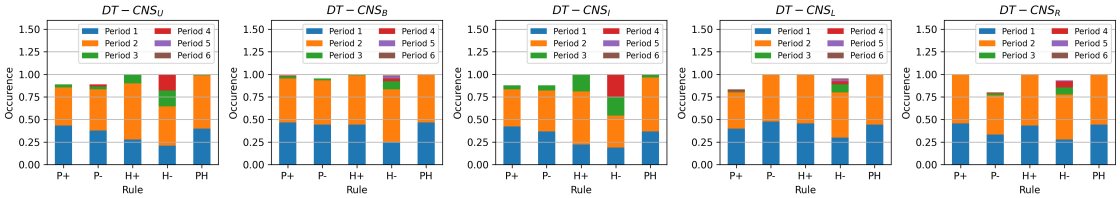
(a) 1.0 transmissibility given an exposure



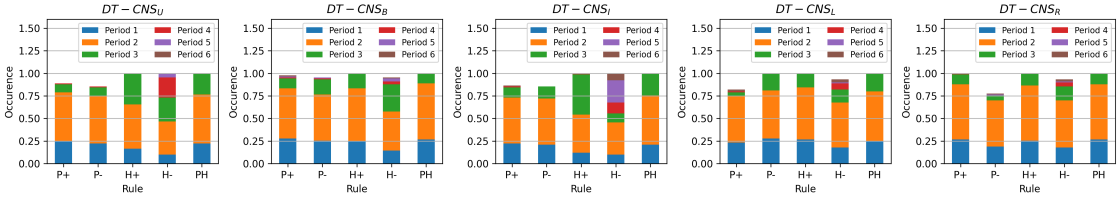
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure



(d) 0.4 transmissibility given an exposure



(e) 0.2 transmissibility given an exposure

Figure 4.8: The infection occurrence over time with different transmissibilities given an exposure.

the transmissibility decreases from 1.0 to 0.2, while the infections increase within Period 2 and Period 3. Given a smaller transmissibility, the nodes are less likely to get infected in Period 1 and thus stay healthy. However, these healthy nodes in Period 2 and Period 3 get exposed to more infected neighbours and finally get infected, leading to a delayed increase in infection numbers. The decrease of transmissibility results in increase of time needed to infect all the nodes and in the same time gives more time to react to the disaster. Compared with the other models, DT–CNS^{H-} models, driven by the homophily principle, have smaller infection occurrences despite different transmissibilities. In addition, the infection occurrences for $DT-CNS^{H-}$ models also increase slowly compared to other models. This is because the seed node clusters around a limited number of similar nodes (Tab. 4.7). This leads to a social distancing between the infectious seed and dissimilar nodes. The increase in transmissibility does not bridge the social distance between the seed node and the dissimilar nodes. Therefore, for DT–CNS^{H-} models, it still takes more time for the epidemic spread to reach and infect these nodes. For infection occurrences in various age distributions over different time steps, $DT-CNS_I$ generally has the fewest infections within the first time step. This also results from limited connections between the seed node and others (Tab. 4.7). The abovementioned phenomenon indicates that the increase in transmissibility can have a greater impact on networks that densely clusters around the seed node.

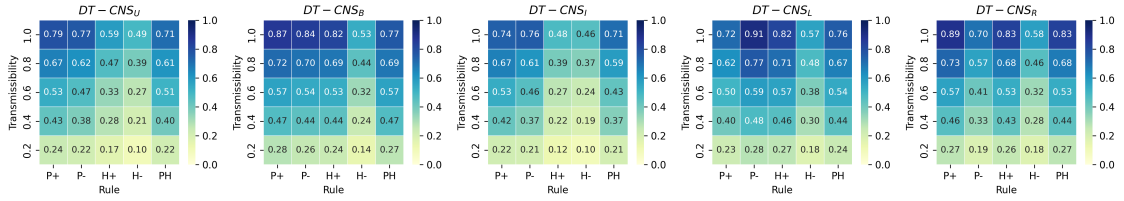
4.4.2.3 People at Risk

We introduce the People at Risk (PaR) measure to have an intuitive representation of the proportion of people at risk of a disaster:

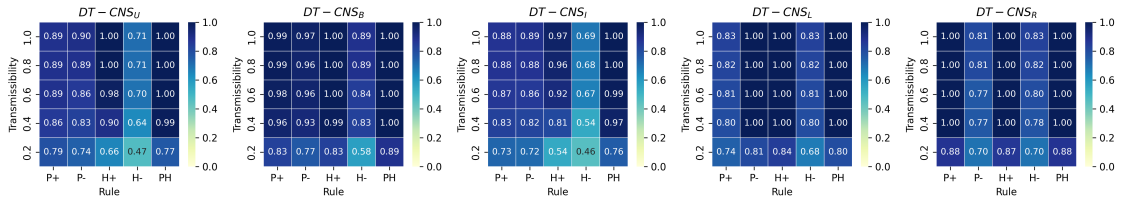
$$PaR(T,D) = \frac{\sum\limits_{v_{i,t} \in V_t} \mathbf{r}(v_{i,t}) \delta_{\mathrm{l}(v_{i,t},\mathbf{s}_t),D} \delta_{t,T}}{N}, \quad D \leq T \tag{4.10}$$

which is defined as $PaR(T,D)$ to help identify the easiest-to-be-infected proportion of the population in the limited time $[0,T]$ and the space within the distance $[0,D]$ away from the seed. $\delta_{t,T}$ and $\delta_{\mathrm{l}(v_{i,t},\mathbf{s}_t),D}$ each represents the Kronecker functions related to time t and the distance $\mathrm{l}(v_{i,t},\mathbf{s}_t)$ between node $v_{i,t}$ and seed $\mathbf{s}_t$. N represents the number of nodes.

Fig. 4.9 shows the $PaR(1,1)$ and the $PaR(2,2)$ values that vary with transmissibilities, rules of network formation and the age feature distributions. First, the $PaR(1,1)$ decreases as transmissibility decreases. $DT-CNS^{P+}$, $DT-CNS^{P-}$, $DT-CNS^{H+}$ and $DT-CNS^{PH}$ generally have a higher $PaR(1,1)$, and this is due to the higher number



(a) $PaR(1,1)$



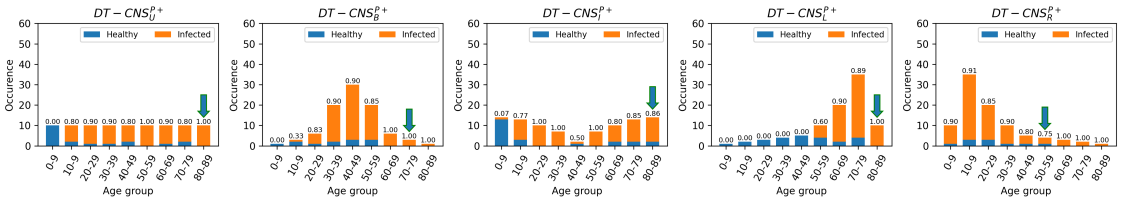
(b) $PaR(2,2)$

Figure 4.9: The $PaR(1,1)$ and $PaR(2,2)$ values with different transmissibilities, age distributions and rules of network formation.

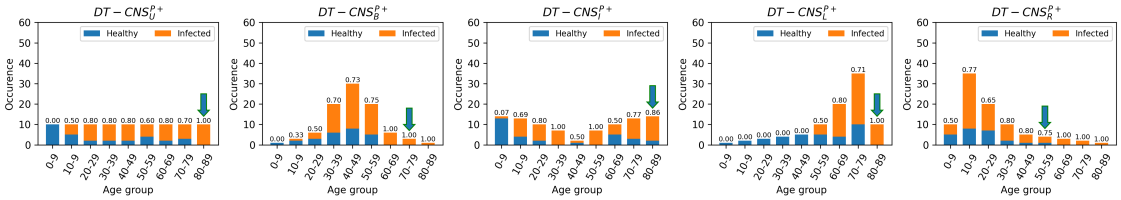
of nodes connected to the seed (Fig. 4.7, where there are more occurrences of infections within the distance 1). In addition, $DT-CNS^U$ and $DT-CNS^I$, due to a higher age diversity, have smaller $PaR(1,1)$ values. The $PaR(2,2)$ values increase significantly as the transmission propagates one step further. Almost all the connected nodes get infected within two time steps, despite the differences in transmissibilities (Fig. 4.8). This can be caused by the significant number of direct connections between the popular seed node and others (Tab. 4.6). Similar with the case of $PaR(1,1)$ values, the $PaR(2,2)$ values for the $DT-CNS^{H-}$ models are also smaller than for the other models, resulting from the limited number of connections with the seed node (Tab. 4.6). The above-mentioned differences between $PaR(1,1)$ and $PaR(22)$ values indicate the necessity of epidemic control in primary time and distance to the seed node.

To better understand the diversity of age features and their influence on the infection status, we respectively calculate the $PaR(1,1)$ of each age group based on different transmissibilities, age groups and rules (Fig. 4.10, Fig. 4.11, Fig. 4.12, Fig. 4.13 and Fig. 4.14). We also identify the age group where the epidemic starts with a green arrow in the corresponding $PaR(1,1)$ figures.

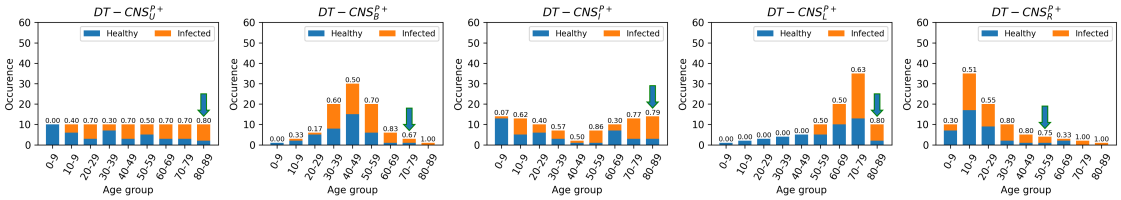
In Fig. 4.10, the $PaR(1,1)$ increases with the increasing transmissibility. In addition, there is a significant upward trend of $PaR(1,1)$ when the transmissibility changes from 0.2 to 0.4. This indicates an increasing level of difficulty to resist an epidemic outbreak given an increased transmissibility, because a higher $PaR(1,1)$ value represents a higher proportion of infected nodes and more resources that are needed to keep the epidemic



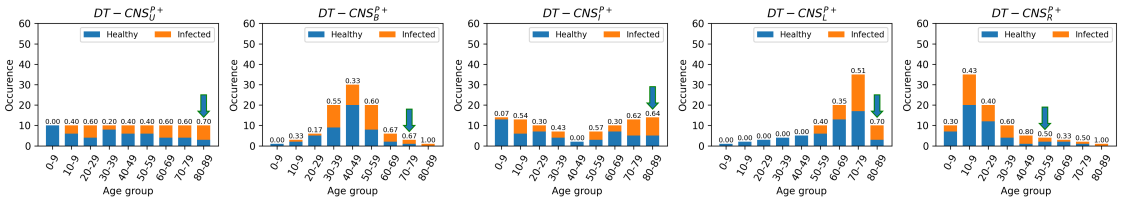
(a) 1.0 transmissibility given an exposure



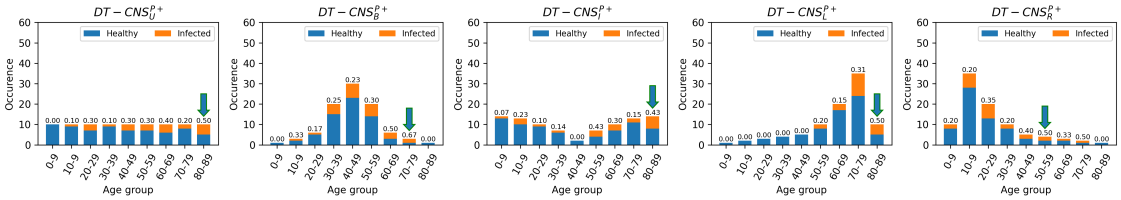
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure



(d) 0.4 transmissibility given an exposure



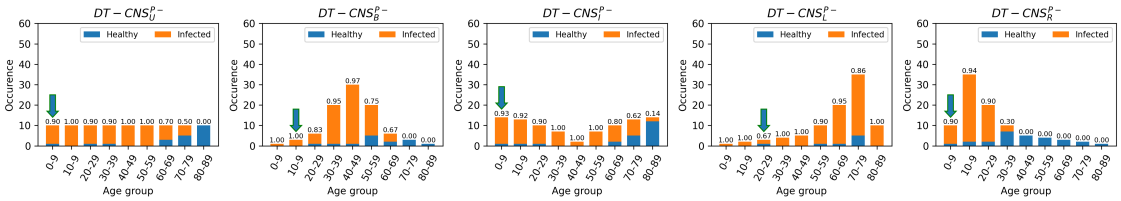
(e) 0.2 transmissibility given an exposure

Figure 4.10: The PaR(1,1) for each age group given different transmissibilities, age distributions and the $P+$ rule.

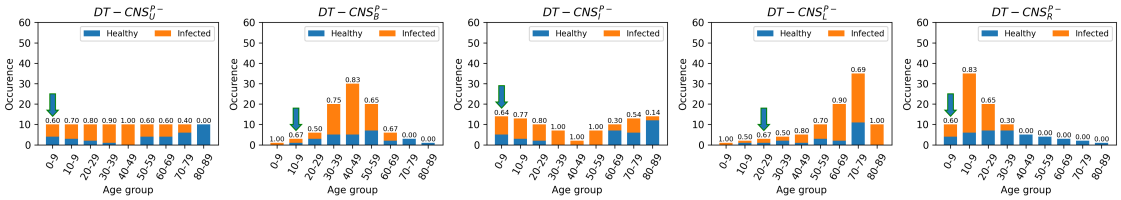
under control. Given the $P+$ rule (preferential attachment to old ages) of network formation but different age distributions, older age groups and the denser age groups, each characterised with higher node degrees and node clustering coefficient, tend to have higher $PaR(1,1)$ values (Fig. 4.4, Fig. 4.5 and Tab. 4.6). Due to the features they prefer and the similar preferences within the same age group, these nodes get directly connected with the seed node, exposing the corresponding age group to much higher infection risk. For example, with the $DT{-}CNS_R$ paradigms and the epidemic transmission from older age groups around the age of $50-59$, the younger and denser age groups around the age of $0-39$ have higher $PaR(1,1)$ values due to the significant number of direct connections with the seed node. Therefore, the preference for old ages with the $P+$ rule (positive preferential attachment to age) may pose the old at higher risk. Similarly, the density of age groups can also lead to higher infection risks. However, this is not always the case in reality and is constrained by our assumptions related to features and the corresponding preferences.

As is shown in Fig. 4.11, given $P-$ rule (preferential attachment to young ages) of network formation, young and dense age groups have higher $PaR(1,1)$ values due to the similarly preferred features within the corresponding age group. Most of these nodes get directly connected with the seed node and have higher infection risks. This indicates a relatively lower resistance level to such an epidemic outbreak as a higher proportion of the respective age groups tend to be infected and treated in epidemic control. The higher $PaR(1,1)$ values in young-age groups given the $P+$ rule contrasts with the lower ones given the $P-$ rule (Fig. 4.10). However, the preferential attachment principles, either $P+$ or $P-$, induces higher infection risks for dense age groups.

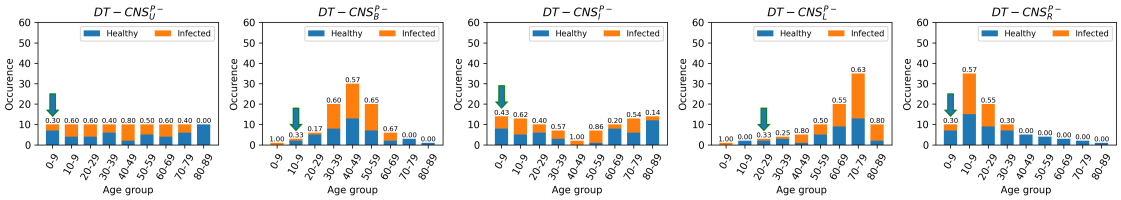
As is shown in Fig. 4.12, given the $H+$ rule (preferences for dissimilar ages) of network formation, the age groups, which are denser and have larger age differences with the seed node, have higher $PaR(1,1)$ values due to their similar preferences for dissimilar ages and thus, the significant number of direct connections across the age groups (between the age group of the seed node and the other age groups). Compared with the $DT{-}CNS^{P^-}$ models built on negative preferential attachment in age values, the $DT{-}CNS^{H+}$ models have a different seed node but similar $PaR(1,1)$ value distributions in the respective age groups. This is because, in both these cases, the young and dense age groups prefer to be connected with the seed node. Therefore, in the shortest time and distance to the first infection, the differences in $P-$ and $H+$ rules can not lead to significant differences in infection risks. This also indicates that the preference for the seed node for each age group influences their $PaR(1,1)$ values directly.



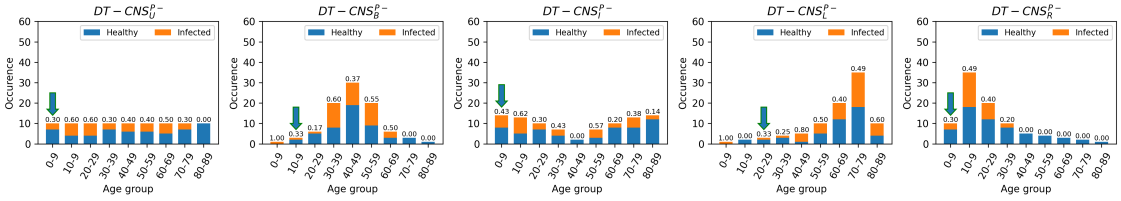
(a) 1.0 transmissibility given an exposure



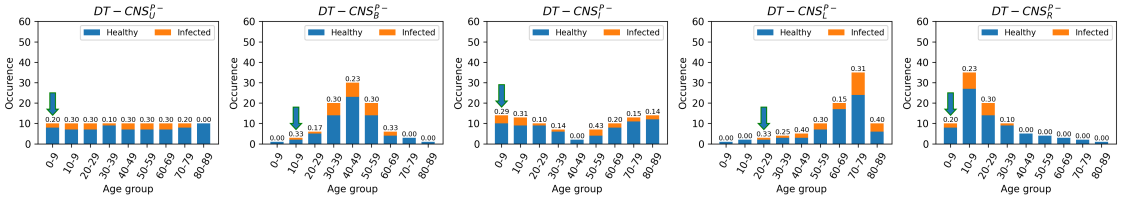
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure

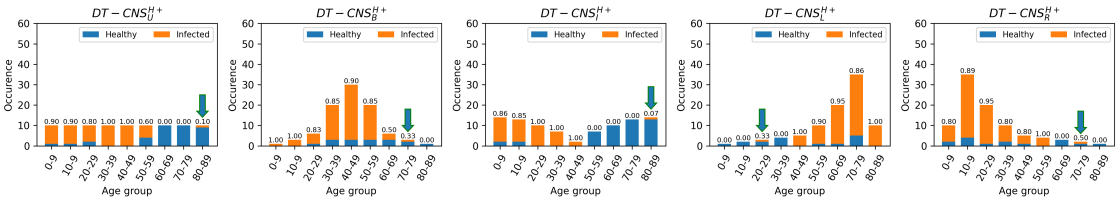


(d) 0.4 transmissibility given an exposure

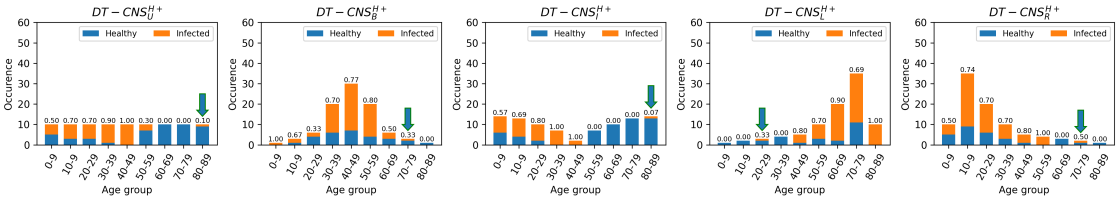


(e) 0.2 transmissibility given an exposure

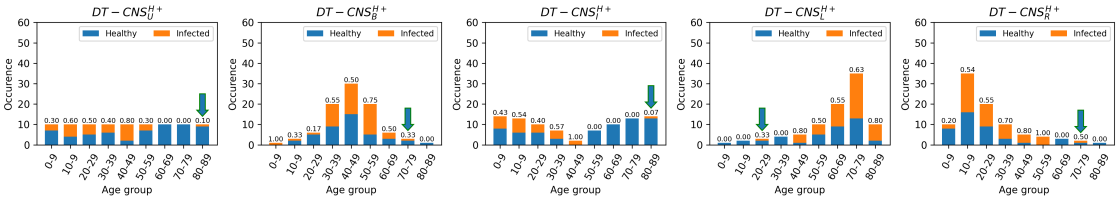
Figure 4.11: The PaR(1,1) for each age group given different transmissibilities, age distributions and the $P-$ rule.



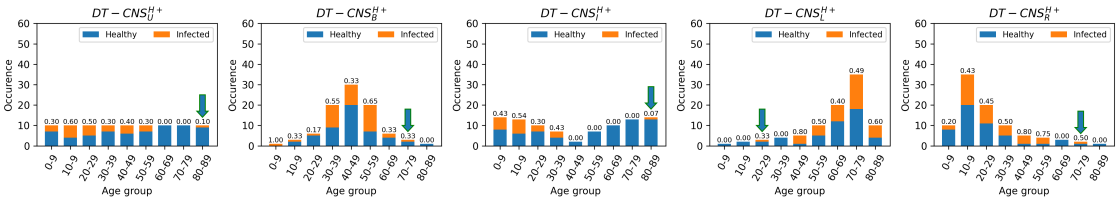
(a) 1.0 transmissibility given an exposure



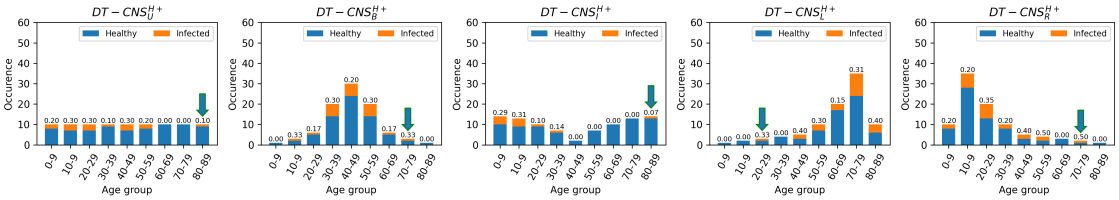
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure

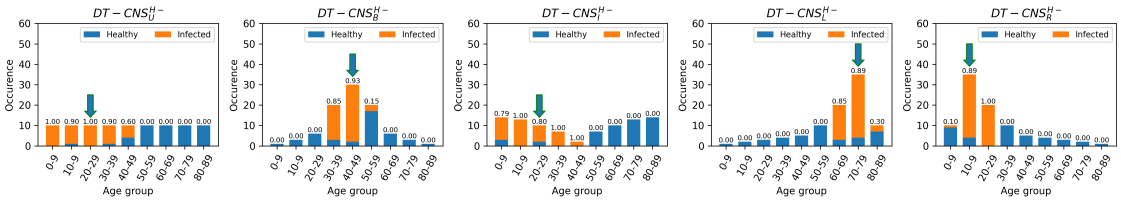


(d) 0.4 transmissibility given an exposure

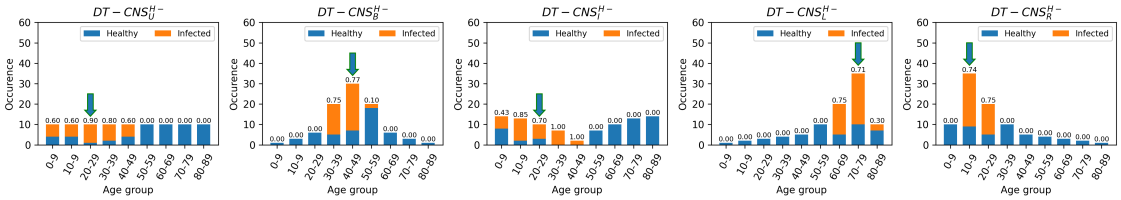


(e) 0.2 transmissibility given an exposure

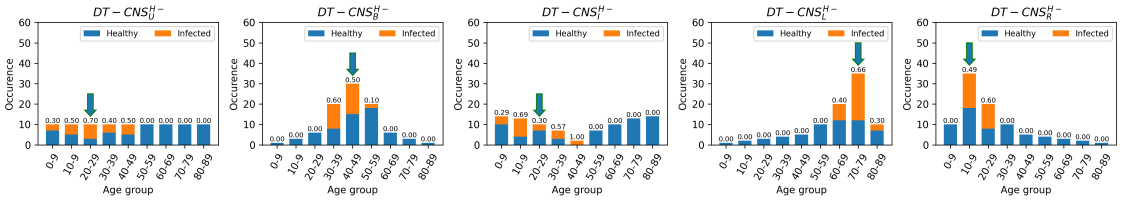
Figure 4.12: The PaR(1,1) for each age group given different transmissibilities, age distributions and the $H+$ rule.



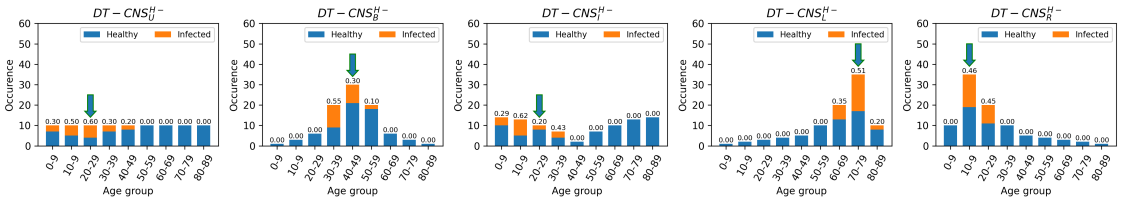
(a) 1.0 transmissibility given an exposure



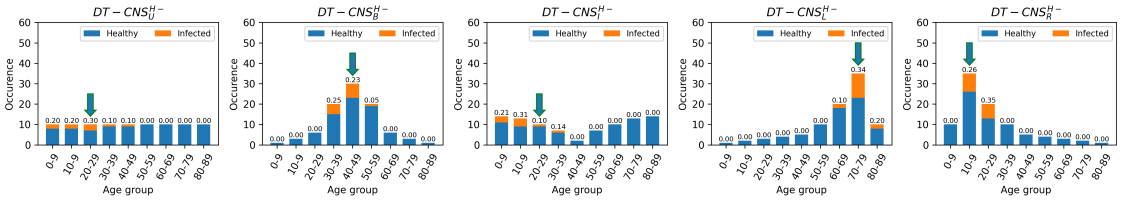
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure



(d) 0.4 transmissibility given an exposure



(e) 0.2 transmissibility given an exposure

Figure 4.13: The PaR(1,1) for each age group given different transmissibilities, age distributions and the $H-$ rule.

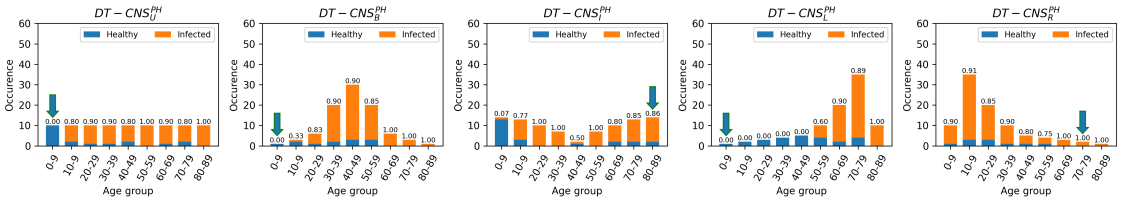
In Fig. 4.13, given the $H-$ rule (preferences for similar ages) of network formation, fewer age groups get involved in the epidemic spreading process since the homophily effect limits the interactions with the seed node to similar age groups (Tab. 4.6). The $PaR(1,1)$ values are higher for age groups which are dense and in a similar age with the seed node due to the similar preferences for similar features ($H-$ rule). In contrast with the other rules (Fig. 4.10, Fig. 4.11, Fig. 4.12 and Fig. 4.13), the lower infection risk, as indicated by smaller $PaR(1,1)$ values, represents a higher resistance level to the epidemic outbreak in the early stages of epidemics and distance to the first infection.

In Fig. 4.14, the PH rule (optimised preferences for ages and age difference) of network formation optimally combines the effect of preferential attachment and homophily principles to approach the target degree distributions of the scale-free networks. The $DT{-}CNS_U^{PH}$ and the $DT{-}CNS_B^{PH}$ models combine the negative preferential attachment in ages and the heterophily effect related to the preferences for different ages. For these two models, the old age groups, dissimilar to the seed node and sharing similar preferences for young nodes $(0-9)$, have high $PaR(1,1)$ values. In contrast, the $DT{-}CNS_I^{PH}$, $DT{-}CNS_L^{PH}$ and the $DT{-}CNS_R^{PH}$ models combine the positive preferential attachment in ages and the heterophily effect related to the preferences for different ages. For these models, the age groups, which are younger (younger than 60) and denser, have high $PaR(1,1)$ values. The phenomena mentioned above indicate the influence of density in age groups on the corresponding $PaR(1,1)$ values. Denser age groups, despite the differences and complexities in preferences, generally have higher infection risks and correspondingly lower resistance levels to the epidemic outbreak.

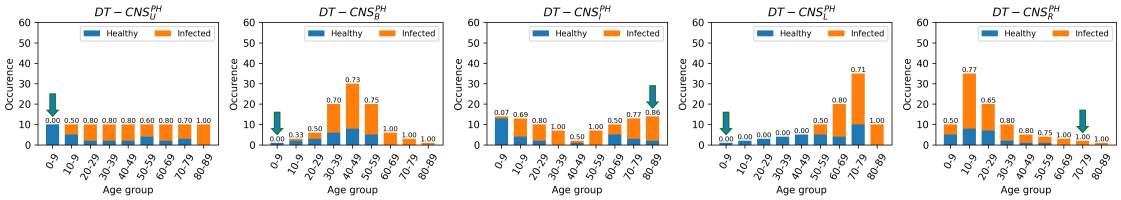
4.4.3 Faithfulness

In this section, we evaluate the simulation-based DT-CNS models based on their network representations' faithfulness. This involves the similarity of simulated network with the target network (Tab. 4.8), the simulation runtime (Fig. 4.15) and the reproducible network statistics.

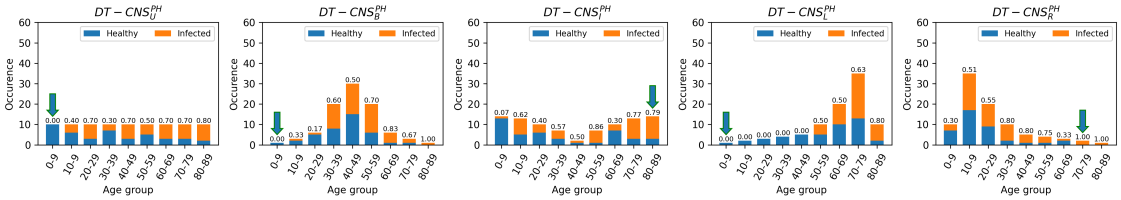
We evaluate the similarity of simulated networks with the target network from a global perspective using the degree distribution. In addition, we apply the Jensen‚ÄìShannon divergence (JS divergence) [412] to measure the similarity of degree distributions. A smaller JS divergence indicates a higher similarity level. In Tab. 4.8, $DT-CNS_U$, $DT{-}CNS_B$ and $DT{-}CNS_I$ paradigms, with more diverse age distributions where nodes are more evenly distributed in the age groups with a similar density, tend to generate more similar degree distributions than the paradigms built on other distributions (the



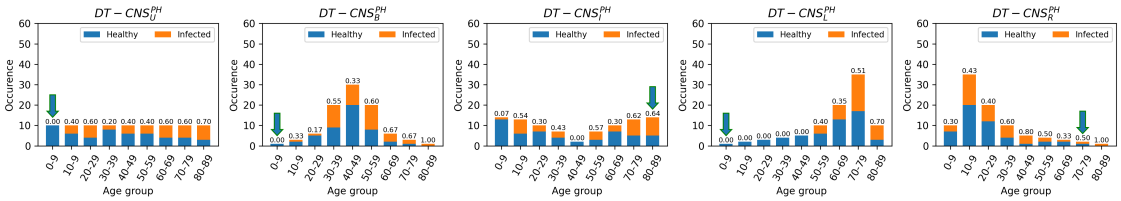
(a) 1.0 transmissibility given an exposure



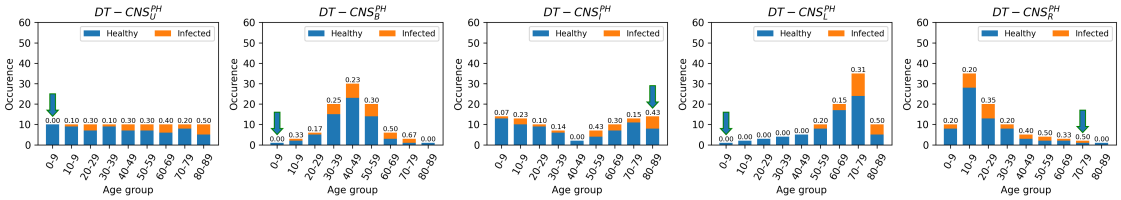
(b) 0.8 transmissibility given an exposure



(c) 0.6 transmissibility given an exposure



(d) 0.4 transmissibility given an exposure



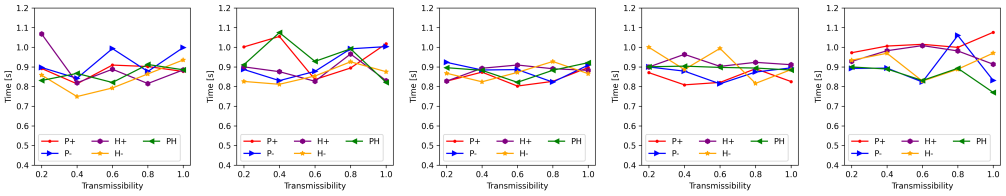
(e) 0.2 transmissibility given an exposure

Figure 4.14: The PaR(1,1) for each age group given different transmissibilities, age distributions and the PH rule.

Table 4.8: The similarity of degree distributions with the target degree distribution, as measured by the Jensen–Shannon divergence.

Paradigms	$P+$	$P-$	$H+$	$H-$	PH
$DT-CNS_U$	0.51	0.53	0.36	0.43	0.26
$DT-CNS_B$	0.50	0.50	0.45	0.50	0.27
$DT-CNS_I$	0.56	0.54	0.46	0.42	0.33
$DT-CNS_L$	0.56	0.53	0.49	0.53	0.28
$DT-CNS_R$	0.51	0.59	0.47	0.55	0.28

blue texts in Tab. 4.8). $DT-CNS^{PH}$ paradigms, with an optimised preference principles have smaller JS values (red texts in Tab. 4.8). $DT-CNS_U^{PH}$ paradigm achieves the most similar degree distribution with the target. Therefore, the scale-free network pattern, described by its degree distribution from a global perspective, can be approached based on node features and the nodes' preferences for connecting with others. The consideration of node features is a good way to built up the complexity of a model and approach the reality.



(a) $DT-CNS_U$ (b) $DT-CNS_B$ (c) $DT-CNS_I$ (d) $DT-CNS_L$ (e) $DT-CNS_R$

Figure 4.15: The runtime of simulation-based DT-CNS models given different features and rules.

We employ the runtime of simulations to evaluate the efficiency of the $DT-CNS$ paradigms, where, as is shown in Fig. 4.15, the time spent on simulation fluctuates with varying transmissibility. The $DT-CNS^{H-}$ paradigms generally have the lowest runtime due to the relatively lower node degrees and the resulting inactivity of epidemic spread. Overall and as expected, the diversity of age distributions and the different rules of network formation, as well as the resulting patterns from the epidemic spreading process, do not make much difference, indicating a similar efficiency level for all the involved $DT-CNS$ paradigms.

Regarding reproducibility, we ensure the repetition of the same resulting patterns with the involved DT-CNS models by setting up a random seed to generate the same

random distributions in the process of feature simulation, network formation and process simulation. These random distributions include (i) the uniform distribution that generates age values within an age group (e.g. $\mathscr{U}(20,29)$ for the [20-29] age group), (ii) the Bernoulli distribution $\mathscr{B}(0.08)$ that generates binary values to represent encounters of node pairs based on the 0.08 encounter rate in network simulation, (iii) the normal distribution $\mathscr{N}(0,0.005^2)$ that generates the random interference in the mutual evaluation between node pairs in the network simulation and (iv) the Bernoulli distribution that generates binary values to represent infections of nodes based on the aggregated infection risks from the exposures in process simulation.

4.4.4 Summary of results

In this section, we summarise the experiment results on the simulation-based DT-CNS models in disaster resilience scenarios connected with the emerging network patterns and the infection status.

In the network dimension, the diversity of the population increases as the age distribution gets more even. Nodes at an older/younger age are more popular than others given $P+$ or $P-$ rules. In contrast, nodes closer/further away from the average age are more popular with $H+$ or $H-$ rules. The $DT-CNS_U^{PH}$ paradigm, based on the most diverse age distribution and driven by an optimised PH rule that prefers a younger and dissimilar age, better approaches the target scale-free network patterns. This indicates that introducing features and combining corresponding preferences enable the DT-CNS models to approach the target states better with a higher complexity level concerning the network attributes and dynamics. Our proposed DT-CNS modelling framework enables the extension of DT-CNS models with a flexible complexity levels and the evaluation of their respective distances to the target.

In the process dimension, we study the infection status in the disaster resilience scenario under the influence of network structures driven by various age features and interaction rules. DT-CNS model driven by $P+$, $P-$ and the $H+$ rules can generally achieve the most infections within two steps (edges) away from the seed. The decrease of transmissibility means that more time is needed to reach an infection peak, preserving time to react to the epidemic outbreak. DT-CNS model paradigms driven by $P+$, $P-$ or $H+$ rules tend to take less time for the final infection status. This is because the seed node has more direct connections with other nodes, in contrast with the nodes who connect with limited number of similar others under the influence of homophily effect. The heterogeneous features and respective preferences lead to complex network

topologies, inducing different infection patterns that occur sooner or later due to various transmissibilities. Therefore, social networks have different infection risks/occurrences and, correspondingly, different resistance levels to an epidemic outbreak within a specific time, dependent on transmissibilities and network topologies driven by features and preferences. As a result, nodes with different features and preferences also suffer different infection risks, present different resistance levels to the epidemic and thus can be treated with different mitigation policies.

To promote disaster resilience, the diversity of the population and the interaction rules are important drivers of case-dependent policy-making. We employ the people at risk $PaR(1,1)$ to represent the proportion of the infected people to be treated by the policymakers with the highest priority. The experiment results suggest that larger transmissibility, a less diverse population, and the $P+$ (positive preferential attachment in age), $P-$ (negative preferential attachment in age) and $H+$ (heterophily effect) rules contribute to a higher $PaR(1,1)$. Among each age group, the $P+$ and $P-$ rules each lead to higher $PaR(1,1)$ for the older and the younger, while $H+$ and $H-$ lead to clustering infection among several age groups due to their connections to similar and dissimilar others respectively. Therefore, the control of epidemic outbreak requires mitigation policies targeted at heterogeneous population. The age groups with higher PaR values given an upcoming epidemic outbreak should be vaccinated in priority to contain the spread of disease. The social networks characterised with higher PaR values within specific time and space require more stringent and urgent isolation policies when the epidemic spreads.

4.5 Conclusion

This chapter answers the **RQ3** by proposing a modelling framework for DT-CNSs based on heterogeneous node features and interaction rules. We also create an evaluation protocol for a faithful representation of reality. Under the modelling framework and evaluation protocol, we conduct a case study on disaster resilience, where we build and compare the simulation-based DT-CNS models given various features and rules.

We build the DT-CNS modelling framework from a top-down perspective concerned with the network and the process dynamics (section 4.3.1). We first employ an inner-rule-based network model to represent the network faithfully and mimic its growth. This network model is built on heterogeneous features (ascribed and topological features – Tab. 4.4) and various feature preference principles (homophily and preferential attach-

ment principles – see section 4.3.3). We also devise the (epidemic) process model based on a seed selection strategy and node adoptability, which all vary with specific conditions concerned with node features and the infection status. Based on these features and rules, the network growth and dynamic process are parameterised with sDNA and pDNA vectors, each incorporating the feature preferences and the consideration of conditions related to transmissibility.

We also propose an evaluation protocol on faithfulness concerned with similarity, efficiency and reproducibility (section 4.3.2). Under this protocol, the definition of faithfulness varies with the simulation-based DT-CNS models, hybrid DT-CNS models and data-driven DT-CNS models due to the ground-truth availability. There are pinpoint-, local- and global-level similarity employed, each focusing on the network components, networks local structures, and the global network statistics. We also use the runtime of the simulation and/or modelling to measure the efficiency and evaluate the reproducibility of the same data, statistics or phenomena.

We finally conduct a case study on disaster resilience and conclude with three findings. First, the diversity of the population, people preferences with who they interact and a change of transmissibility can significantly influence infection patterns and serve as important indicators for policymakers in a disaster resilience scenario. Second, the complexity of network dynamics increases when more feature preference representation principles (e.g. preferential attachment and homophily) are introduced. Such an increase in complexity can improve the faithfulness of network representation by preserving necessary heterogeneous patterns observed in reality. Third, age diversity influences the network structures and induces the epidemic outbreak among specific people. This implies the necessity of targeting the easily infected populations with heterogeneous mitigation policies.

Overall, we answer **RQ3** by proposing a modelling framework and an evaluation protocol that each enables the extension of DT-CNS models with a flexible complexity level and the evaluation of their respective distances to the target. The research contributions, regarding the corresponding research objectives, can be summarised as follows:

- **RQ3**: How to build and evaluate DT-CNS model considering the feature preferences?

    1. **RO3.1**: Develop an extensible DT-CNS model based on heterogeneous node features and rules of feature preference.

RC3.1: Construction of DT-CNS model based on heterogeneous node features and feature preference principles.

2. **RO3.2**: Develop evaluation protocol for DT-CNS model.

   RC3.2: Evaluation protocol to assess the faithfulness of DT-CNS model, considering similarity, efficiency and reproducibility.

3. **RO3.3**: Evaluate the proposed DT-CNS model.

   RC3.3: A case study on DT-CNS models built on different node features, feature preferences and transmissibility of epidemic spread.

In addition, the proposed modelling framework provides the methodological basis of realising the conceptual ideas of DT-CNS models, as presented in chapter 2, for social networked systems and can also be employed in our future study to generate more realistic social networks by incorporating more complex and real-world information. Our simulation-based experiments on age features and related interaction rules indicate the complexity of real-world interactions and reveal the challenge of approaching reality where people's features and preferences can be uncertain and, thus, even more complex. This poses a research question to be addressed in the next chapter, **RQ4:** How do we represent uncertain features with DT-CNS models.

CHAPTER 5

HETEROGENEOUS FEATURE REPRESENTATION FOR DT-CNS MODEL

5.1 Introduction

Research community's efforts have resulted in significant progress in terms of Complex Networked Systems (CNSs) modelling. The accuracy of the approaches consistently increases with the ultimate goal to build a Digital Twin of a real-world Complex Networked System that aims to faithfully represent real systems. The recent advancements are related to the increasing modelling complexity that includes introducing heterogeneous features and realistic rules of network growth and dynamic processes on networks [3] (as presented in chapter 4). The features include both topological and ascribed characteristics connected with the topology and nodes attributes while the rules aim at capturing such phenomena as the preferential attachment and homophily.

Digital Twin-Oriented Complex Networked Systems (DT-CNSs) is an approach proposed to deal with different levels of complexity when it comes to dynamics of and on the network with the aim of faithfully representing and modelling real complex networked systems [1] (chapter 2). In the space of social networks, the aim of the DT-CNS model is to model interactions between people and dynamics of any processes propagating over the resulting from the interactions social network. The information that is utilised to build DT-CNS model includes people's characteristics (e.g. age, gender) and preferences when it comes to who they want to interact with (i.e. preference of people towards specific

age of others).

In the network dimension, many studies developed network models on preferential attachment to topological features [45, 54, 55] (e.g. degree centrality and clustering coefficient, etc.). In contrast, some other studies drive the network growth based on the homophily of nodes considering their similarity in terms of the ascribed features [57–59] (e.g. age, sex, etc.) and the latent representation space [417, 418]. In the process dimension, the dynamic processes on the networks generally select a seed for the first contagious nodes based on topological features [421, 422] (e.g. centrality measures, etc.) and model their transmissibility conditional on specific features about interactions (e.g. multiple/single exposure [18, 333, 406], close/air contact [14, 67], etc.).

CNS models in the current studies get increasingly complex, considering more realistic information about the network and process dimensions. As mentioned before, CNS models first focused on the least complex scenario: static networks without introducing network attributes or dynamic processes [23, 45, 48, 61, 97, 192, 202, 399, 404]. Only recently, the researchers incorporated realistic information for dynamic processes to investigate the impact of different static network topologies on the processes [49–52, 436–438]. For example, [437] validated the impact of network community structure on the epidemic dynamics by modelling the epidemic spreading process on the clustered scale-free networks. Meanwhile, researchers also included more realistic network features in CNS models [34, 135, 136, 386, 439]. [34] proposed a social network simulator based on node features and nodes' preferences for connecting with others. [439] considered the multi-relation information attributed to the edges and studied their impact on the epidemic spreading process. chapter 4 investigates the impact of node features on the networks and the epidemic spreading processes on these networks.

The complexity of CNS models increases even more as the network and the process dimensions evolve with the interrelated changes in the network topology and transmissibility. Current studies generally ignore the interrelated dynamics in CNS models and most of them only focus on the dynamic networks without considering dynamic processes [34, 62–64, 66, 203, 250]. There are just few studies that investigate the impact of dynamic networks on the dynamic processes [67, 68]. In addition, some studies include nodes' features in the CNS model evolution process. Only [34] simulates dynamic networks based on features and the changeable connection principles. Overall, the heterogeneous node features and the changes in CNS models remain a research challenge.

In chapter 4, we proposed a DT-CNS modelling framework for social networked

systems based on the heterogeneous node features and feature preference principles, including preferential attachment and homophily. However, all the features studied in the CNS space are specific qualitative features. The uncertain, subjective and vague inductions of qualitative and quantitative features, such as different views of the linguistic descriptions for height (short, tall, very tall, etc.), emerge as another source of complexity for DT-CNS models and a research gap to be addressed in this chapter.

More specifically, current studies generally use fuzzy sets to express vague qualitative or quantitative data based on various fuzzy membership functions, such as Gaussian and triangular membership functions [60]. They investigate the influence of membership functions on the expressive power as part of sensitivity checks [440–442], and optimise their parameters to improve expressiveness [443, 444]. In contrast, current studies on CNS models generally employ fuzzy sets to represent fuzzy communities [445] or fuzzy trust relations [446, 447]. None of these studies uses fuzzy sets to describe the CNS features of nodes.

Therefore, we propose **RQ4** and in this chapter aim to improve the expressive power of the proposed DT-CNS modelling framework by fulfilling the three research objectives concerning the heterogeneous feature representation principles in DT-CNS models.

- **RQ4**: How to represent uncertain features with DT-CNS models?
    1. **RO4.1**: Develop flexible rules of feature representation for the DT-CNS models.
    2. **RO4.2**: Validate the influence of different ways of feature representation on DT-CNS models.
    3. **RO4.3**: Evaluate the proposed DT-CNS models built on different feature representation principles.

This chapter proposes to model uncertain features with fuzzy sets (**RO4.1**) and improves the expressive power of the current generation 1 DT-CNS models for social networked systems (as presented in chapter 4 and can also be seen in [3]) by introducing heterogeneous feature representation principles (**RO4.2**), including both the crisp and fuzzy value representations. We first extend the proposed generation 1 DT-CNS models by introducing both crisp and fuzzy representations of features with the goal of recreating physical contact networks of different countries from real features, including Belgium, Finland, Germany, Luxembourg, Italy and Poland [1]. Next, we validate the impact of various features and feature representation principles on the DT-CNS models' performance.

[1] `http://www.socialcontactdata.org/socrates/`

Finally, we identify each country's most realistic physical contact networks and investigate their respective disaster resilience given an epidemic spread from the most popular node. The extended DT-CNS models involved in this chapter focus on the static networks and the dynamic processes on these networks. They fall in generation 1 DT-CNS model category and compared with the DT-CNS models proposed in chapter 3, progress with higher dynamics complexity level by introducing feature representation principles. The extended DT-CNS models also provide the methodological basis for DT-CNS models' further development towards a DT.

The main contributions of this chapter are:

- Enabling heterogeneous node feature representation from the perspectives of entities in their decisions on evaluating and connecting with othersby introducing fuzzy representation principles (**RC4.1**);
- Recreating real-world social contact patterns based on real features and principles of feature representation and preference (**RC4.2**);
- Validating the impact of heterogeneous feature representation on network growth and epidemic spread (**RC4.3**);
- Suggesting mitigation policies targeted at heterogeneous populations for disaster resilience (**RC4.3**).

The rest of this chapter is structured as follows. Section 5.2 presents the methodology of DT-CNS models driven by heterogeneous features and rules. It also presents the feature representation principles, including the crisp value and the fuzzy set representations. Following this, section 5.3 presents the empirical analysis, discussion and validation of DT-CNS models that recreate social contact networks from real features. Finally, we conclude with section 5.4.

5.2 Digital Twin-Oriented Complex Networked Systems built on fuzzy representation of features

In this section, we propose a DT-CNS modelling approach for real-world social networked systems based on heterogeneous features, feature representation principles and feature preference principles.

To facilitate an understanding of the concepts used the following example is used. The nodes (entities) in the social networks can have features such as sex or age. These nodes observe and evaluate the features of others to determine their preferred connections. The specific and uncertain features can be represented (feature representation) and preferred (feature preference) differently in this process. In the used datasets, the gender feature take specific binary values between males and females in the nodes' mutual evaluation processes, where they can have specific preferences for males or females (crisp values). In contrast, age, though it can be described with specific values, can only be perceived and preferred by other nodes with uncertainty in the evaluation process. The middle-aged nodes may prefer to contact a young-age node in the social interaction process due to parenthood. In this case, the respective age range can be uncertain and represented with fuzzy sets and the nodes have heterogeneous preferences for each fuzzy set.

We represent this real system as $\mathbf{GP}_t = (\mathbf{G}_t, \mathbf{P}_t)$, composed of a social network $\mathbf{G}_t = (\mathbf{V}_t, \mathbf{E}_t)$ and an epidemic spreading process $\mathbf{P}_t = (\mathbf{S}_t, \mathbf{R}_t, \gamma_t, \eta_t)$ that takes place on this social network at the time t.

The social entities (nodes) are represented as $\mathbf{V}_t = \{\cdots, v_{i,t}, \cdots\}$, with their undirected pairwise interactions denoted in the edge set $\mathbf{E}_t = \{\cdots, \mathbf{e}(v_{i,t}, v_{j,t}), \cdots | v_{i,t}, v_{j,t} \in V_t, i \neq j\}$. Among these entities, the epidemic spread starts from specific social entities $\mathbf{S}_t = \{\cdots, v_{i,t}, \cdots | v_{i,t} \in V_t\}$ with a transmissibility at the level of γ_t and recovery rate of η_t and results in changes of the others' infection status $\mathbf{R}_t = \{\mathbf{r}(v_{i,t}) | v_{i,t} \in V_t\}$. The infection status is represented by a binary value, 0 or 1, each representing the node's healthy and infected status.

The social contact patterns and the epidemic propagation, as represented with $\mathbf{G}_t = (\mathbf{V}_t, \mathbf{E}_t)$ and $\mathbf{P}_t = (\mathbf{S}_t, \mathbf{R}_t)$, are each driven by heterogeneous features $\mathbf{f}(v_{i,t})$ related to each social entity $v_{i,t}$ and the corresponding principles of feature preference $\mathbf{fp}(v_{i,t})$ and feature representation $\mathbf{fr}(v_{i,t})$ as explained in the sections below. The features, preferences related to features and the respective representation principles vary for the entity $v_{i,t}$ and can change with the time t.

5.2.1 Feature representation

If we assume there are k features considered in society, each social entity $v_{i,t}$ has a k dimensional feature vector:

$$\mathbf{f}(v_{i,t}) = [f_1(v_{i,t}), f_2(v_{i,t}), \cdots, f_k(v_{i,t})] \tag{5.1}$$

which, correspondingly, is followed with a k dimensional vector $\mathbf{fr}(v_{i,t})$ to determine the feature representation principles. The representation of the k-th feature is determined by $fr_k(v_{i,t}) = (cf_k^p(v_{i,t}), \mathbf{m}_k^p(v_{i,t}), cf_k^h(v_{i,t}), \mathbf{m}_k^h(v_{i,t}))$, including a binary indicator $cf_k^p = 0 \quad \text{or} \quad 1$ for the usage of crisp feature values or fuzzy sets respectively, and, if using fuzzy sets, the l_k^p–length vectors $\mathbf{m}_k^p(v_{i,t})$ that define the respective membership function of each fuzzy set. Similarly, the binary indicator $cf_k^h = 0 \quad \text{or} \quad 1$ decides the usage of crisp feature differences or fuzzy feature differences, and is followed with the l_k^h–length vectors $\mathbf{m}_k^p(v_{i,t})$ that determines the corresponding Membership functions.

Specifically, for any entity $v_{i,t}$, his perception of the feature $f_k(v_{j,t})$ of any other node such as $v_{j,t}$ is unfolded as:

$$f_k^u(v_{j,t}|v_{i,t}) = \begin{cases} f_k(v_{j,t}) & \text{if } cf_k^p(v_{i,t}) = 0. \\ [f_{k,1}(v_{j,t}|v_{i,t}), f_{k,2}(v_{j,t}|v_{i,t}), \cdots, f_{k,l_k^p}(v_{j,t}|v_{i,t})] & \text{if } cf_k^p(v_{i,t}) = 1. \end{cases} \tag{5.2}$$

where we unfold an uncertain perception of feature $f_k(v_{j,t})$ into a series of membership degrees according to the Membership functions of fuzzy sets.

We can calculate membership degree to the y_{th} fuzzy set based on a Gaussian Membership function as:

$$f_{k,y}(v_{j,t}|v_{i,t}) = exp\{-\frac{1}{2}(\frac{f_k(v_{j,t}) - \mu_y^p(v_{i,t})}{\sigma_y^p(v_{i,t})})^2\} \tag{5.3}$$

where $\mu_y^p(v_{i,t})$ and $\sigma_y^p(v_{i,t})$, included in the vector $\mathbf{m}_k^p(v_{i,t})$, respectively represents the location and the spread of the involved Gaussian distribution for the specific focal node $v_{i,t}$.

As a result, the k-dimensional feature vector $\mathbf{f}(v_{j,t})$ of each node is unfolded into a z-dimensional vector $\mathbf{f}^u(v_{j,t}|v_{i,t})$, dependent on the feature representation principle of the other node. Imagining a k-length feature vector $\mathbf{f}(v_{j,t})$ which only has one uncertain feature $f_2(v_{j,t})$, it is unfolded in to the following z-length vector $\mathbf{f}^u(v_{j,t}|v_{i,t})$:

$$\begin{aligned} \mathbf{f}^u(v_{j,t}|v_{i,t}) = [&f_1(v_{j,t}), f_{2,1}(v_{j,t}|v_{i,t}), f_{2,2}(v_{j,t}|v_{i,t}), \\ &\cdots, f_{2,l_2}(v_{j,t}|v_{i,t}), f_3(v_{j,t}), \cdots, f_k(v_{j,t})] \end{aligned} \tag{5.4}$$

Similarly, the k-dimensional vector $\Delta\mathbf{f}(v_{i,t}, v_{j,t})$ represents the difference between entities in society:

$$\Delta\mathbf{f}(v_{i,t}, v_{j,t}) = [\Delta f_1(v_{i,t}, v_{j,t}), \Delta f_2(v_{i,t}, v_{j,t}), \cdots, \Delta f_k(v_{i,t}, v_{j,t})] \tag{5.5}$$

The difference $\Delta f_k(v_{i,t}, v_{j,t}) = |f_k(v_{i,t}) - f_k(v_{j,t})|$ between the entities in terms of feature f_k considering the fuzzy set denoted with l_k^h, from the subjective view of $v_{i,t}$, is unfolded

as:

$$\Delta f_k^u(v_{i,t}, v_{j,t}|v_{i,t}) = \begin{cases} \Delta f_k(v_{i,t}, v_{j,t}) & \text{if } cf_k(v_{i,t}) = 0. \\ [\Delta f_{k,1}(v_{j,t}|v_{i,t}), \Delta f_{k,2}(v_{j,t}|v_{i,t}), \cdots, \Delta f_{k,l_k^h}(v_{j,t}|v_{i,t})] & \text{if } cf_k(v_{i,t}) = 1. \end{cases} \tag{5.6}$$

where an uncertain perception of feature difference $\Delta f_k(v_{j,t}, v_{i,t})$ can also be unfolded from the perspective of entity $v_{i,t}$ into a series of membership degrees.

Similarly, we can also calculate the membership degree to the y_{th} fuzzy set based on a Gaussian Membership function as:

$$\Delta f_k^u(v_{i,t}, v_{j,t}|v_{i,t}) = exp\{-\frac{1}{2}(\frac{\Delta f_k(v_{i,t}, v_{j,t}) - \mu_y^h(v_{i,t})}{\sigma_y^h(v_{i,t})})^2\} \tag{5.7}$$

where $\mu_y^h(v_{i,t})$ and $\sigma_y^h(v_{i,t})$ each represents the location and the spread of the involved Gaussian distribution for the specific focal node $v_{i,t}$ and are included in the parameter vector $\mathbf{m}_k^h(v_{i,t})$.

Correspondingly, the difference between entities $v_{i,t}$ and $v_{j,t}$, from the subjective view of $v_{i,t}$ can be unfolded into the vector $\Delta\mathbf{f}^u(v_{i,t}, v_{j,t}|v_{i,t})$:

$$\begin{aligned} \Delta\mathbf{f}^u(v_{i,t}, v_{j,t}|v_{i,t}) = [&\Delta f_1(v_{i,t}, v_{j,t}), \Delta f_{2,1}(v_{i,t}, v_{j,t}|v_{i,t}), \Delta f_{2,2}(v_{i,t}, v_{j,t}|v_{i,t}), \\ &\cdots, \Delta f_{2,l_2}(v_{i,t}, v_{j,t}|v_{i,t}), \Delta f_3(v_{i,t}, v_{j,t}), \cdots, f_k(v_{i,t}, v_{j,t})] \end{aligned} \tag{5.8}$$

To help illustrate, we again take the age and sex features of nodes as an example. According to the crisp representation principle, the sex features can be represented with specific binary values such as 0 and 1, each denoting the males and the females, respectively. Similarly, the age feature takes specific values and can be represented with the respective crisp value. However, rather than considering the specific age values and determining whether to prefer a specific age or age difference, the nodes can prefer middle-aged nodes around 35 and the nodes with an age difference of around 30. In this case, a single age can be represented with three membership values, each denoting the respective membership to the young-aged (age around 15), middle-aged (age around 35) and the old-aged (age around 65). The age difference between a pair of nodes can also be represented by the degree of membership to the fuzzy set (concerning the age difference around 30).

5.2.2 Feature preference

The nodes determine their connection preferences based on the others' features and their feature differences in comparison to others. These preferences are respectively captured

by the preferential attachment and homophily principles, each respectively concerning the preferences for features and feature differences.

Referring to our previous study on DT-CNS model [3], the preference for the unfolded features $\mathbf{f}^u(v_{i,t})$ and the unfolded feature differences $\mathbf{f}^u(v_{i,t}, v_{j,t}|v_{i,t})$ is defined with four same length vectors:

$$\mathbf{fp} = \{\mathbf{p}, \mathbf{w}^p, \mathbf{h}, \mathbf{w}^h\} \tag{5.9}$$

As an example, the uncertain deductions on a single feature of age can be unfolded with three fuzzy sets, each describing the respective membership to the young-age, middle-age and old-age group.

The trinity value vectors $\mathbf{p}$ and $\mathbf{h}$ indicate the negative/zero/positive preference for features and feature differences. $\mathbf{w}^p$ and $\mathbf{w}^h$ each represents the weight of preference, valuing within $(0, 1]$.

Following the toy example in section 5.2.1, the nodes have the corresponding preferences for the respective features and feature differences. The preferential attachment principles can e.g. describe the nodes' preferences for connecting with male and middle-aged nodes. The homophily effect can describe the nodes' preferences for example for the same gender and an age difference of around 30, which involves comparing the features between any node pair.

5.2.3 Network formation

In our proposed DT-CNS models, social networks grow based on the preference for the unfolded features. The pairwise interactions between entities are each scored as:

$$\pi(v_{i,t}, v_{j,t}) = (\frac{1}{2}\pi^p(v_{i,t}, v_{j,t}) + \frac{1}{2}\pi^h(v_{i,t}, v_{j,t}) + \epsilon_{ij,t}) * I(\zeta) \quad v_i, v_j \in V, i \neq j \tag{5.10}$$

where the encounter factor $I(\zeta)$ determines whether two nodes encounter each other based on a binary value between 0 and 1, following a Bernoulli distribution at the probability of ζ. The node pair can encounter, denoted as $I(\zeta) = 1$, by the chance of ζ and then evaluate each other for a connection. Tn contrast, the node pair, without an encounter (as denoted by $I(\zeta) = 0$), cannot conduct mutual evaluations and stays unconnected. The random interference $\epsilon_{ij,t} \sim \mathcal{N}(0, \sigma^2)$ follows a random normal distribution and causes variations of scores of node pairs when they are similarly scored through mutual evaluation.

Constrained by the encounters, the network growth is driven by ranking dynamics based on the score $\pi(v_{i,t}, v_{j,t})$ of any node pair through mutual evaluation concerned about their preferences for features and feature differences.

The preferential attachment score $\pi^p(v_{i,t}, v_{j,t})$ incorporates the preference for the other nodes with larger/smaller crisp/fuzzy features:

$$\pi^p(v_{i,t}, v_{j,t}) = \mathbf{f}^u(v_{j,t}|v_{i,t})^\tau(\mathbf{p}(v_{i,t}) \odot \mathbf{w}^p(v_{i,t})) + \mathbf{f}(v_{i,t}|v_{j,t})^\tau(\mathbf{p}(v_{j,t}) \odot \mathbf{w}^p(v_{j,t})) \tag{5.11}$$

The homophily score $\pi^h(v_{i,t}, v_{j,t})$, concerned with feature differences, incorporates the preference for the other nodes with similar/dissimilar crisp/fuzzy features:

$$\pi^h(v_{i,t}, v_{j,t}) = \quad \Delta\mathbf{f}(v_{i,t}, v_{j,t}|v_{i,t})^\tau(\mathbf{h}(v_{i,t}) \odot \mathbf{w}^h(v_{i,t})) + \Delta\mathbf{f}(v_{i,t}, v_{j,t}|v_{j,t})^\tau(\mathbf{h}(v_{j,t}) \odot \mathbf{w}^h(v_{j,t})) \tag{5.12}$$

The score list $\Pi_t = \{\pi(v_{i,t}, v_(j,t))|v_{i,t}, v_{i,t} \in V, i \neq j\}$ keeps frozen with fixed features and feature preference and representation principles. Given the required edge number $\lambda_{e,t}$, we create a static network to represent the first $\lambda_{e,t}$ discrete interactions at this fixed time point.

Based on the respective set-ups in section 5.2.1 and section 5.2.2, the nodes in the toy example evaluate and select preferred connections considering the sex and age features. The male and middle-aged nodes around 35 are preferred and, thus, likely to have more connections than the other nodes. Meanwhile, the nodes with the same gender and an age difference of around 30 tend to cluster in the network formation process. However, it is hard to determine the most popular nodes with the superposition of preferential attachment and homophily principles. We must create more realistic networks based on real features to explore feature-driven network patterns.

5.2.4 Epidemic transmission

The epidemic spreads from the seed nodes $\mathbf{S}_t = \{\cdots, v_{i,t}, \cdots | v_{i,t} \in V_t\}$. A healthy node, given a connection with an infected node, gets infected at the probability of γ_t. Correspondingly, the infection probability $\mathscr{P}^s$ of a healthy node $v_{i,t}$ increases with the increased connections with the infected nodes.

$$\mathscr{P}^s(v_{i,t}) = 1 - \prod_{v_{j,t} \in \mathbf{V}_t} (1-\gamma_t)^{e(v_{i,t}, v_{j,t})} \tag{5.13}$$

where $\prod_{v_{j,t} \in \mathbf{V}_t} (1-\gamma_t)^{e(v_{i,t}, v_{j,t})}$ denotes the probability of keeping healthy given the connections $e(v_{i,t}, v_{j,t})$ with any other node. In contrast, the recovery rate η_t denotes the

probability of the recovery of any infected node despite of the connections with infected others.

The features and the corresponding preferences determine the network patterns, which impact the epidemic spreading process and the well-being of any node in a social network. Based on the toy example presented in section 5.2.1, section 5.2.2, and section 5.2.3, the popular nodes, with preferred features such as males at the age of around 30, tend to have more epidemic risks due to higher number of connections with others. Meanwhile, nodes with similar genders and an age difference of around 30 can also have higher infection risks due to the clustering network patterns. To create realistic social networks and to investigate the influence of features on the epidemic spreading process, we conduct an empirical analysis and present its results and discussion in section 5.3.3.

5.3 Empirical Analysis

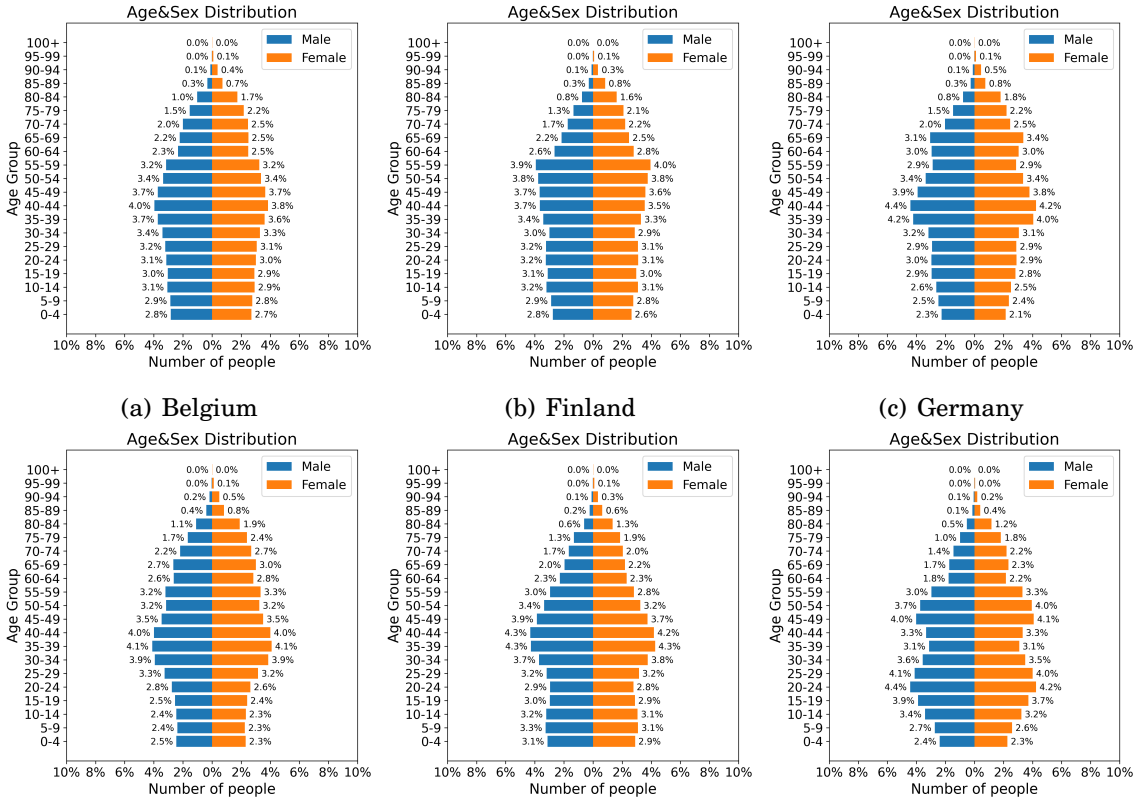
In this section, we use the proposed DT–CNS model to generate the physical contact networks for six countries (data available for Belgium, Finland, Germany, Luxembourg, Poland and Italy from 2005) and investigate their respective disaster resilience to an epidemic spread. Given partial observable information about this real scenario (as described in section 5.3.1), we build hybrid DT-CNS models based on both real and simulated data, including (i) a hybrid network built on real nodes' features and simulated edges and (ii) a simulation-based epidemic spreading process.

We build hybrid network models with available age & sex features and feature representation principles (section 5.3.2) and evaluate their performance in simulating social networks and simulating realistic social contact matrices in section 5.3.3. The selected model is validated using sensitivity analysis considering the increasingly flexible feature representation principles in section 5.3.4. In section 5.3.6, we simulate the social networks of each country and simulate the epidemic spreading process as a reference for policy-making in disaster resilience.

5.3.1 Data Description

We have partial observable information about the social contact patterns in different countries, including the social contact matrices between different age groups, as disclosed on the website [1], and the age & sex distributions of the population gathered for year 2005 (Fig. 5.1).

We obtain the age and sex distributions of the population in each country from the website [2] and, referring to these distributions, generate age features for each country. As shown in Fig. 5.1, the age &sex distributions of the five countries have a similar shape. There are more people in the middle age ranges (25-60) than in other age groups, and more males in the older groups (60-90+) than females. The entire population's age is between 0 and 90.



(a) Belgium (b) Finland (c) Germany

(d) Italy (e) Luxembourg (f) Poland

Figure 5.1: The age & sex-structured population profiles in 2005 of countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy (e) Luxembourg and (f) Poland.

We obtain social contact matrices for countries, including Belgium, Finland, Germany, Luxembourg, Italy and Poland, from the social contact data website [1]. The social contact matrix $\{m_{i,j}\}$ describes the average number of physical contacts between each pair of age classes reported weekly between May 2005 and September 2006 in a specific country's survey [448]. Considering the reciprocal nature of social connections, the social contact matrices $\{m_{i,j}\}$ are adjusted based on the population size in the respective age

[1] http://www.socialcontactdata.org/socrates/

[2] https://www.populationpyramid.net/

groups [449]:

$$m_{ij}^{reciprocal} = \frac{m_{ij}N_i + m_{ji}N_j}{2N_i} \tag{5.14}$$

where N_i and N_j each represents the population size in age group i and j. $m_{ij}^{reciprocal}N_i$ represents the average number of social contacts between age group i and j. Given a specific number of nodes and the corresponding age features, we can create their social contact matrix by adjusting the country's social contact matrix $m_{i,j}$ according to equation 5.14 based on the information related to the population size N_i and N_j in the network simulations.

In this chapter, we assume 90 nodes for the simulated social networks in each country. Given the social contact matrix, we derive a number of edges for each network. As shown in Tab. 5.1, the social networks generally have around 700 edges. However, the social network of Finland only has 339 edges, and in contrast, the social network of Poland has 800 edges.

Table 5.1: Number (No.) of nodes and edges.

Country	Belgium	Finland	Germany	Italy	Luxembourg	Poland
No.Nodes	90	90	90	90	90	90
No.Edges	614	336	462	770	687	793

5.3.2 Network Models Initialisation

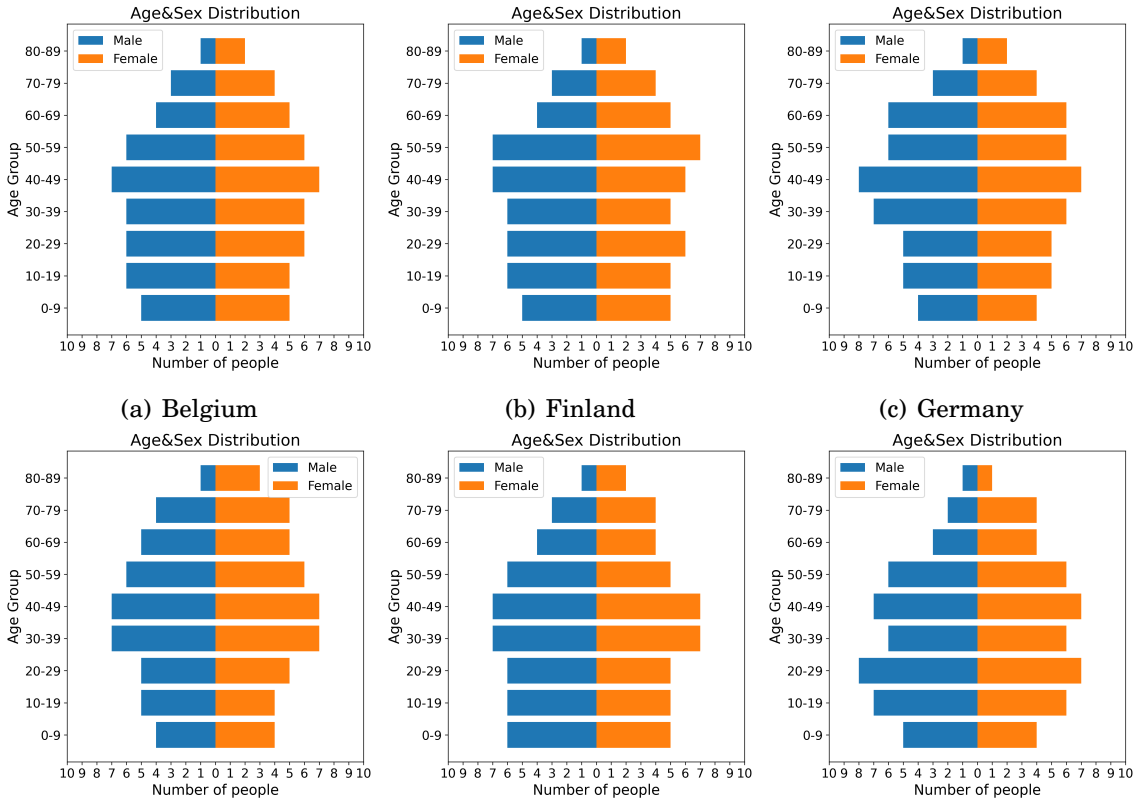
This section provides a description of the process to build network models to uncover the unobservable social preferences behind complex social interactions and unfold the social contact matrices into social networks (Tab. 5.1). We also introduce different features and their representation principles to investigate their respective impact on the faithfulness of network representation (Table. 5.2). This involves the random network model RN built on the random encounters of each node pair and the hybrid network models HN that introduce various features (age and sex) and feature representation principles (crisp/fuzzy) and preferences to those features (a mixture of preferential attachment and homophily).

Table 5.2: Model description.

Model Name	Features	Rules	
		Representation	Preference
RN	—-	—-	—-
$HN-A_c$	Age	Crisp	Preferential attachment & Homophily
$HN-A_f$	Age	Fuzzy	Preferential attachment & Homophily
$HN-A_c-S_c$	Age	Crisp	Preferential attachment & Homophily
	Sex	Crisp	Preferential attachment & Homophily
$HN-A_f-S_c$	Age	Fuzzy	Preferential attachment & Homophily
	Sex	Crisp	Preferential attachment & Homophily

5.3.2.1 Features

Features, including the age and the sex of nodes, are generated according to the age and sex-distributions of each country in 2005 (Fig. 5.1 for real data and Fig. 5.2 for simulated data). As the population age is between 0 and 90, we assume the age range in the simulations as $[0,89]$ and simulate the age and sex features accordingly. We classify the 90 nodes into eighteen age and sex groups, considering the age classes separated every ten years and the differences between females and males. Within each age and sex class, we randomly simulate the age features for each of the nodes.



(a) Belgium (b) Finland (c) Germany

(d) Italy (e) Luxembourg (f) Poland

Figure 5.2: The simulated age and sex-distributions in 2005 for countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland.

Based on the differences in the simulated age distribution, sex distribution and age and sex distribution, we analyse the diversity of each country's population by calculating the Hill numbers (Fig. 5.3). The Hill numbers are calculated with an order value q. A higher q value indicates a higher sensitivity level to the relative frequencies of the species (different groups considering the features of age and sex). In addition, higher Hill numbers represent a higher diversity level, characterised with higher abundance of the respective species and more even distribution of nodes in the respective species [432]. Therefore, the sex diversity of the countries is similar, which approaches the value of 2, which is the most diverse case (Fig. 5.3(b)). The age diversity of the countries varies depending on the country, which deviates from the most diverse case possible at the value of 9 (Fig. 5.3(a)). The diversity of the population increases a lot when considering the differences in age and sex. The differences in each country's population diversity have also increased (Fig. 5.3(c)). Generally speaking, Belgium has the highest level of diversity given any diversity references (age, sex, or age&sex) considering the calculated

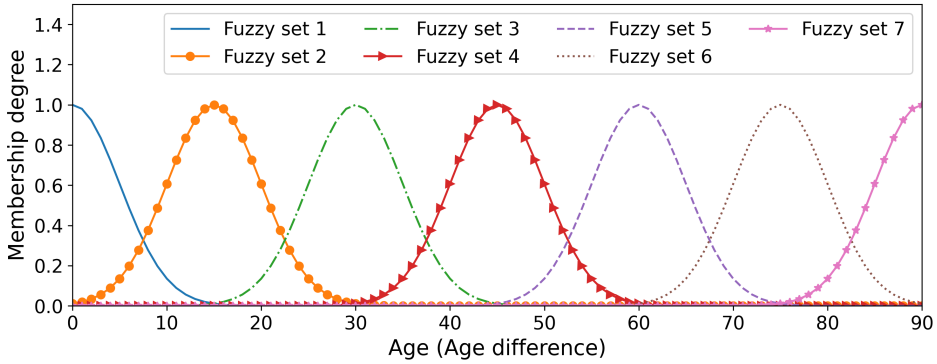
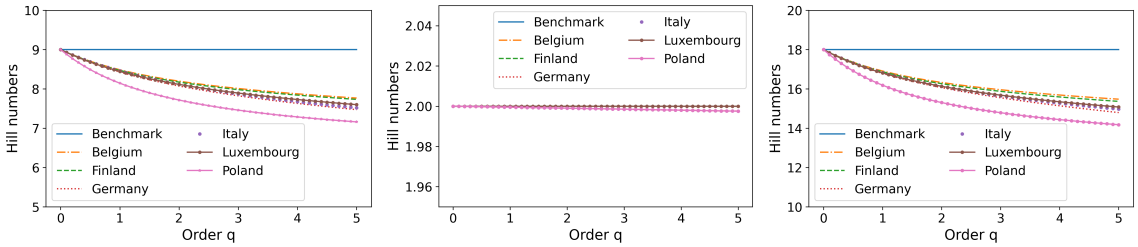


Figure 5.4: The fuzzy representation of age (age difference).

Hill numbers.



(a) Age Group (b) Sex Group (c) Age&Sex Group

Figure 5.3: The diversity of the population in each country, as measured by the Hill numbers based on age, sex and age&sex groups.

5.3.2.2 Feature representation and preference

Principles involved in the hybrid modelling of social networks cover two aspects: (i) feature representation principles and (ii) feature preference principles.

In this chapter, we represent the sex features with binary values, where females and males are each represented respectively as ones and zeros with a crisp boundary. The representation of age features follows either a crisp or fuzzy principle. Given a crisp principle, we represent the age and the age differences with their corresponding crisp values. Given a fuzzy principle, we initialise seven fuzzy sets to represent the uncertain inductions of the ages and the age differences. Each fuzzy set describes the respective membership of age/age difference to each abstract age/age difference group (Fig.5.4).

As shown in Fig. 5.4, we represent age (age difference) value with a vector incorporating seven membership values for the corresponding age ranges. We employ Gaussian membership functions to describe the membership of a single age (age difference) value to the age ranges that centre around the list of μ values, each initialised as 0, 15, 30,

45, 60, 75 and 90. In addition, we initialised the standard deviations of each Gaussian membership function with an identical value at 5.

We employ preferential attachment and homophily principles to describe the preference for features and generate social networks. With a crisp representation principle, the homophily principle describes the preference for similar/dissimilar features by preferring small/large feature differences. In contrast, the preferential attachment principle describes the preference for small/large features values. In contrast, with a fuzzy representation principle, the homophily principle transforms to the preference for fuzzy feature difference ranges. The preferential attachment principle describes the preference for fuzzy feature value ranges.

5.3.3 Model Selection considering Features and Rules

Given the ground-truth information about social contact matrices and age and sex distributions, we optimise the preferences and weights of preferences for the age and sex to generate social networks. Based on an assumed encounter rate at 0.8, we can also recreate the corresponding social contact matrices based on the social networks simulated by the network models. Referring to our previous report [3], we measure the faithfulness of the network representation by local-level similarity, reproducible network topology and modelling efficiency.

In our empirical analysis, we ensure the **reproducibility** of the same networks by setting a random seed for the Bernoulli distribution that generates binary values to represent encounters. We can recreate the same networks with the same sDNA and encounter rate. In addition, we measure the **efficiency** of generating hybrid networks with the runtime of a simulation. As shown in Tab. 5.3, the runtime of simulations does not vary much with the feature values changes, given a crisp representation principle. In contrast, the runtime increases when a fuzzy representation principle is employed, caused by the fuzzification process. In addition, the runtime of simulation differs depending on the country, which increases with the required number of edges (Tab. 5.1).

For the **faithfulness**, we measure the local similarity between real and hybrid networks based on the Euclidean distance (EU distance) between the social contact matrices. As shown in Tab. 5.3, the EU distance generally decreases with the increasing number of features and the introduction of fuzzy age (age difference). This implies the positive impact of increasing structural complexity and dynamics complexity on social network simulations, each concerning the introduction of heterogeneous features and

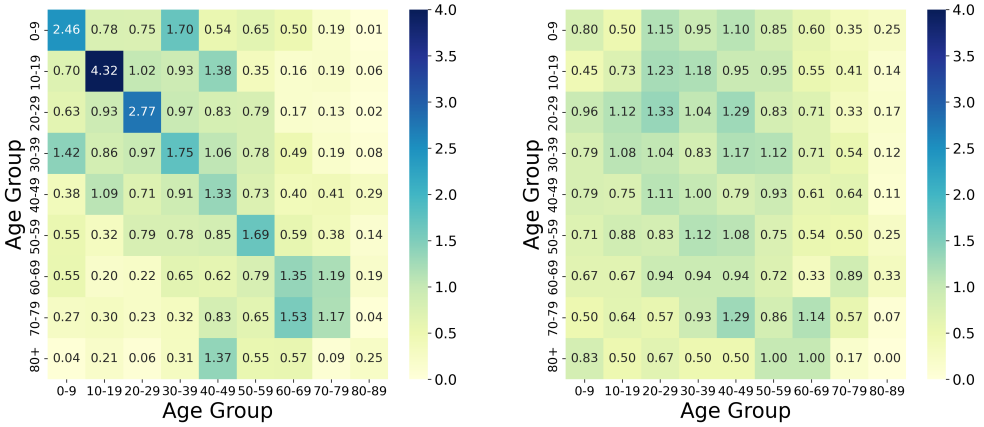
Table 5.3: The similarity and the efficiency of network representation with the network models, as measured by the Euclidean distance (EU distance) between social contact matrices and the Runtime (second).

	Country	Belgium	Finland	Germany	Italy	Luxembourg	Poland
RN	EU distance	5.51	4.80	4.60	11.62	8.48	9.37
	Runtime	0.31	0.23	0.22	0.24	0.23	0.23
$HN-A_c$	EU distance	5.76	4.16	4.68	10.39	7.36	8.06
	Runtime	0.24	0.22	0.43	0.24	0.23	0.23
$HN-A_f$	EU distance	6.30	4.08	4.15	10.40	7.19	7.78
	Runtime	0.62	0.62	0.63	0.62	0.66	0.63
$HN-A_c-S_c$	EU distance	4.51	3.65	3.82	9.19	5.84	6.63
	Runtime	0.33	0.31	0.31	0.32	0.32	0.32
$HN-A_f-S_c$	EU distance	4.02	2.55	3.32	9.05	5.40	5.62
	Runtime	0.71	0.71	0.76	0.73	0.71	0.74

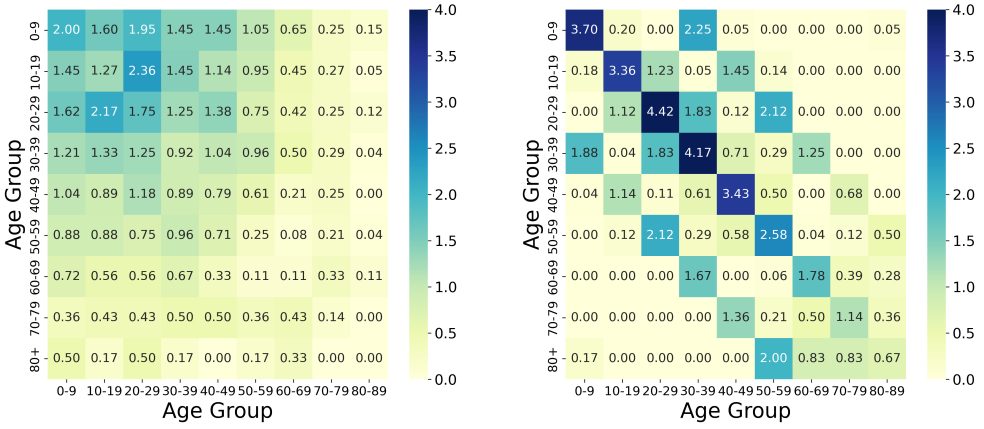
flexible fuzzy representation principles. Generally, $HN-A_f-S_c$ models enable the creation of the most similar hybrid social networks to those in real world for all the countries with the highest complexity level in both structural and dynamics dimensions.

To investigate the influence of structural and dynamics complexity on network generation, we simulate the social networks for each country with the optimised HN models and recreate the social contact matrices (Fig. 5.5 and Fig. 5.6, Fig. 5.7, Fig. 5.8, Fig. 5.9 and Fig. 5.10).

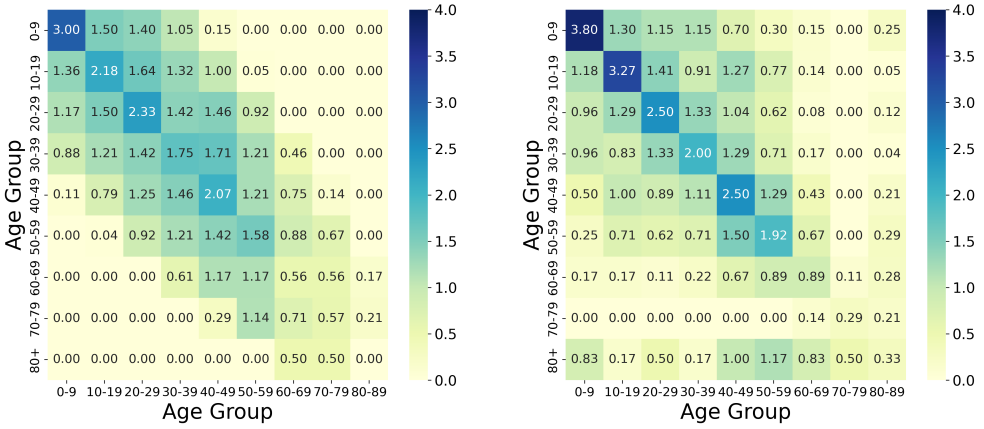
Overall, the social contact matrix of real social network (the target) in each country shows people's preference for similar ages, younger ages and age differences at around 30. With an optimised encounter rate and sex preference, The RN model manages to fill in the cells of contact matrices by random encounters and recreate a low social contact rate for large age groups due to the small number of people. With an optimised age preference, the $HN-A_c$ model only captures the homophily to age but leaves two blank corners for the recreated social contact matrix. In contrast, the $HN-A_f$ model, with the introduction of fuzzy age representation, fills in more cells and can better capture the preference for similar ages and age differences around 30. Compared with the $HN-A_c$ and $HN-S_c$ models, the $HN-A_c-S_c$ model combines both advantages by incorporating the preference for age values and similar ages. However, there are still blank cells in the two corners. This problem is solved with the $HN-A_f-S_c$ model by



(a) Target (b) RN

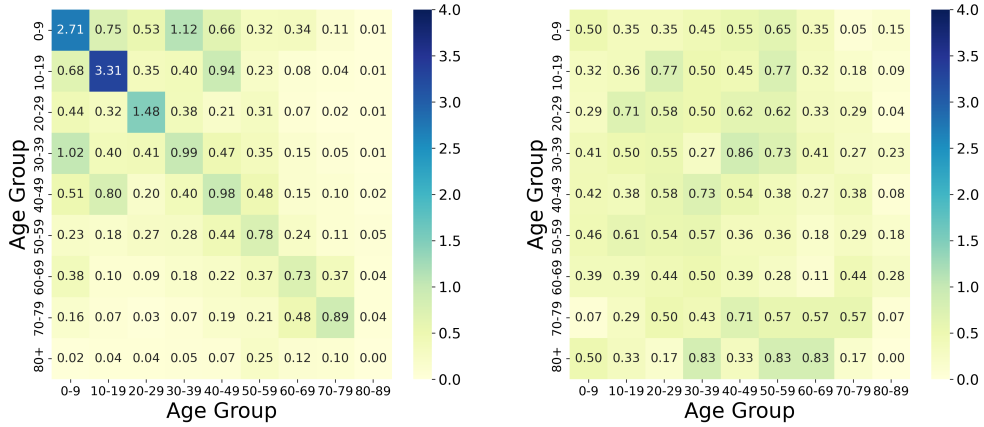


(c) $HN-A_c$ (d) $HN-A_f$



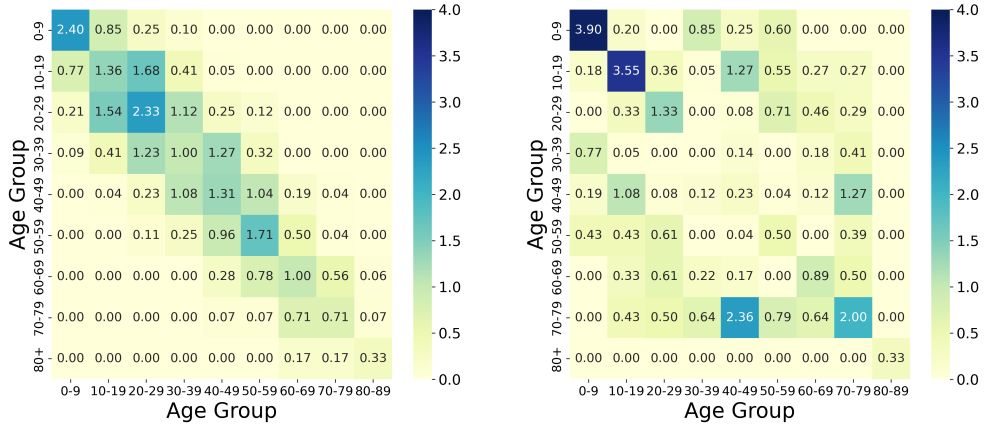
(e) $HN-A_c-S$ (f) $HN-A_f-S$

Figure 5.5: The target social contact matrix ((a) target) and the recreated social contact matrices of Belgium, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models.



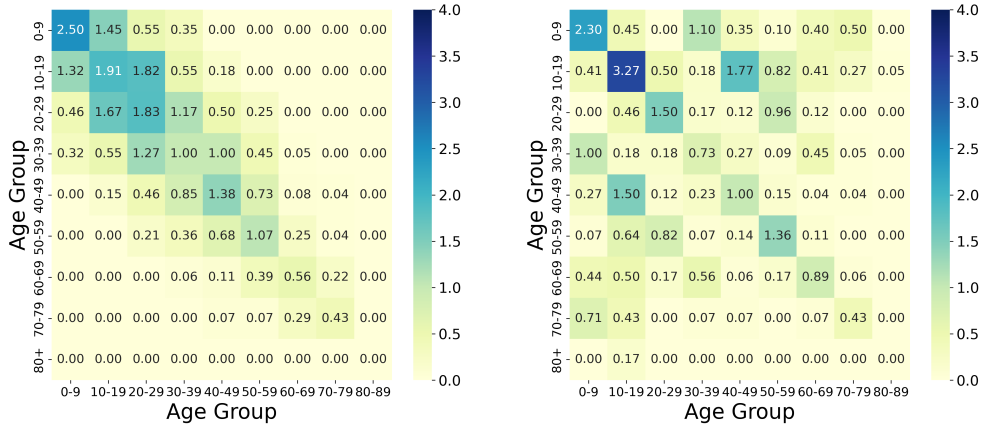
(a) Target

(b) RN



(c) $HN - A_c$

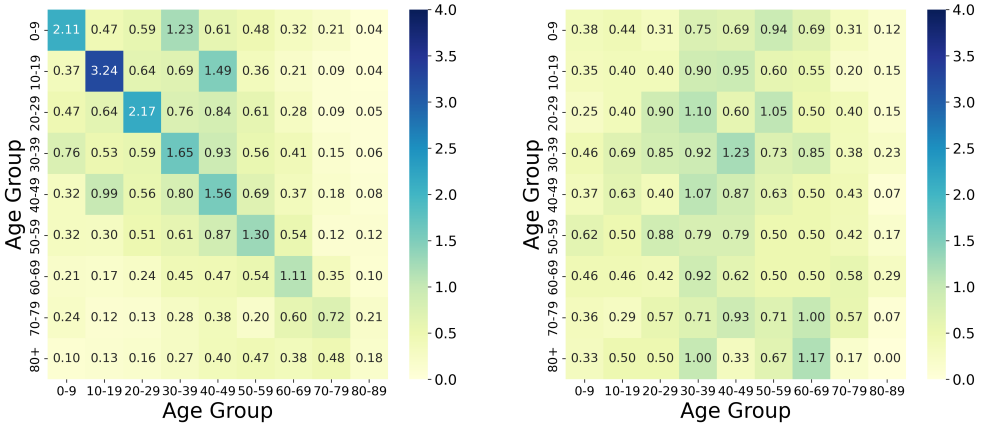
(d) $HN - A_f$



(e) $HN - A_c - S$

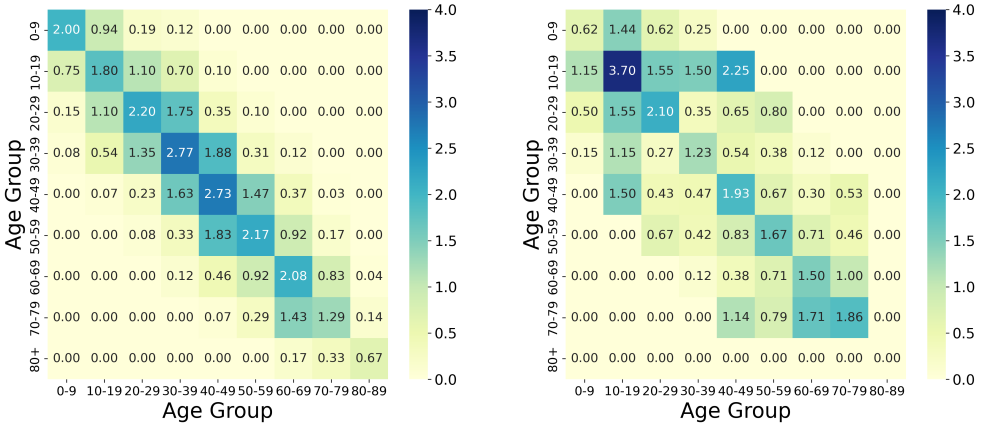
(f) $HN - A_f - S$

Figure 5.6: The target social contact matrix ((a) target) and the recreated social contact matrices of Finland, which are respectively generated by the (b) RN, (c) $HN - A_c$, (d) $HN - A_f$, (e) $HN - A_c - S$ and (f) $HN - A_f - S$ models.



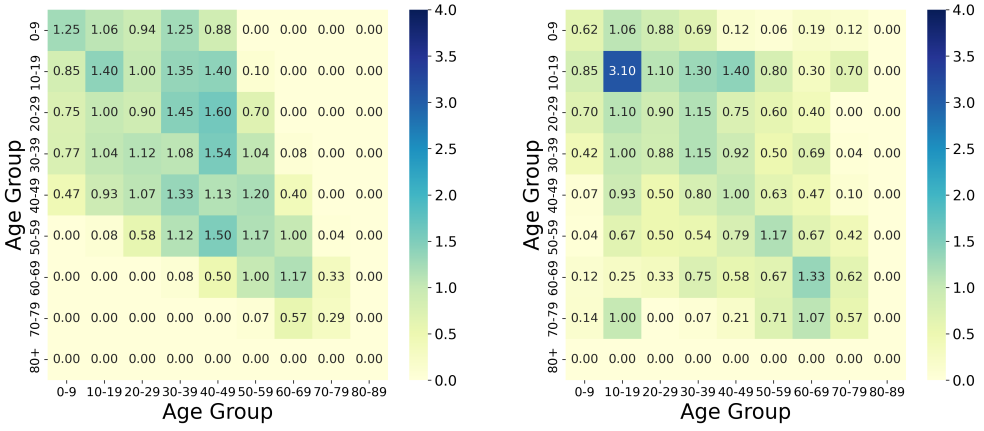
(a) Target

(b) RN



(c) $HN-A_c$

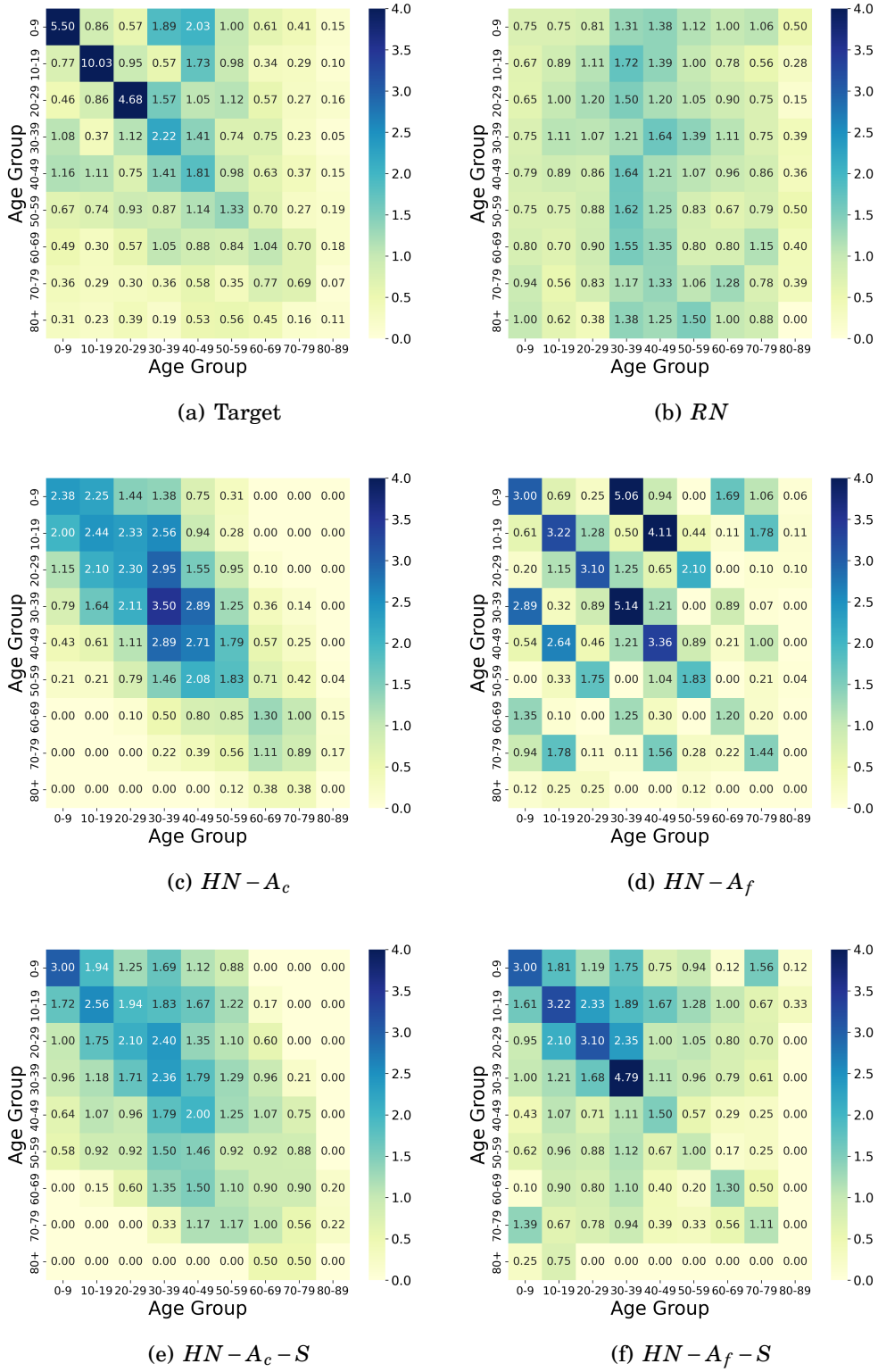
(d) $HN-A_f$



(e) $HN-A_c-S$

(f) $HN-A_f-S$

Figure 5.7: The target social contact matrix ((a) target) and the recreated social contact matrices of Germany, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models.

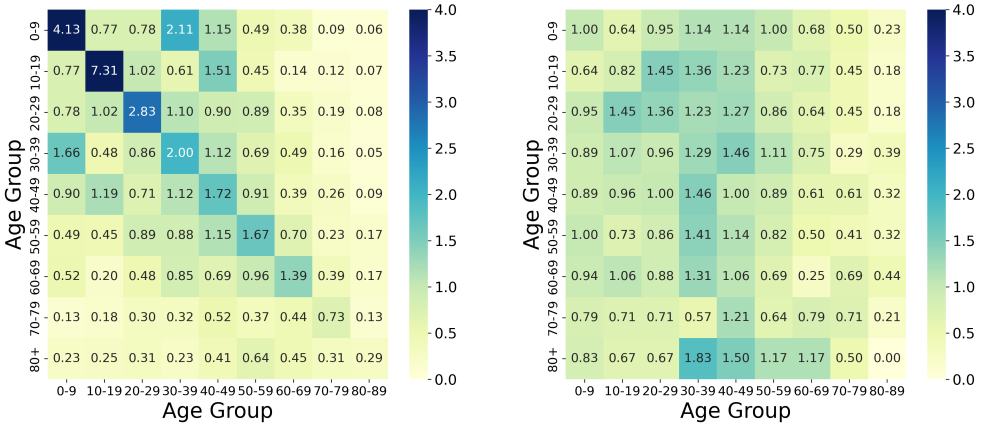


(a) Target (b) RN

(c) $HN-A_c$ (d) $HN-A_f$

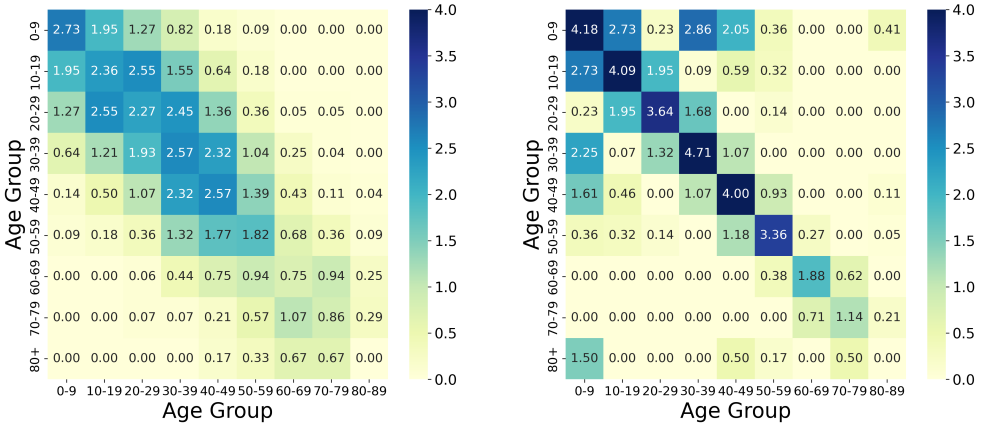
(e) $HN-A_c-S$ (f) $HN-A_f-S$

Figure 5.8: The target social contact matrix ((a) target) and the recreated social contact matrices of Italy, which are respectively generated by the (b) RN, (c) $HN-A_c$, (d) $HN-A_f$, (e) $HN-A_c-S$ and (f) $HN-A_f-S$ models.



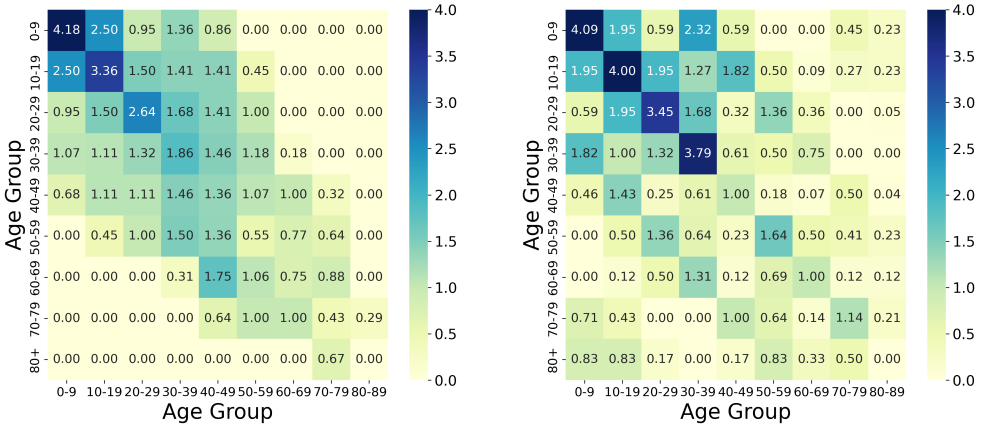
(a) Target

(b) RN



(c) $HN - A_c$

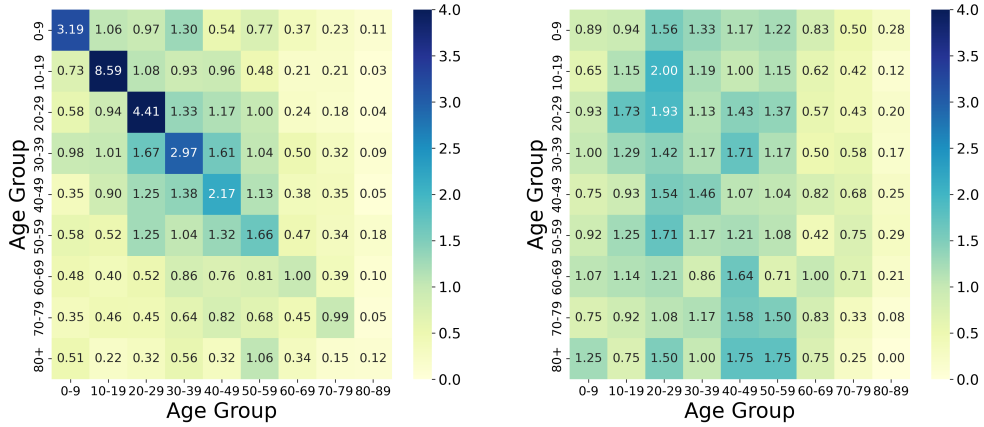
(d) $HN - A_f$



(e) $HN - A_c - S$

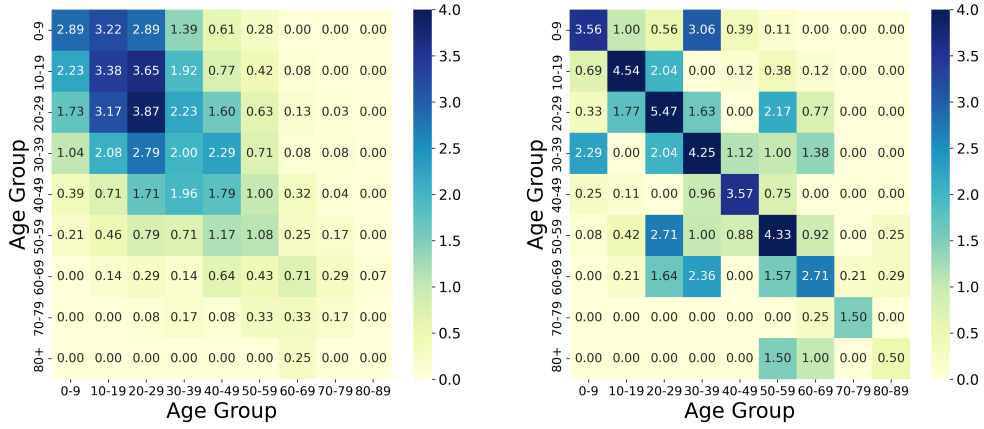
(f) $HN - A_f - S$

Figure 5.9: The target social contact matrix ((a) target) and the recreated social contact matrices of Luxembourg, which are respectively generated by the (b) RN, (c) $HN - A_c$, (d) $HN - A_f$, (e) $HN - A_c - S$ and (f) $HN - A_f - S$ models.



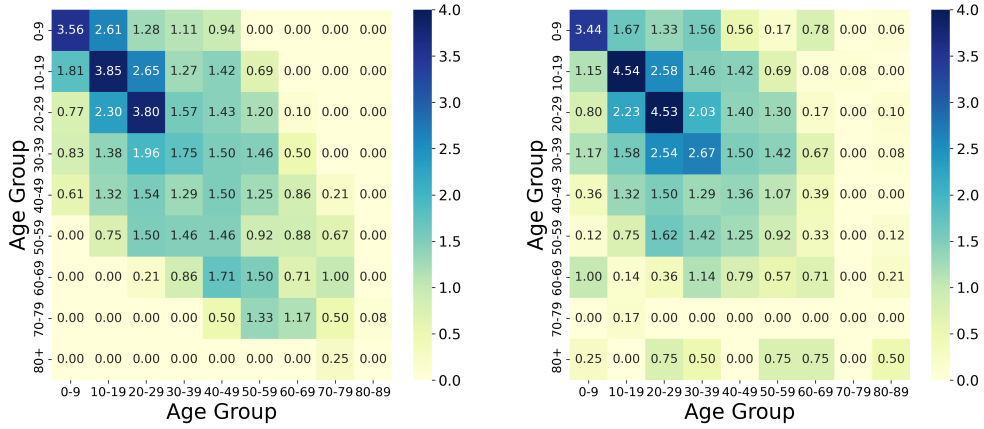
(a) Target

(b) RN



(c) $HN - A_c$

(d) $HN - A_f$



(e) $HN - A_c - S$

(f) $HN - A_f - S$

Figure 5.10: The target social contact matrix ((a) target) and the recreated social contact matrices of Poland, which are respectively generated by the (b) RN, (c) $HN - A_c$, (d) $HN - A_f$, (e) $HN - A_c - S$ and (f) $HN - A_f - S$ models.

the fuzzy representation of age and the corresponding preference for a fuzzy feature (feature difference). The $HN-A_f-S_c$ model generates the most similar social networks, as suggested by the local-level similarity when recreating the social contact matrices. The increased diversity of the population and the fuzzy representation of features lead to an increased similarity level. This conclusion holds for all the networks involved in the experiments – the DT-CNS models built on more realistic features can create more realistic networks. However, the selection of appropriate fuzzy sets to recreate social networks remains a research challenge to address. Therefore, we select the $HN-A_f-S_c$ model based on the model performance presented in this section and the next section extends this model with flexible fuzzy sets to investigate the impact of flexible fuzzy representation principles on the model performance in approaching realistic networks.

5.3.4 Model Extension With Flexible Fuzzy Representation Principles

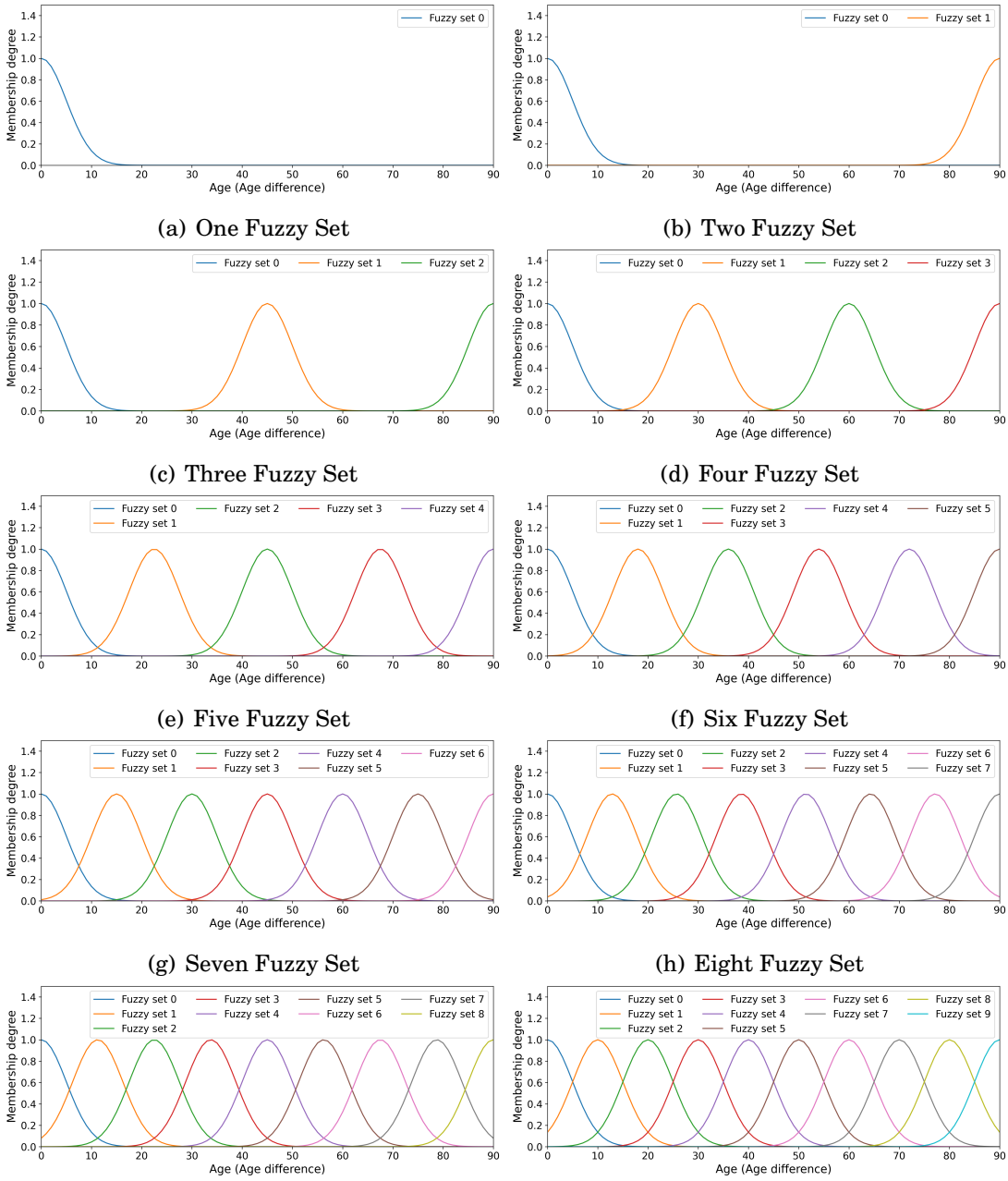
In this section, we investigate the impact of flexible fuzzy representation principles on the model performance of DT-CNS models. We increase the flexibility of fuzzy representation principles by introducing different numbers of fuzzy sets and tuning the parameters of the fuzzy sets. This increases the complexity of modelling DT-CNS models and influences the accuracy of creating realistic social networks. As an example of extending DT-CNS model with higher flexibility and complexity, we extend the $HN-A_f-S_c$ model (network model built on the fuzzy representation of age features and the crisp representation of sex features) selected in section 5.3.3 towards better model performance in recreating social contact matrices.

As presented in the model initialisation process (section 5.3.2, the fuzzy representation of the ages and age differences follow the assumptions of (i) seven fuzzy sets, which are determined by the same (ii) Gaussian Membership functions, parameterised with the μ list $[0,15,30,45,60,75,90]$ and the σ list $[5,5,5,5,5,5,5]$. This section investigates whether the expressiveness of the age features can be improved when the fuzzy sets become more flexible. We select the best-performing fuzzy sets based on this enlarged searching space in section 5.3.4.1. We also conduct a global sensitivity analysis of the selected fuzzy sets and select parameters to be optimised for model performance improvement in section 5.3.4.2.

5.3.4.1 Selection of Fuzzy Sets

We first introduce a flexible number of fuzzy sets to the respective representation of age and age difference. In this space, we re-initialise the $HN-A_f-S_c$ models with different fuzzy set number, ranging from 1 to 10. In addition, we initialise the μ list by equally dividing the age (age difference) value range $[0,90]$ and set up a σ list of the same values at 5 (Fig. 5.11). The initialised σ value at 5 enables each fuzzy set to cover an age range of around 20 years that centres around the respective μ value, corresponding to a generation that nearly excludes parenthood relationships.

As shown in Fig. 5.11, the combination of fuzzy sets covers more value range with more overlapping areas as the fuzzy set number increases. The representation of features gets more detailed as we unfold a single feature value into a longer vector concerned with more specific age groups. In a real-world scenario, the increasing number of fuzzy sets for age indicates that people pay more attention to age. In a network composed of kids, adults and kinship relations, people focus on the age difference of around 30. In contrast, people attend to more fuzzy age (age difference) ranges when they develop friendships. However, there is a challenge in optimising the corresponding same-length preference principles and the parameters of membership functions. This chapter aims to find the optimal fuzzy sets for age and age differences from these initialised fuzzy representation principles.



(a) One Fuzzy Set (b) Two Fuzzy Set

(c) Three Fuzzy Set (d) Four Fuzzy Set

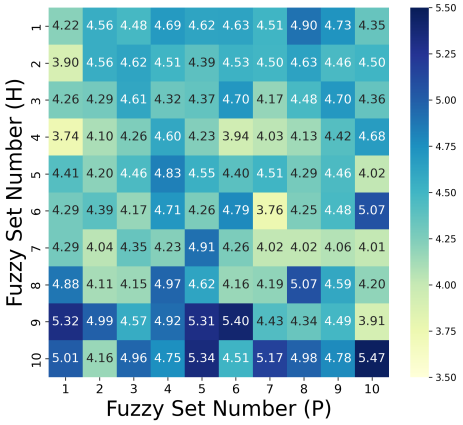
(e) Five Fuzzy Set (f) Six Fuzzy Set

(g) Seven Fuzzy Set (h) Eight Fuzzy Set

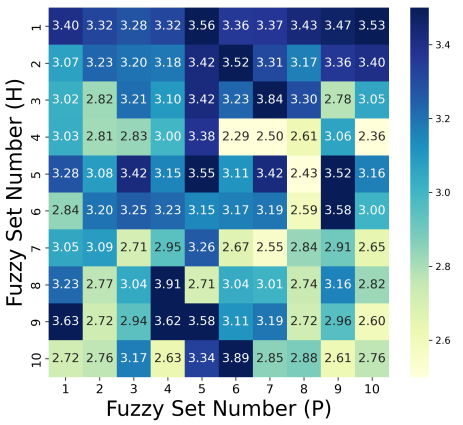
(i) Nine Fuzzy Set (j) Ten Fuzzy Set

Figure 5.11: The searching space of fuzzy sets for age and age difference. In the experiments, both the age and age difference initialise the fuzzy sets based on the same parameter set-ups. Therefore, this figure

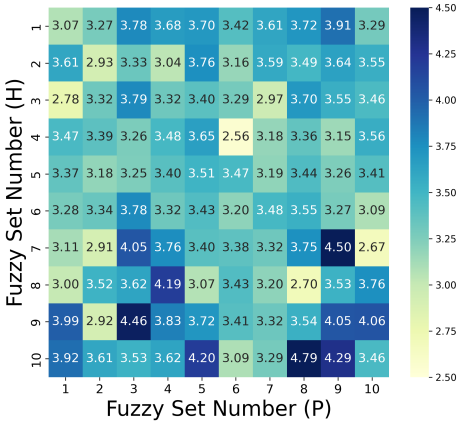
We visualise the influence of the respective fuzzy set number for age and age difference on the *EU* distance between the target and the recreated social contact matrices (Fig.5.12). For each country, the *EU* distance varies with the fuzzy sets without a clear trend, where an increasing number of fuzzy sets does not necessarily improve the model



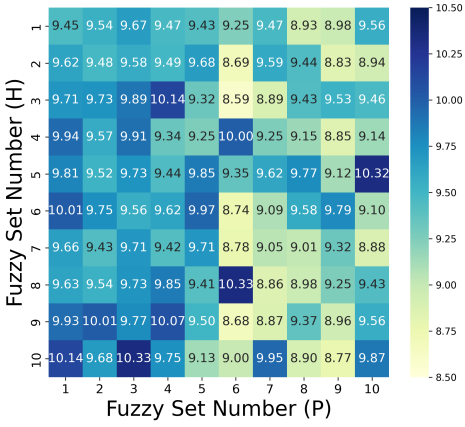
(a) Belgium



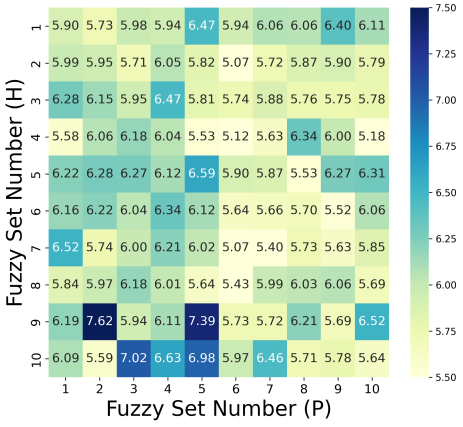
(b) Finland



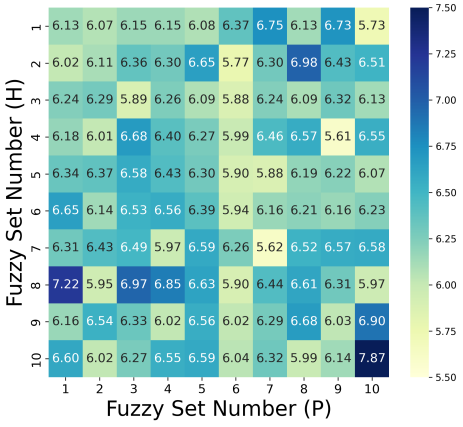
(c) Germany



(d) Italy



(e) Luxembourg



(f) Poland

Figure 5.12: The *EU* distance achieved with different settings of fuzzy sets (Fuzzy Set Number (H) for age difference on the y-axis and Fuzzy Set Number (P) for age on the x-axis) in experiments on (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland.

performance. This indicates that the number of fuzzy sets impacts the expressive power of DT-CNS models but does not directly correlates with the accuracy of recreating realistic social networks. In addition, selecting appropriate fuzzy sets for realistic feature representation is a research challenge to address in DT-CNS modelling.

Table 5.4: The EU distance achieved with the respective fuzzy sets selected for each country.

Country	Number of fuzzy sets		EU distance
	Age	Age difference	
Belgium	7	7	4.02
	1	4	3.74
Finland	7	7	2.55
	6	4	2.29
Germany	7	7	3.32
	6	4	2.56
Italy	7	7	9.05
	6	3	8.59
Luxembourg	7	7	5.40
	6	2	5.07
Poland	7	7	5.62
	9	4	5.61

Compared with the EU distance achieved with seven fuzzy sets (section 5.3.3), we improve the model performance by introducing more flexible fuzzy sets. Based on Tab. 5.4, we select the best-performing fuzzy sets with the smallest EU distance for each country. Generally, smaller number of fuzzy sets is needed to represent age differences. Particularly, the "four fuzzy set" covers the uncertain perception around age differences around 0, 30 and 60, which reveals people's preferences for kinship relations in Belgium, Finland, Germany and Poland. In addition, we select the "six fuzzy set" to represent the age values around 0, 18, 36, 54, 72 and 90. It captures people's preferences for age values in Finland, Germany and Italy. In addition, we select a more significant number of fuzzy sets to represent age values than age differences, except for the case of Belgium. This indicates that people in each country put more importance to age differences rather than age. People unfold a single age value into much longer vectors in this condition by considering more fuzzy sets. In contrast, they have limited but focused consideration of age differences, employing fewer fuzzy sets to unveil their consideration for age differences.

5.3.4.2 Selection of Parameters

We conduct a global sensitivity analysis of the Membership Functions' parameters for the selected fuzzy sets of each country in section 5.3.4.1. We use the Variogram Analysis of Response Surfaces (VARS) to estimate parameters' sensitivities. This method uses directional variograms to estimate Integrated Variogram Across a Range of Scales (IVARS), by integrating the directional variogram up to a given perturbation scale of interest (e.g, 10%, 30% or 50%). In this space, these directional variograms represent the sensitivity of the model response to each parameter across the full range of perturbation scales [450, 451]. In this chapter, we measure the parameter sensitivity with $IVARS_{10}$ (scale range = 0-10%), $IVARS_{30}$ (scale range = 0-30%) and $IVARS_{50}$ (scale range = 0-50%). We also measure the variance-based total-order effect (denoted as $Sobol$) for each parameter, which quantifies the respective effects of parameters on the variance of the response of a mathematical model [452, 453]. In our experiment, the response of the mathematical model in the sensitivity analysis refers to the changes in similarity level (measured by the EU distance between the target and recreated social contact matrices). The larger values of $IVARS$ and $Sobol$ metrics indicate a more noteworthy impact of the corresponding parameter on the similarity of network generation.

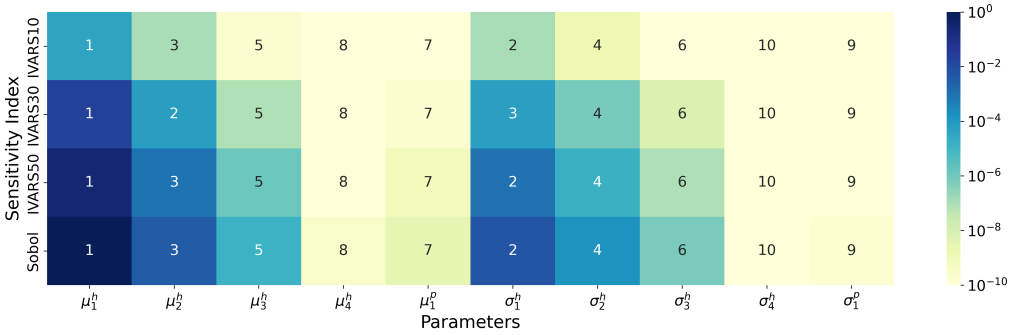
Each metric generates a set of parameter importance rankings (please see the parameter rankings for Belgium in Fig. 5.13(a); Finland in Fig. 5.14(a); Germany in Fig. 5.15(a); Italy in Fig. 5.16(a); Luxembourg in Fig. 5.17(a); and Poland in Fig. 5.18(a)). Based on the aggregated ranks of these sensitivity metrics, we can select the most important parameters in model calibration. As there are no general and definitive criteria in the parameter selection, we select the 6 most important parameters to ensure the flexibility of parameter changes while preserving the computation efficiency of optimising multiple parameters. In this part, we use the Bayesian optimisation algorithm [454] to probe parameter values based on the initial parameter set-ups. Tab. C.1 in appendix C.1 shows the bounds of parameters in the sensitivity analysis and the optimisation process. Tab. 5.5 presents the aggregated ranks of parameter importance, the selected parameters in the model calibration and the enhanced model performance.

Tab. 5.5 shows the parameter rankings for each country based on the aggregated ranks of sensitivity metrics obtained from the global sensitivity analysis. The selected parameters for each country all incorporate μ_1^h and σ_1^h. These two parameters initially describe people's interest in age difference at around 0. This indicates that the homophily effect, related to the age difference around 0, can have a greater impact on the social contact matrices than the other characteristics. In addition, the selected parameters for

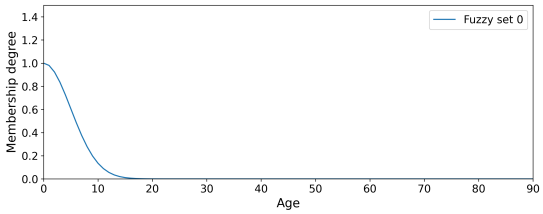
Table 5.5: The parameter ranking and selection for the selected fuzzy sets and the resulted model performance evaluated with EU distance. The parameter ranking is determined based on the aggregated ranks of sensitivity indices including $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$.

Country	Number of fuzzy sets		Initial EU	Parameter rankings	Parameter selection	Improved EU
	Age	Age difference				
Belgium	1	4	3.74	$\mu_h^1 > \sigma_h^1 > \mu_h^2 > \sigma_h^2 > \mu_h^3 > \sigma_h^3 > \mu_p^1 > \mu_h^4 > \sigma_p^1 > \sigma_h^4$	$[\mu_h^1, \sigma_h^1, \mu_h^2, \sigma_h^2, \mu_h^3, \sigma_h^3]$	3.47
Finland	6	4	2.29	$\mu_h^1 > \sigma_h^1 > \mu_h^2 > \sigma_h^2 > \mu_p^4 > \sigma_p^4 > \mu_p^5 > \sigma_p^6 > \mu_p^6 > \sigma_p^5 > \mu_p^3 > \mu_h^3 > \mu_p^2 > \sigma_p^3 > \sigma_h^3 > \sigma_p^2 > \mu_p^1 > \sigma_p^1 > \mu_h^4 > \sigma_h^4$	$[\mu_h^1, \sigma_h^1, \mu_h^2, \sigma_h^2, \mu_p^4, \sigma_p^4]$	2.10
Germany	6	4	2.56	$\mu_h^1 > \mu_p^2 > \sigma_h^1 > \mu_p^3 > \sigma_p^2 > \sigma_p^3 > \mu_h^2 > \mu_p^1 > \mu_p^4 > \sigma_p^4 > \sigma_p^1 > \mu_h^3 > \sigma_h^2 > \mu_p^5 > \sigma_h^3 > \sigma_p^5 > \mu_p^6 > \sigma_p^6 > \mu_h^4 > \sigma_h^4$	$[\mu_h^1, \mu_p^2, \sigma_h^1, \mu_p^3, \sigma_p^2, \sigma_p^3]$	2.28
Italy	6	3	8.59	$\mu_h^1 > \sigma_h^1 > \mu_p^2 > \sigma_p^2 > \mu_p^1 > \sigma_p^1 > \mu_p^3 > \sigma_p^5 > \sigma_p^3 > \mu_h^3 > \mu_h^2 > \mu_p^5 > \sigma_h^2 > \sigma_p^4 > \mu_p^4 > \sigma_h^3 > \mu_p^6 > \sigma_p^6$	$[\mu_h^1, \sigma_h^1, \mu_p^2, \sigma_p^2, \mu_p^1, \sigma_p^1]$	8.34
Luxembourg	6	2	5.07	$\mu_h^1 > \sigma_h^1 > \mu_p^2 > \mu_p^1 > \sigma_p^2 > \mu_p^5 > \mu_p^6 > \sigma_p^1 > \sigma_p^5 > \sigma_p^6 > \mu_h^2 > \mu_p^3 > \mu_p^4 > \sigma_p^3 > \sigma_p^4 > \sigma_h^2$	$[\mu_h^1, \sigma_h^1, \mu_p^2, \mu_p^1, \sigma_p^2, \mu_p^5]$	4.77
Poland	9	4	5.61	$\mu_h^1 > \sigma_h^1 > \mu_h^2 > \mu_p^2 > \sigma_p^2 > \sigma_p^5 > \sigma_h^2 > \mu_p^5 > \mu_p^3 > \sigma_p^7 > \mu_h^4 > \mu_h^3 > \sigma_h^3 > \sigma_p^3 > \sigma_p^4 > \sigma_p^9 > \mu_p^7 > \mu_p^4 > \mu_p^1 > \sigma_p^1 > \mu_p^9 > \sigma_h^4 > \mu_p^8 > \sigma_p^8 > \sigma_p^6 > \mu_p^6$	$[\sigma_h^1 > \mu_h^2, \mu_p^2, \sigma_p^2, \sigma_p^5, \sigma_h^2]$	5.04

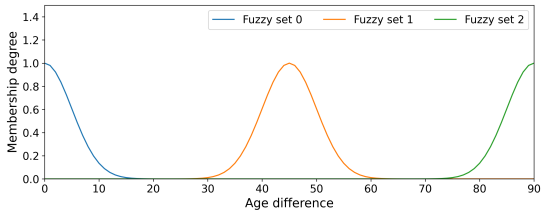
Germany, Italy and Luxembourg include $\mu_p 2$ and σ_p^2 based on the "six fuzzy sets" of age. These two parameters initially describe people's interest in age values around 18 and have a greater impact on the social contact matrices than the other characteristics. The other selected parameters vary with countries and have different meanings considering the number of fuzzy sets. Thus to recreate realistic social networks, we tune the unique parameter set for each country to optimise the model performance, as measured by the similarity between real and simulated social contact matrices.



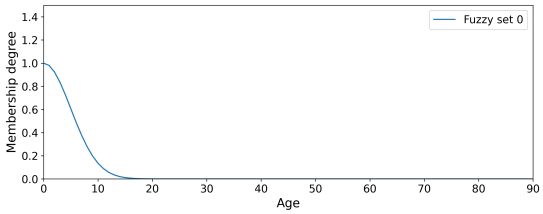
(a) Parameter rankings based on the log-scale parameter sensitivity



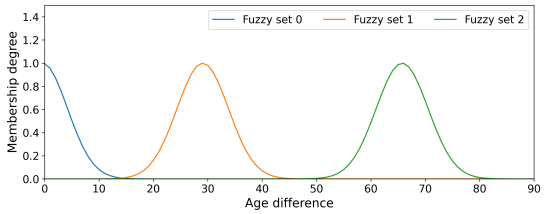
(b) Initial representation of age



(c) Initial representation of age difference



(d) Optimised representation of age

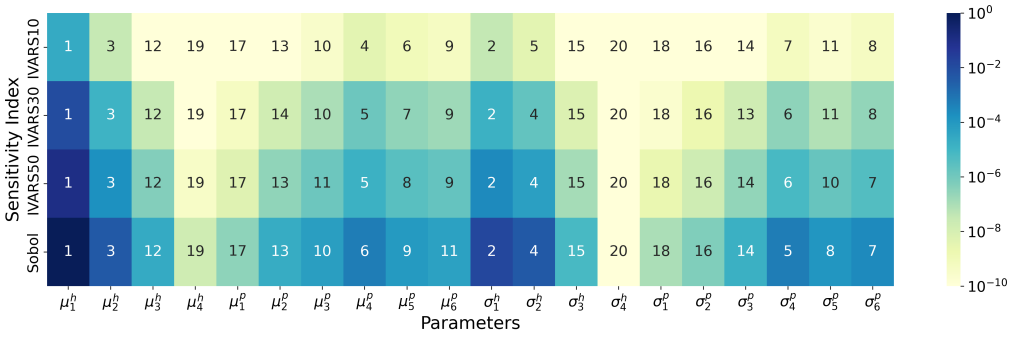


(e) Optimised representation of age difference

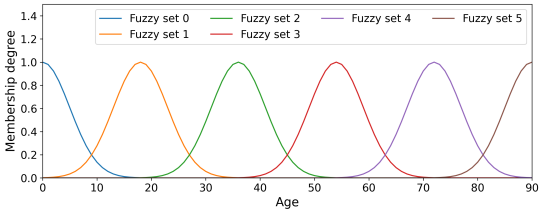
Figure 5.13: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$ (a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Belgium. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h) and parameters of the third fuzzy set of age difference (μ_3^h and σ_3^h) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

Fig. 5.13 (a) shows the ranks and the values of each parameter's sensitivity metrics (in log scale), including the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$ measures for the parameters of the fuzzy sets (one fuzzy set to represent age and six fuzzy sets to represent age difference) for Belgium. The sensitivity metrics enable the same parameter rankings and presents a decreasing impact of the μ_i^h and σ_i^h values as their focus moves from age differences around 0 to larger age differences. The log value of sensitivity levels decreases as $IVARS_{10}$ transits to $IVARS$ at larger scales, which phenomenon can also be seen in the experiments on other countries. More specifically, $\mu_1^h = 0$ and $\sigma_1^h = 5$ are the parameters of the Gaussian Membership function for fuzzy set that initialise the age difference around the focus of 0 within the spread around $[-3\sigma_1^h, 3\sigma_1^h]$ (initialised around the range of [-15,15]). They significantly impact the model performance, as characterised with much higher importance than that of other parameters considering all types of sensitivity indices. In addition, μ_2^h and σ_2^h (parameters for fuzzy set that describes the age difference around 30) and μ_3^h and μ_3^h (parameters for fuzzy set that describes the age difference around 60) also rank high in terms of each sensitivity index. This indicates that tuning the parameters of μ_1^h, σ_1^h, μ_2^h, σ_2^h, μ_3^h and σ_3^h can leads to a more significant change of EU distance. Their optimisation can potentially improve the model performance. We optimise the μ_1^h, σ_1^h, μ_2^h, σ_2^h, μ_3^h and σ_3^h values based on the initialised parameter set up of the fuzzy sets. As shown in Fig. 5.13 (c) and Fig. 5.13 (e), the initialised fuzzy set defined with μ_1^h and σ_1^h, which describes the uncertain induction of age difference around 0, is optimised with a less extensive spread from 5.00 to 4.65 and the transition of the fuzzy set's location to the left from 0.00 to -0.4198. The negative values of age difference do not make sense but induce a smaller membership values based on the Gaussian Membership functions and adjusts the model performance. More specifically, based on the changes mentioned above, people in Belgium tend to assign smaller membership degrees to any age difference over 0 for the first fuzzy set.The second fuzzy set of age difference, initialised at around $\mu_2^h = 30.00$ with a standard deviation at $\sigma_h^2 = 5$, centres around $\mu_2^h = 29.80$ with a less extensive spread at $\sigma_h^2 = 4.74$ after optimisation. The third fuzzy set of age difference, initialised at around $\mu_3^h = 60.00$ with a standard deviation at $\sigma_h^2 = 5$, centres around $\mu_3^h = 65.75$ with a smaller spread at $\sigma_h^2 = 4.79$ after optimisation. The changes mentioned above imply that people in Belgium tend to assign less membership degrees to the age differences at around 0 and higher membership degrees to the age difference at around 29 and 65 (rather than 30 or 60). This indicates that people in Belgium pay more attention to their peers or people around their parents'/grandparents' age. They assign larger membership degrees to these differences and consider them

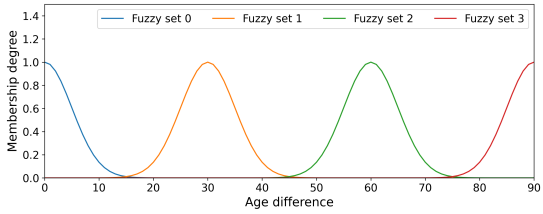
when making social contact. These changes finally lead to a decrease in the EU distance from 3.74 to 3.47 (Tab. 5.5).



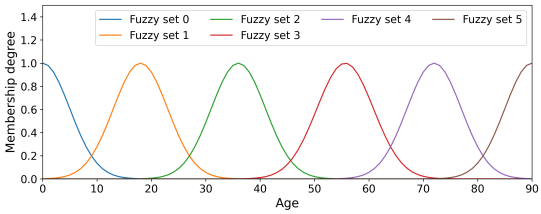
(a) Parameter rankings based on the log-scale parameter sensitivity



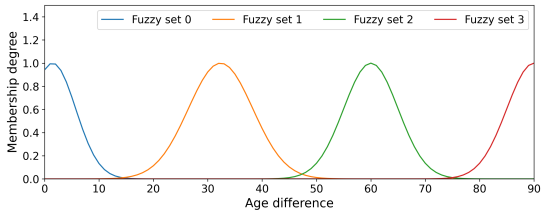
(b) Initial representation of age



(c) Initial representation of age difference



(d) Optimised representation of age



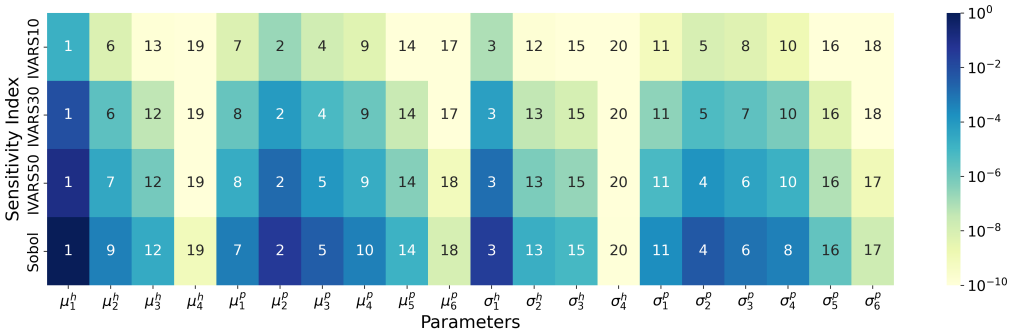
(e) Optimised representation of age difference

Figure 5.14: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Finland. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h) and parameters of the fourth fuzzy set of age difference (μ_4^p and σ_4^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

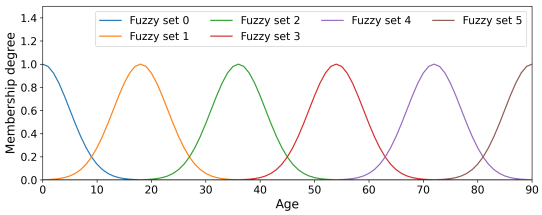
Fig. 5.14 (a) shows the ranks and the log values of sensitivity metrics for the parameters involved in fuzzy sets (six fuzzy sets to represent age and four fuzzy sets to represent age difference) for Finland. The sensitivity metrics enable the same parameter rankings in the first three places and have similar ranks for the less important parameters. Generally, the impact of the μ_i^h and σ_i^h values decreases as their focus moves

from age differences around 0 to larger age differences. In contrast the μ_i^p and σ_i^p values fluctuates without a clear trend. μ_1^h, σ_1^h, μ_2^h and σ_2^h, concerned with the age difference around 0 and 30, have a more significant impact on the model performance and thus are optimised. These four parameters are followed by μ_p^4 and σ_p^4, which describe the age values around 54 and also ranks high considering all the sensitivity metrics. We optimise these parameters based on the initialised parameter set up of the fuzzy sets. As shown in Fig. 5.14 (e), the first fuzzy set for age difference, parameterised with μ_1^h and σ_1^h, is tuned in the optimisation stage with a less extensive spread at 4.25 than the initialised spread at 5 and a transition from the location around 0.00 to the location around 1.47. The second fuzzy set for the age difference, parameterised with μ_2^h and σ_2^h, centres around 32.36 rather than 30 with a more extensive spread at 5.926 than the initialised one at 5 after optimisation. The fourth fuzzy set for the age values, parameterised with μ_p^4 and σ_p^4, transits its focus from 54 to 55.64, with a more extensive spread at 5.21 than the initialised one at 5.00. The changes mentioned above imply the people's interest in age difference at around 1 and age difference around 32 (rather than age difference around 0 or 30) and age at around 55 (rather than 54) in Finland, different from people in Belgium. These changes lead to a decrease of EU distance from 2.29 to 2.10 (Tab. 5.5).

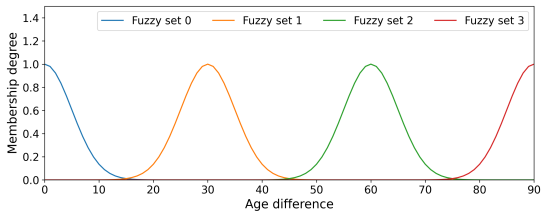
Fig. 5.15 (a) shows the ranks and the log values of sensitivity metrics for each parameter involved in the fuzzy sets (six fuzzy sets to represent age and four fuzzy set to represent age difference) for Germany. Similar with the case of Finland, the sensitivity metrics enable the same parameter rankings in the first three places and presents a decreasing impact of the μ_i^h and σ_i^h values as their focus moves from age differences around 0 to larger age differences. The parameter importance of the μ_i^p and σ_i^p values fluctuates with the transitions of age focuses from young to old. This also indicates that smaller age differences have a larger impact on the model performance in recreating social contact matrices than the larger ones. A more significant impact of the μ_1^h and σ_1^h values on the model performance than the other parameters can be observed. In addition, parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameters of the third fuzzy set of age difference (μ_3^p and σ_3^p) also rank high considering all types of sensitivity indexes. We optimise these parameters and present the changes of fuzzy sets in Fig. 5.15 (c) and Fig. 5.15 (e). The first fuzzy set is optimised with a less extensive spread at 4.34 and a transition from the location around 0.00 to the location around 1.21. The second fuzzy set of age has a more extensive spread at 6.09 and transits the focus from 18.00 to 17.29 after optimisation. The third fuzzy set of age has a more extensive spread at 7.56 and a focus transition from 36.00 to 37.19. These changes indicates people's interest in no



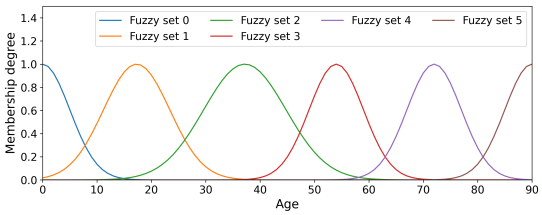
(a) Parameter rankings based on the log-scale parameter sensitivity



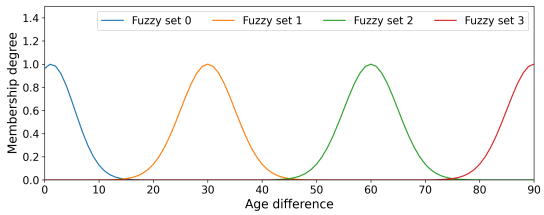
(b) Initial representation of age



(c) Initial representation of age difference



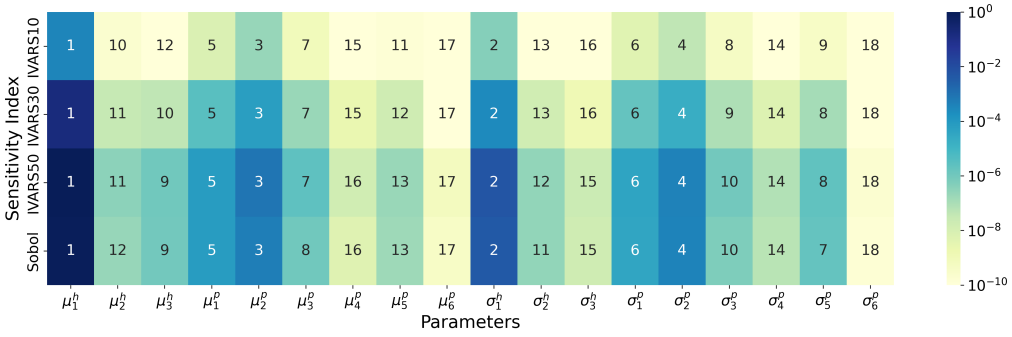
(d) Optimised representation of age



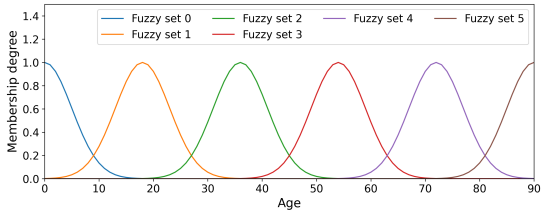
(e) Optimised representation of age difference

Figure 5.15: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$ (a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Germany. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameters of the third fuzzy set of age (μ_3^p and σ_3^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

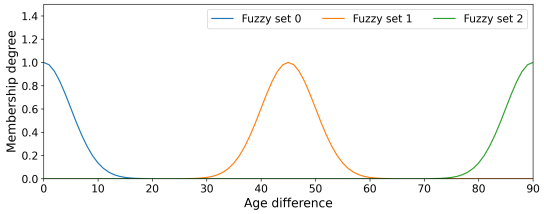
age difference and the age values around 17 and 37 (rather than 18 or 36). In addition, the abovementioned changes lead to a decrease of the EU distance from 2.56 to 2.28 (Tab. 5.5).



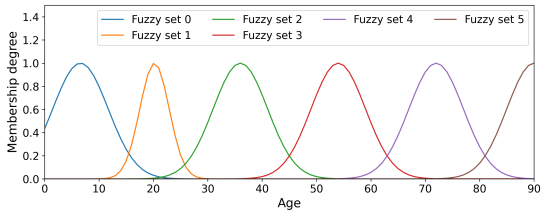
(a) Parameter rankings based on the log-scale parameter sensitivity



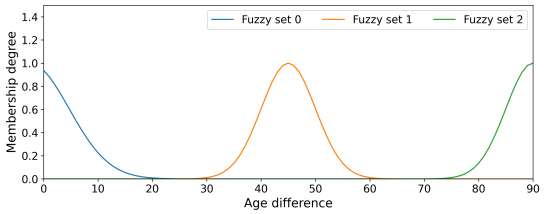
(b) Initial representation of age



(c) Initial representation of age difference



(d) Optimised representation of age



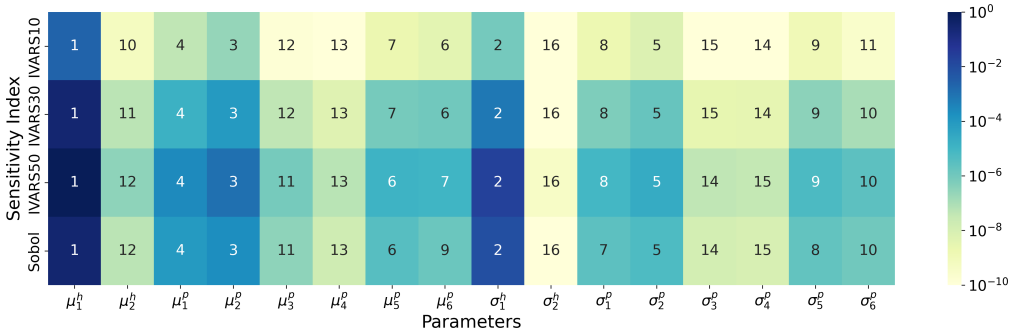
(e) Optimised representation of age difference

Figure 5.16: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Italy. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameters for the first fuzzy set of age (μ_1^p and σ_1^p) and parameters of the second fuzzy set of age (μ_2^p and σ_2^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

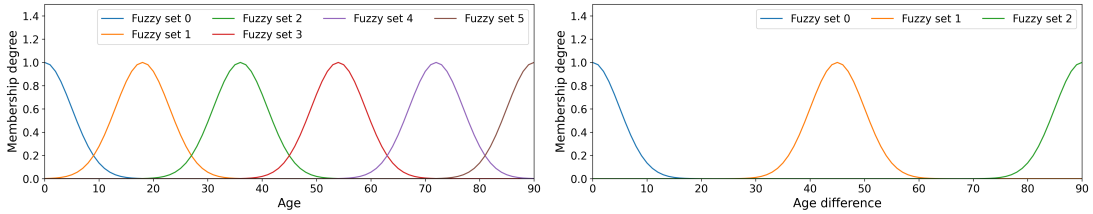
Fig. 5.16 (a) shows the ranks and the log values of sensitivity metrics for each parameter involved in the fuzzy sets (eight fuzzy sets to represent age and one fuzzy set to represent age difference) for Italy. The parameters share the same rankings in the first six places considering each sensitivity index. This implies the validity of

parameter selection in model calibration as the selected parameters all rank high. In addition, the parameters of other parameters have small differences and little impact on model calibration and model performance. Similar with the abovementioned countries, the parameter μ_1^h, σ_1^h and especially μ_1^h significantly impact the model performance, indicating the importance of age difference around 0 (as indicated by parameters for the first fuzzy set of age difference, including σ_1^h and μ_1^h). In addition, the age around 18 and age around 36 (as indicated by the first fuzzy set of age (μ_1^p and σ_1^p) and second fuzzy set of age, related to μ_2^p and σ_2^p) also have greater impact on the decision-making of social contact than the other parameters considering all the sensitivity indexes. We optimise the parameter based on the initialised parameter set up of the fuzzy sets. As shown in Fig. 5.16 (c) and Fig. 5.16 (e), the initialised first fuzzy set of age difference, parameterised with μ_1^h and σ_1^h, describes the uncertain induction of age difference around 0. As shown in Fig. 5.16 (b) and Fig. 5.16 (d), it is optimised with a larger spread of 7.32 and a location transition from location around 0 to location around -2.63, indicating the decrease of membership degrees for the first fuzzy set. In addition, the first fuzzy set for age is optimised with a slightly more extensive spread at 5.09 and a significant focus transition from 0.00 to 6.59. The second fuzzy set for age, initialised around 18 with a spread at 5, is optimised with the transition of focus to 20.21 and a less extensive spread at 2.70. These changes imply the people's interest in age difference around 0 and the age values around 6 and 20 (rather than 0 or 18). The abovementioned changes decrease the EU distance from 8.59 to 8.34 (Tab. 5.5).

Fig. 5.17 (a) shows the ranks and the log values of sensitivity metrics for each parameter involved in the fuzzy sets (six fuzzy sets to represent age and two fuzzy set to represent age difference) for Luxembourg. Considering each sensitivity index, we have same parameter ranks in the first five places and similar parameter ranks following these five parameters. The μ_1^h and σ_1^h value, concerned with the representation of zero age difference, significantly impacts the model performance and indicates the importance of age difference around 0 in the decision-making of social contact. The parameters for the first fuzzy set of age (μ_1^p), the second fuzzy set of age (μ_2^h and σ_2^h) and the fifth fuzzy set of age (μ_5) rank afterwards also have a greater impact on model performance than the other parameters. We optimise these parameters based on the initialised parameter set up for the fuzzy set of age difference. As shown in Fig. 5.17 (c) and Fig. 5.17 (e), the first fuzzy set for age difference is optimised with a more extensive spread of 7.60 and a focus transition from 0 to a slightly lower value at -0.45, leading to slightly smaller membership values of age differences and indicating less interest from people.

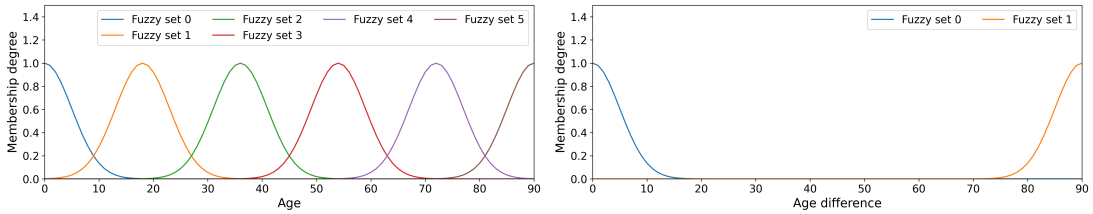


(a) Parameter rankings based on the log-scale parameter sensitivity



(b) Initial representation of age

(c) Initial representation of age difference



(d) Optimised representation of age

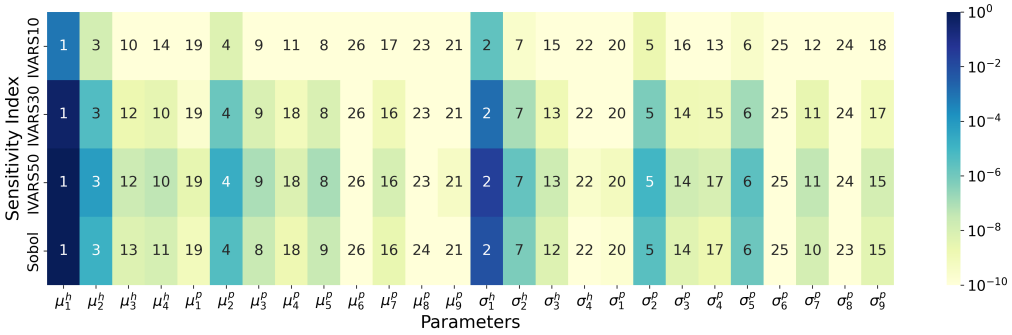
(e) Optimised representation of age difference

Figure 5.17: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Luxembourg. In (a), parameters for the first fuzzy set of age difference (μ_1^h and σ_1^h), parameter for the first fuzzy set of age (μ_1^p), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameter for the fifth fuzzy set of age (μ_5^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

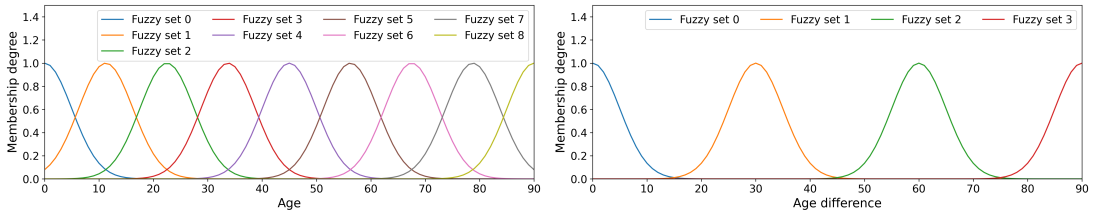
Similarly, the first fuzzy set for age is optimised with a focus transition from 0.00 to a negative value at −1.23, which decreases the membership degrees of age values to the first fuzzy set of age after optimisation. The second fuzzy set for age is optimised with a smaller spread at 4.69 and a significant focus transition from 18 to 18.15. The fifth fuzzy set for age, initialised around 72.00, is optimised with the transition of focus to 73.12. These changes imply the people's decreased interest in age difference around 0 and the age values around 0. They also indicate the people's interest in age around 18 and 73 (similar with the unoptimised values around 18 and 72). The abovementioned changes decrease the EU distance from 5.07 to 4.77 (Tab. 5.5), implying the necessity of optimising membership functions for fuzzy representation of features.

Fig. 5.18 (a) shows the ranks and the log values of sensitivity metrics for each parameter involved in the fuzzy sets (nine fuzzy sets to represent age and four fuzzy set to represent age difference) in Poland. The σ_1^h values, concerned with the representation of age difference around 0, significantly impact the model performance and indicate their importance in the parameter optimisation. The μ_h^2 and σ_h^2 for the second fuzzy set of age difference, among the four fuzzy sets of age difference, rank high in terms of parameter importance. This is the same with the cases of Belgium and Finland. In addition, the parameters for the second fuzzy set of age at around 22 (μ_p^2 and $\sigma_p 2$) and the parameter for the fifth fuzzy set of age around 45 (σ_p^5) also rank high considering all the sensitivity metrics. We optimise the parameters based on the initialised parameter set up of the fuzzy sets. As shown in Fig. 5.18 (c) and Fig. 5.17 (e), the fuzzy set for age difference, parameterised with μ_1^h and σ_1^h, is further optimised with a similar spread at 5.18. The second fuzzy set of age difference is optimised with a significant focus transition from 30 to 26. This implies people's interest in age difference around 26 rather than 30. In addition, the second fuzzy set of age transits its focus from 11.33 to 11.70, which is a slight adjustment of people's interest around the age at 11. The fifth fuzzy set of age is optimised with a significant change in spread, which decreases from 5.00 to 1.00, reaching the lower bound of parameter optimisation. This implies that the age closer to 45 can be assigned with a higher membership degree, indicating the people's interest in the age values that are close to 45. The abovementioned changes decrease the EU distance from 5.61 to 5.04 (Tab. 5.5).

To conclude, the optimised parameters for each country all incorporate the μ_1^h and σ_1^h values. They describe people's interest in age difference at around 0 and are optimised to represent small age differences close to 0. This implies that people in each country have more interest in age differences around these small values and captures the homophily

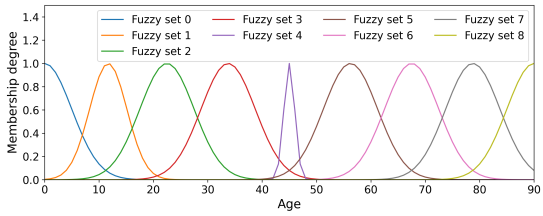


(a) Parameter rankings based on the log-scale parameter sensitivity

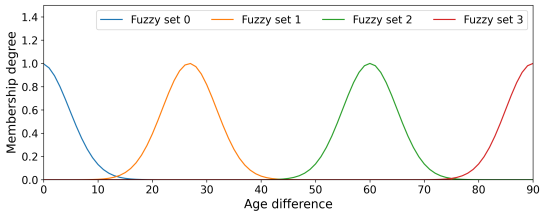


(b) Initial representation of age

(c) Initial representation of age difference



(d) Optimised representation of age



(e) Optimised representation of age difference

Figure 5.18: The parameter sensitivity level (in log scale) and the corresponding rankings for parameters of fuzzy sets based on the $IVARS_{10}$, $IVARS_{30}$, $IVARS_{50}$ and $Sobol$(a), the initialised fuzzy representation of age and age difference ((b) and (c)) and the optimised fuzzy representation of age and age difference ((d) and (e)) for Poland. In (a), parameter for the first fuzzy set of age difference (σ_1^h), parameters for the second fuzzy set of age difference (μ_2^h and σ_2^h), parameters for the second fuzzy set of age (μ_2^p and σ_2^p) and parameter for the fifth fuzzy set of age (μ_5^p) rank high considering all types of sensitivity indexes. Therefore, they are optimised in model calibration. The optimised fuzzy representation principles are shown in (d) and (e).

effect presented in the real-world social contact matrices (Fig. 5.5, Fig. 5.6, Fig. 5.7, Fig. 5.9, Fig. 5.8 and Fig. 5.10). In addition, the optimised parameters for Belgium and Finland both include the parameters for the second fuzzy set of age difference, but each centring around different values at around 24 and 12. They describe people's different interests in age differences and capture the social contacts across age groups (Fig. 5.5 and Fig. 5.6). The optimised parameters in case of Finland and Italy include the parameter for the third fuzzy set of age, each centring around the age of 28 and 22. They focus on people's preferences for age and captures the popularity of the age groups between 20 and 40, as presented in the respective social contact matrices (Fig. 5.6 and Fig. 5.8). Overall, the above-mentioned selection of fuzzy sets and the selection and optimisation of parameters for fuzzy sets increases the accuracy of recreating realistic social networks, as indicated by a decreasing EU distance between the target and real social contact matrices (Tab. 5.5). This implies that the increasingly flexible set-ups of fuzzy representation principles increase the complexity of modelling DT-CNS models and enable to recreate more realistic social networks through the appropriate selection and optimisation of fuzzy sets.

5.3.5 Network Representation

In this section, we analyse the networks generated by the calibrated models from five perspectives, including the features and the related preferences (section 5.3.5.1), social contact matrices (section 5.3.5.2), degree distributions (section 5.3.5.3), clustering coefficient distributions (section 5.3.5.4) and shortest path length distributions (section 5.3.5.5). Tab. 5.6 shows the number of connected and unconnected nodes and the information about the degree distributions, clustering coefficient distributions and the shortest path length distributions that vary depending on the values of the features and rules employed. Specifically, we also consider the non-existing shortest paths between the node pairs and assume their path lengths as 90, which is the upper limit that cannot be achieved by any real path. Generally, all the nodes get connected with the paradigms of each country. Compared with other countries, Finland has a lower average node degree value, a lower clustering coefficient and more non-existing (fake) paths between the node pairs. Italy and Poland have higher average node degrees and fewer fake paths. These differences can result from different numbers of edges and heterogeneous features and rules. To better understand these characteristics, we calculate the social contact matrices of network simulations and explore the differences in degree distributions, clustering coefficient distributions and shortest path length distributions for females and males at

different ages and for different countries.

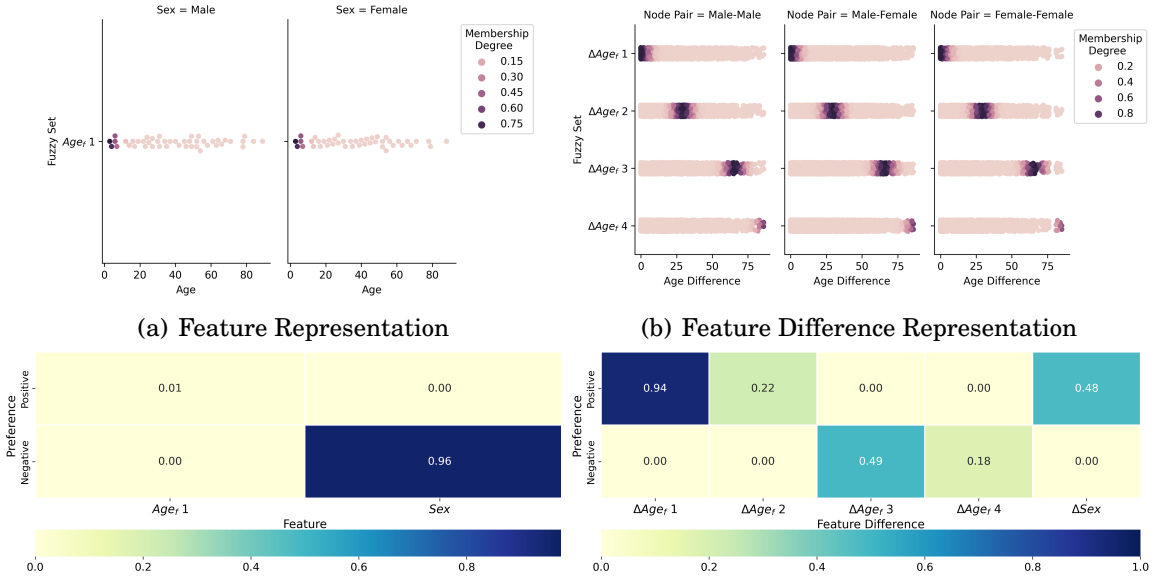
Table 5.6: Topological information of the generated networks.

Country	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
Belgium	90	0	614	13.64	11.84	34	0	0.54	0.31	1.00	0.00	264	8.02	21.79	90	1
Finland	90	0	336	7.47	6.94	22	0	0.32	0.27	1.00	0.00	2287	52.34	43.46	90	1
Germany	90	0	462	10.27	9.8	35	0	0.42	0.30	1.00	0.00	1155	27.57	39.75	90	1
Italy	90	0	770	17.11	13.67	45	0	0.52	0.27	1.0	0.00	89	4.14	12.97	90	1
Luxembourg	90	0	687	15.27	14.20	43	0	0.51	0.33	1.00	0.00	1449	33.81	42.31	90	1
Poland	90	0	793	17.62	11.45	45	0	0.51	0.15	0.83	0.00	89	4.08	12.98	90	1

5.3.5.1 Features, Feature Differences and Social Preferences

In the experiments, we represent the features and feature differences based on the calibrated fuzzy representation principles in section 5.3.4.2 for each country, including Belgium (Fig. 5.19), Finland (Fig. 5.20), Germany (Fig. 5.21), Italy (Fig. 5.22), Luxembourg (Fig. 5.23) and Poland (Fig. 5.24).

As shown in Fig. 5.19 (a) and (b), people's interest in age features and age differences can be respectively represented by one fuzzy set and four fuzzy sets. A single age value, either female or male, is represented by a membership degree to the first fuzzy set of age (Age_f1), where smaller ages around 0 can have higher membership values. In addition, the representation of the age difference between any node pair can also be unfolded into a vector that is composed of four membership degrees to the four fuzzy sets (Denoted by $\Delta Agei$ in Fig. 5.19 (b)). A large number of nodes, centring around the age of 22 and 65 are respectively assigned with higher membership degrees to the second and third fuzzy set of age (ΔAge_f2 and ΔAge_f3), which captures people's interests and is in line with the parameter rankings in the model calibration process (Fig. 5.13(d)). The representation of age features and age differences, considering "males/females" and "male-male/female-female/male-female" relations, does not vary much, except for a few differences in density due to the differences in age distributions of females and males (Fig. 5.2 (a)). In Fig. 5.19 (c) and (d), the age features and age differences, unfolded by these fuzzy sets, are preferred either positively or negatively based on a preference weight. Males and the age difference around 0 (represented by the first fuzzy set) are strongly preferred with a preference weight that is close to 1. In addition, people in Belgium also have a positive preference for age differences around 29 ($\Delta Age2$, the second fuzzy set of age difference; Fig. 5.13 (d)) and a negative preference for age differences



(a) Feature Representation

(b) Feature Difference Representation

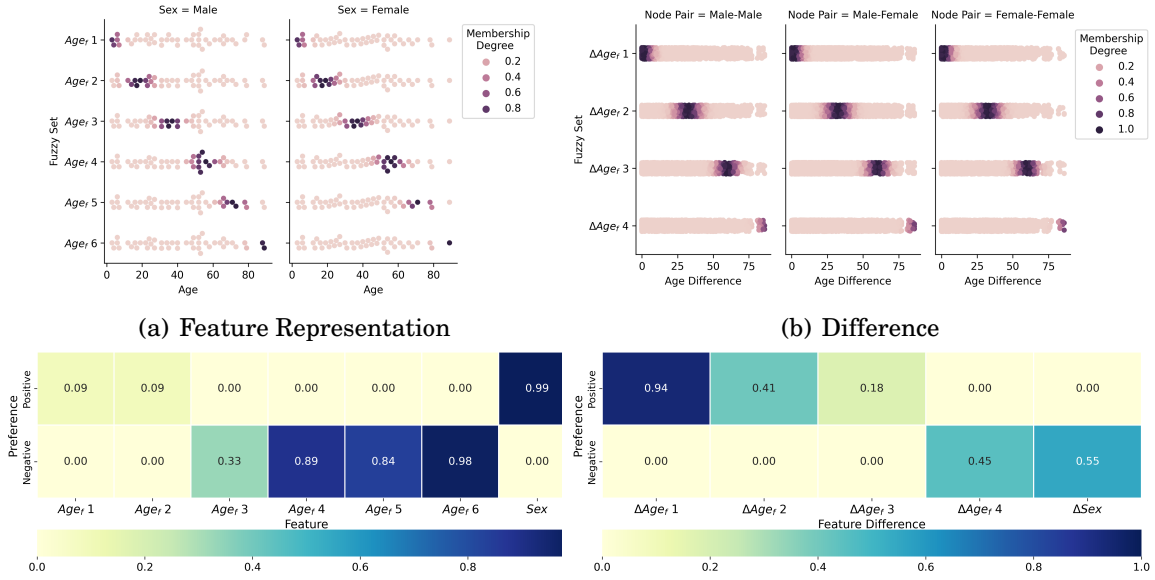
(c) Feature Preference

(d) Feature Difference Preference

Figure 5.19: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Belgium based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), $Age_f i$ and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), $\Delta Age_f i$ and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).

around 65 ($\Delta Age2$, the second fuzzy set of age difference; Fig. 5.13 (d)) and 90. The corresponding preference weights are over 0.15, significantly higher than the preference for age features shown in Fig. 5.19 (c). This indicates that people in Belgium have more interests in age differences and sex features, where they strongly prefer to be connected with males at a similar age in network formation.

As shown in Fig. 5.20 (a) and (b), people's interest in age features and age differences can be each represented by six fuzzy sets and four fuzzy sets. A greater number of nodes are distributed around the age of 50 and are assigned with higher membership degrees to the fourth age fuzzy set ($Age_f 4$, centring around the age of 54). In addition, a significant number of node pairs have an age difference around the age of 30 and 30, which have higher membership degrees to the second and the third fuzzy sets of age difference ($\Delta Age_f 2$ and $\Delta Age_f 3$; Fig. 5.14 (e)). The abovementioned phenomenon indicates people's interests in age values around 50 and the age differences around 0,



(a) Feature Representation

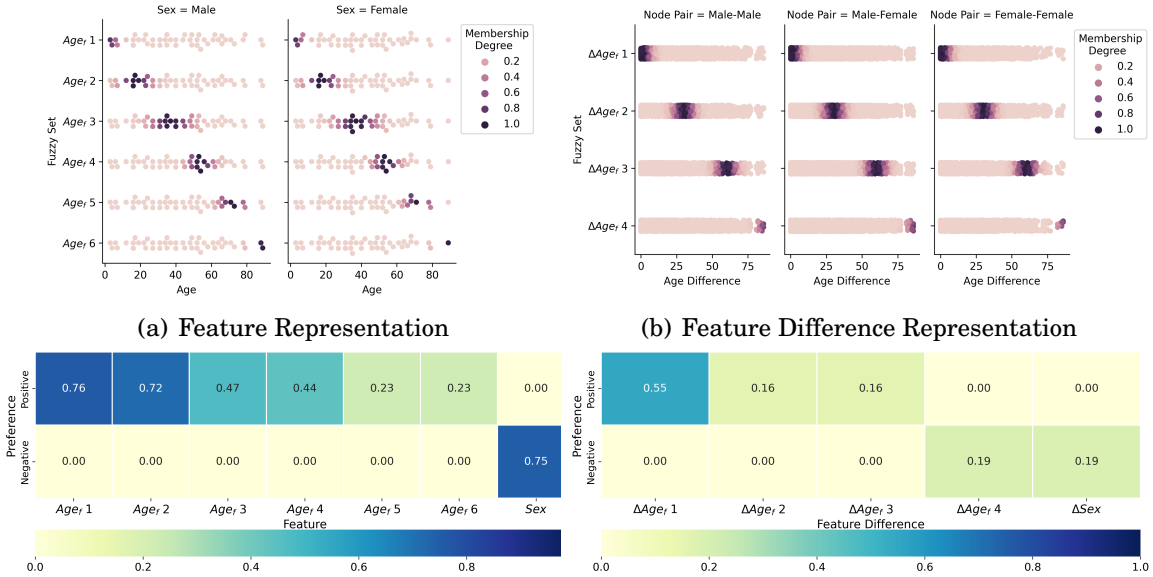
(b) Difference

(c) Feature Preference

(d) Feature Difference Preference

Figure 5.20: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Finland based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0, 1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), $Age_f i$ and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), $\Delta Age_f i$ and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature).

30 and 60. Similar to the case of Belgium, the distributions of the membership degrees do not vary much with the sex features due to the similar age distributions of females and males (Fig. 5.2 (b)). In Fig. 5.20 (c) and (d), females and age differences around 0 are strongly preferred with a preference weight over 0.90. In contrast, people in Finland have a strong negative preference for ages around 54, 72 and 90, with a preference weight over 0.85. This indicates that people have more interest in age differences around 0 and seek connections with people of a similar age, in contrast with the attention and negative preference for age values around 54. In addition, people in Finland also show positive preference for age difference around 32 (ΔAge_f^2, the second fuzzy set of age difference) and age difference around 60 (ΔAge_f^3, the third fuzzy set of age difference) and negative preference for age around 36 (Age_f^3, the third fuzzy set of age). They are characterised with a preference weight over 0.15, which, multiplied with greater membership degrees to the corresponding fuzzy set, can also have a great impact on the interactions.



(a) Feature Representation

(b) Feature Difference Representation

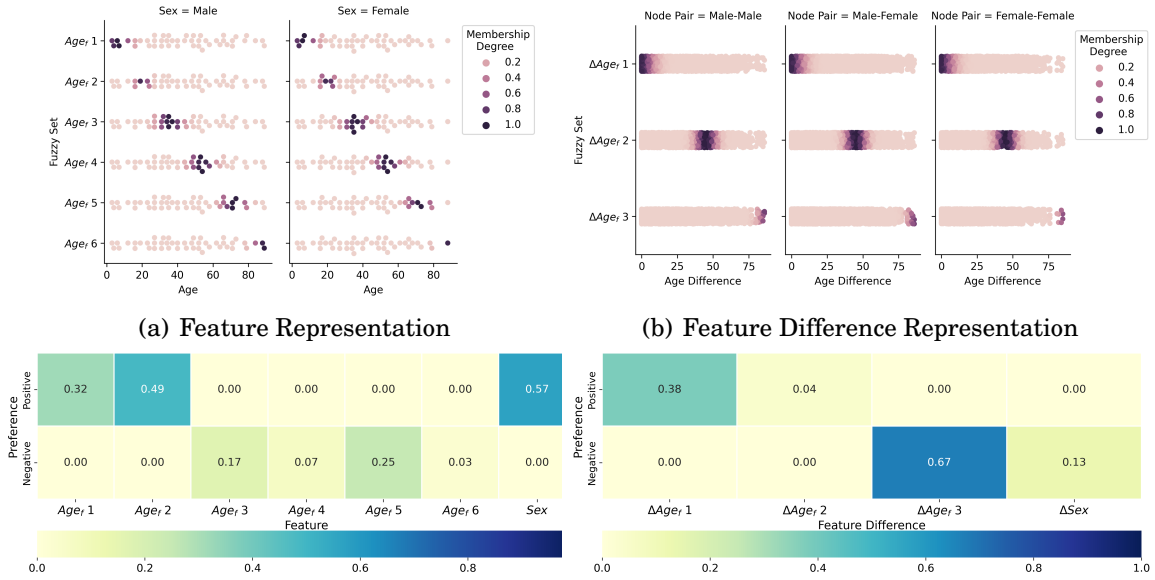
(c) Feature Preference

(d) Feature Difference Preference

Figure 5.21: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Germany based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), $Age_f i$ and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), $\Delta Age_f i$ and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature)..

As shown in Fig. 5.21 (a) and (b), people's interest in age features and age differences can also be represented by six fuzzy sets and four fuzzy sets. Similar to the cases of Finland, a greater number of nodes are distributed around the age range of $[25-60]$ and have higher membership degrees to the corresponding fuzzy set, including the second and the third fuzzy set of age ($Age_f 2$ and $Age_f 3$, each centring around 37 and 54; Fig. 5.2 (e)). There are also a significant number of node pairs allocated around the age difference of 30 and 60, characterised with higher membership degrees to the second and the third fuzzy sets of age difference ($\Delta Age_f 2$ and $\Delta Age_f 3$; Fig. 5.15 (e)). The abovementioned phenomenon indicates people's interests in age values and age differences. Similar to the cases of the abovementioned countries, the distributions of the membership degrees do not vary much with the sex features due to the similar age distributions of females and males (Fig. 5.2 (b)). In Fig. 5.21 (c) and (d), males and age values around 0, 17 and 37 (each related to the first, second and third fuzzy sets of age, denoted by $Age_f 1$, $Age_f 2$

and Age_f3) are strongly preferred with a preference weight over 0.70. This is followed by relatively weaker positive preferences for other age values, positive preferences for age differences around 30 and 60 (each related to the second and third fuzzy sets of age difference, denoted by ΔAge_f2 and ΔAge_f3) and the negative preference for age difference around 90 (related to ΔAge_f4, the fourth fuzzy set of age) and the same sex features. Both the heterogeneous patterns of feature representation and preferences can influence the network formation process, which makes it hard to find a decisive interaction rule and detriments the interpretability of resulting network patterns.



(a) Feature Representation

(b) Feature Difference Representation

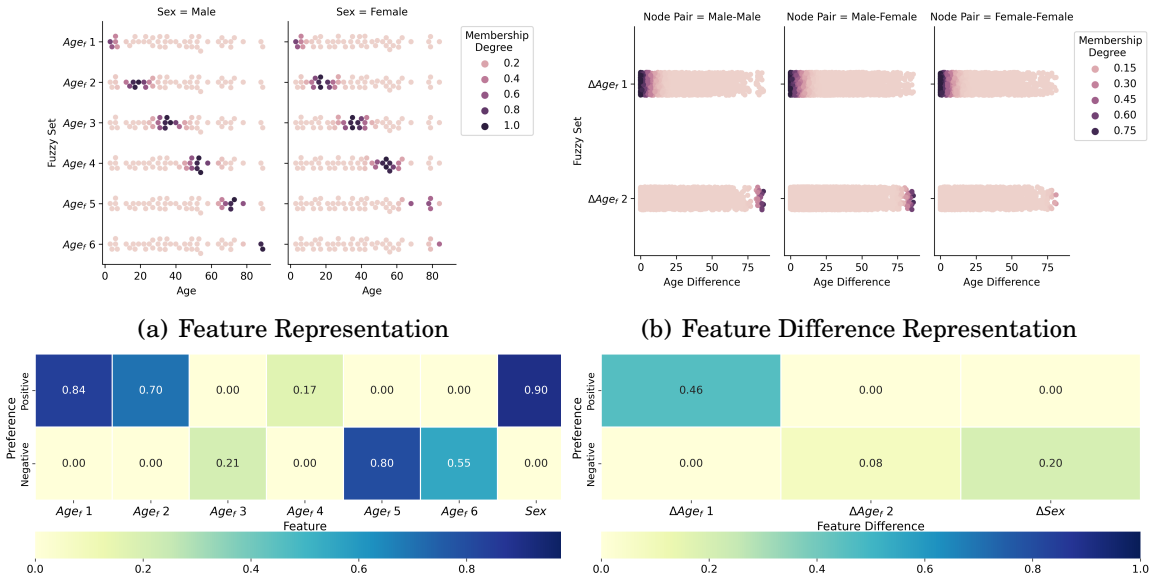
(c) Feature Preference

(d) Feature Difference Preference

Figure 5.22: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Italy based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), Age_fi and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), ΔAge_fi and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature)..

As shown in Fig. 5.22 (a) and (b), people's interest in age features and age differences can also be represented by six fuzzy sets and three fuzzy sets. Similar to the cases of Finland and Germany, a greater number of nodes are distributed around the age range of $[25-60]$ and have higher membership degrees to the corresponding fuzzy set, including the second and the third fuzzy set of age (Age_f2 and Age_f3, each centring around

36 and 54; Fig. 5.2 (d)). In contrast with the cases of Finland and Germany, people in Italy pay more attention to the age difference of 45 and thus assign higher membership degrees to the second fuzzy set of age difference $\Delta Age_f 2$ (Fig. 5.16 (e)). In Fig. 5.22 (c) and (d), people in Italy have a stronger positive preference, characterised by a preference weight around 0.50 for females, age values around 20 ($Age_f 2$, the second fuzzy set of age) and age differences around 0 ($\Delta Age_f 1$, the first fuzzy set of age difference). In addition, they also have a negative preference for age difference around 90 ($\Delta Age_f 3$, the third fuzzy set of age difference) with a preference weight at 0.57, which, however, have limited impact on the network formation because smaller number of node pairs have higher membership degrees to the third fuzzy set of age difference. People in Italy show negative preferences for age values around 36 and 72 with slightly weaker preference weights around 0.20, which, multiplied with people's interests and feature distributions for the third and fifth fuzzy set of age ($Age_f 3$ and $Age_f 5$), can also have a significant impact on network formation.



(a) Feature Representation

(b) Feature Difference Representation

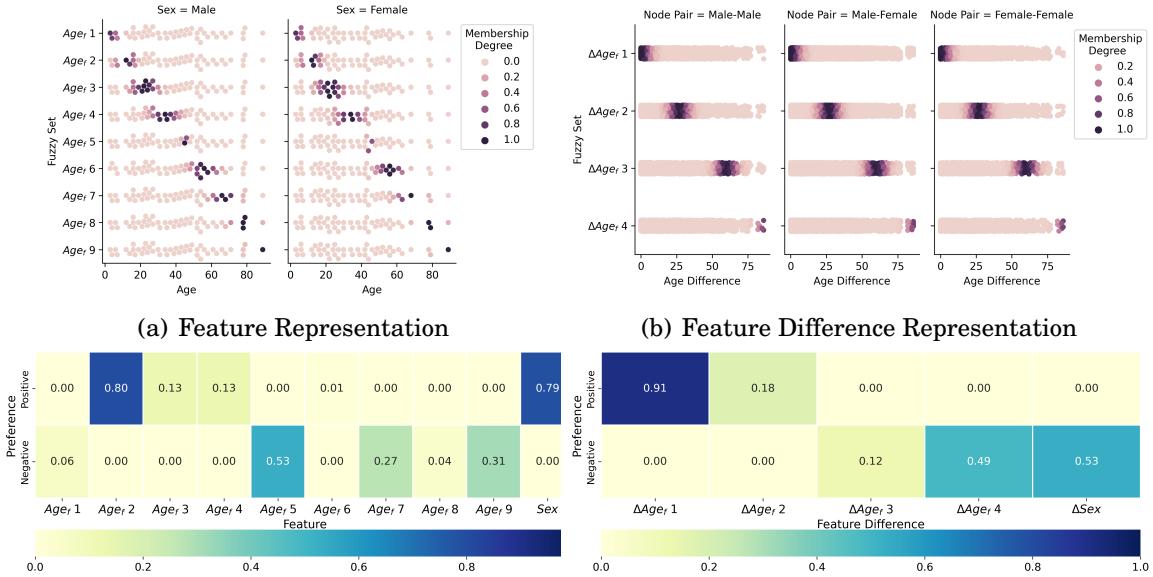
(c) Feature Preference

(d) Feature Difference Preference

Figure 5.23: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Luxembourg based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0, 1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), $Age_f i$ and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), $\Delta Age_f i$ and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature)..

As shown in Fig. 5.23 (a) and (b), people's interest in age features and age differences can also be represented by six fuzzy sets and two fuzzy sets. Similar to the cases of Finland, Germany and Italy, a significant number of nodes are distributed around the age 40 and have higher membership degrees to the corresponding fuzzy sets, including the third and fourth fuzzy set of age (each denoted by Age_f3 and Age_f4). In contrast with the cases of the abovementioned countries, people in Luxembourg pay more attention to the age difference around 0 and thus assign higher membership degrees to the corresponding first fuzzy set of age difference ΔAge_f1 (Fig. 5.17 (e)). In Fig. 5.23 (c) and (d), people in Luxembourg have a strong positive preference, characterised by a preference weight over 0.70, for females, age values around 0 (Age_f1, the first fuzzy set of age) and age values around 18 (Age_f2, the second fuzzy set of age). They have a strong negative preference around 72 (Age_f5, the fifth fuzzy set of age), with a preference weight of 0.80. In addition, they also have a positive preference for age difference around 0 (ΔAge_f1, the first fuzzy set of age difference), which, despite its slightly lower preference weight at 0.46, can also have a significant impact on the network formation considering people's interests in similar ages (as shown in Fig. 5.23 (b)).

As shown in Fig. 5.24 (a) and (b), people's interest in age features and age differences can be represented by nine fuzzy sets and four fuzzy sets. Different from the abovementioned countries, people in Poland have a more extensive interest in age values over various ranges. Based on the distribution of age features (Fig. 5.2 (f)), a significant number of nodes have higher membership degrees to the third, fourth, sixth and seventh fuzzy sets (each denoted by Age_f3, Age_f4, Age_f6 and Age_f7, centring around 22, 33, 56 and 67). Similar to the cases of Belgium, Finland and Germany, a significant number of people have an age difference of around 26 and 60, characterised by higher membership degrees to the second and the third fuzzy set of the age difference (denoted by ΔAge_f2 and ΔAge_f3; Fig. 5.18 (f)). In Fig. 5.24 (c) and (d), people in Poland have a stronger positive preference, characterised by a preference weight over 0.75, for females, age values around 0 (Age_f1, the first fuzzy set of age) and age differences around 0 (ΔAge_f1, the first fuzzy set of age difference). They also have a significant negative preference, with a preference weight around 0.50, for age around 45, (Age_f5, the fifth fuzzy set of age) and age difference around 90 ($\Delta Age_4 6$, the sixth fuzzy set of age difference). In addition, people in Poland have a slightly positive preference for age difference around 30 (ΔAge_f2, the second fuzzy set of age difference) and similar sex features. These two preferences can also have a significant impact on the network formation, considering people's great interests in the similar ages and the specific differences between females



(a) Feature Representation

(b) Feature Difference Representation

(c) Feature Preference

(d) Feature Difference Preference

Figure 5.24: The representation of age features for females and males (a) and the representation of age features between female-male, male-male and male-female nodes (b) in Poland based on the membership degrees to fuzzy sets. The corresponding preferences (positive or negative) and the weights of preference (ranging within $[0,1]$) for each fuzzy membership degree is presented in (c) and (d). In (c), $Age_f i$ and Sex on the x-axis each denotes the i_{th} fuzzy set for age and the crisp representation of sex with binary values (1 for female and 0 for male). In (d), $\Delta Age_f i$ and ΔSex represents the i_{th} fuzzy set for age difference and the crisp representation of sex differences with binary values (1 for the same sex feature and 0 for different sex feature)..

and males (as shown in Fig. 5.24 (b)).

To conclude, people in each country have different interests and preferences in social contact. They generally have great interests or strong preferences in the sex features, age differences around 0, 30 and 60. The variety of heterogeneous features and interaction rules add to the complexity of network formation but gives us a clue to decipher the complex social network patterns.

5.3.5.2 Social Contact Matrices

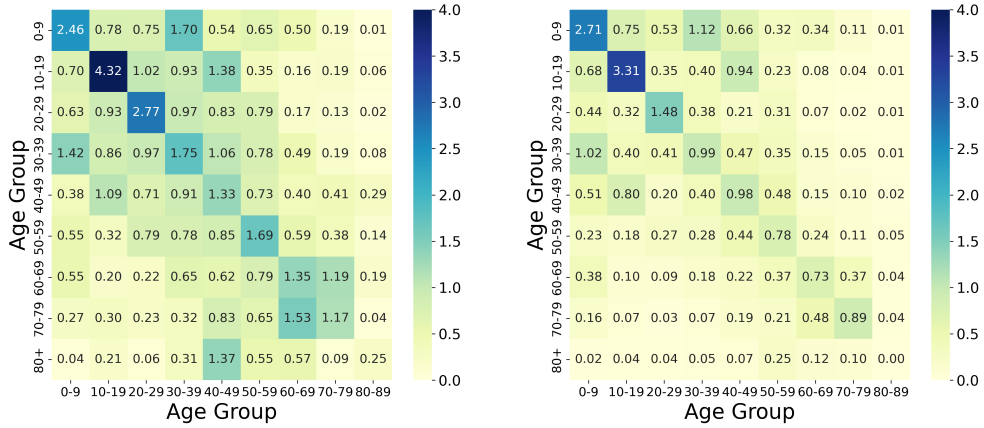
Based on the features and preferences presented in section 5.3.5.1 and towards the target social contact matrices presented in Fig. 5.25, we can generate social network simulations and calculate the corresponding social contact matrices based on the age features for Belgium (Fig. 5.26 (a)), Finland (Fig. 5.26 (b)), Germany (Fig. 5.26 (c)), Italy (Fig. 5.26 (d)), Luxembourg (Fig. 5.26 (e)) and Poland (Fig. 5.26 (f)). The social

contact matrices generally capture the characteristics of the real ones, including people's connections with similar ages and people with an age difference of around 30. This results from the great interest and strong preference for the corresponding fuzzy representation of age differences (e.g. $\Delta Age_f 1$ for the first fuzzy set of age difference around 0 and $\Delta Age_f 2$ for the second fuzzy set of age difference around 30 in Belgium (Fig. 5.19), Finland (Fig. 5.20), Germany (Fig. 5.19) and Poland (Fig. 5.19)). Among these countries, Finland has less average social contacts between different ages than the other countries. In addition, the average social contact number approaches zero for people over 80 in Finland, which leads to lower average node degrees and clustering coefficient and a greater number of non-existing paths between nodes. In the social contact matrices of the simulated social networks, there is no social contact between people over 80 and people younger than 80, even though there is a smaller number of social contacts in the real contact matrix. This is because people have a strong negative preference for older nodes around the age of 55, 72 and 90 (each related to the fourth, fifth and sixth fuzzy set of age differences; Fig. 5.20 (c)). However, this is unrealistic and more detailed set-ups related to social contacts are required to model the limited number of connections developed by specific node pairs.

5.3.5.3 Degree Distribution

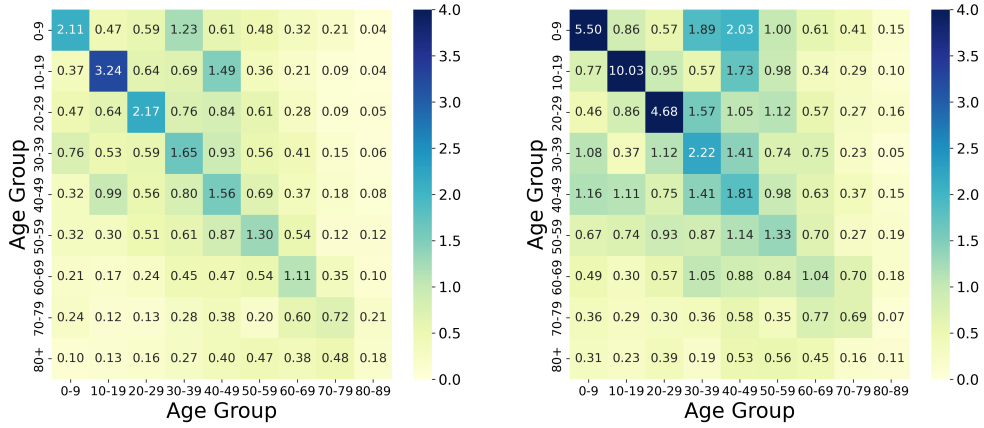
We explore the distributions of node degrees for females, males and both females and males in different ages based on the network simulations of each country, including Belgium (Fig. 5.27), Finland (Fig. 5.28, Germany (Fig. 5.29), Italy (Fig. 5.30), Luxembourg (Fig. 5.31) and Poland (Fig. 5.32). The degree distributions generally have a left tail shape, with a significant number of nodes' degrees fall in [0-10]. The degree distributions of the age groups, shown in each cell of the matrix, form a pyramid shape, where younger nodes generally have a higher node degrees than the older ones. This results from the interests and preferences for young ages and smaller age differences.

We explore the differences in degrees distributions in each country. We find that the people with preferred sex and age features have higher node degrees. For example, in Belgium and Germany, young males around the age of $[0-50]$ have higher node degrees (at around $10-30$) than females, whose node degrees fall in $[0-9]$. This can be caused by the strong preference for males and young age close to 0 (Fig. 5.19 (c) and Fig. 5.21 (c)). This is in contrast with the case of Finland, young females around the age of $[0-50]$ have higher node degrees than males due to the preference for young and female features (Fig. 5.20 (c)). In addition, people in dense age groups, with preferred



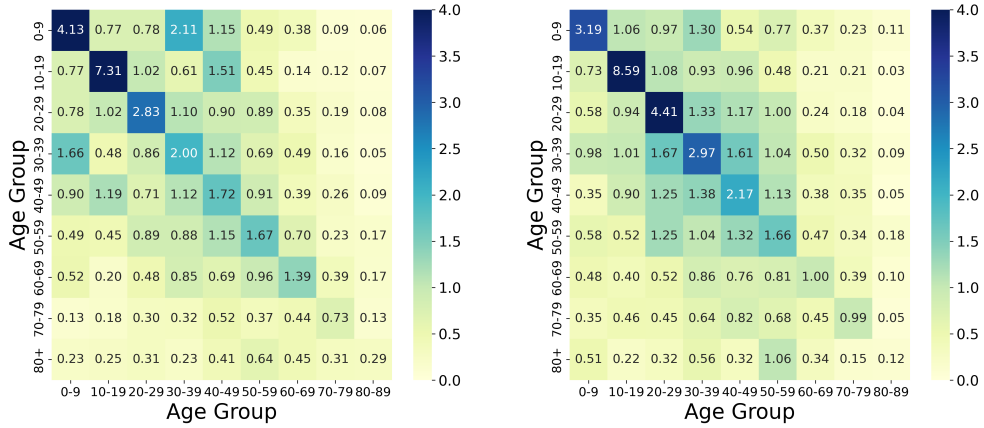
(a) Belgium

(b) Finland



(c) Germany

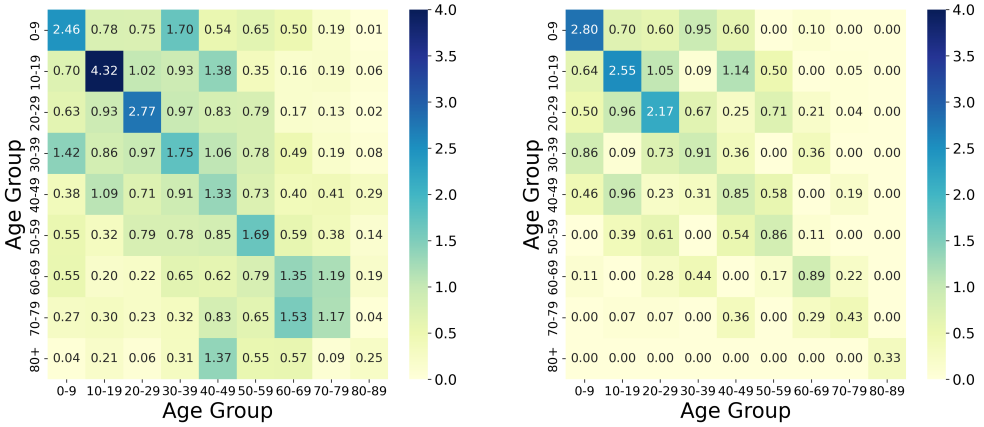
(d) Italy



(e) Luxembourg

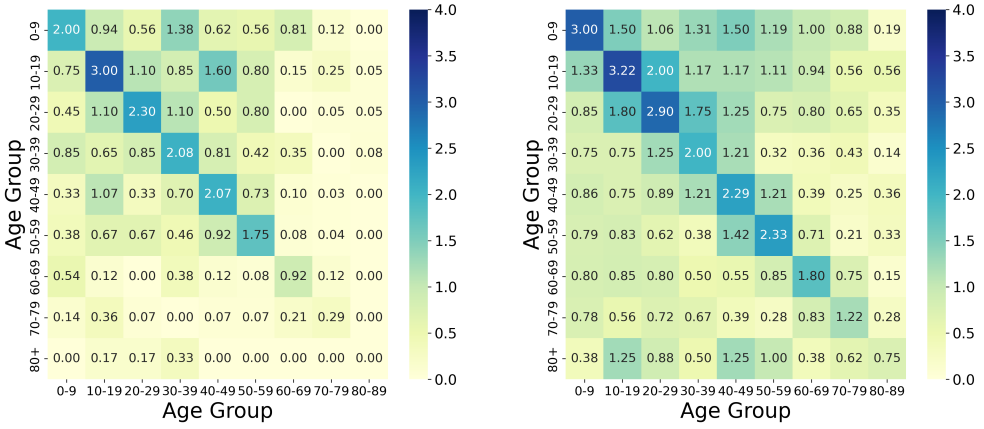
(f) $HN-A_f-S$

Figure 5.25: The target social contact matrices of (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland.



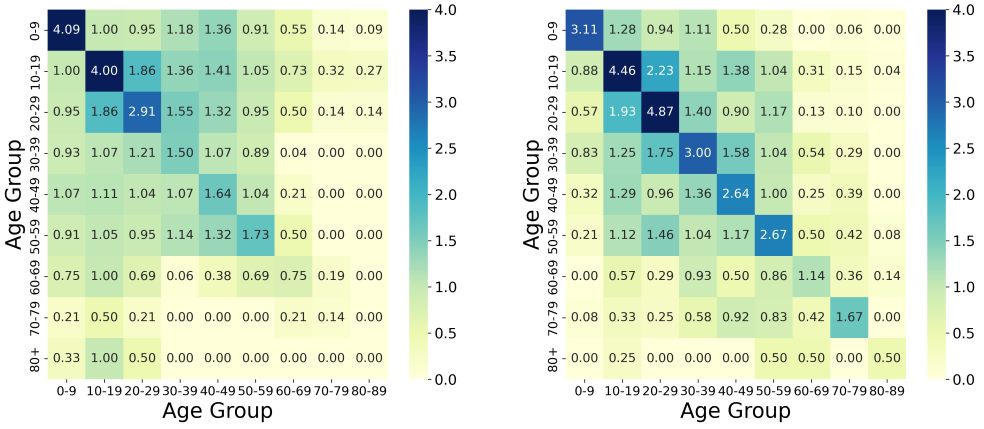
(a) Belgium

(b) Finland



(c) Germany

(d) Italy

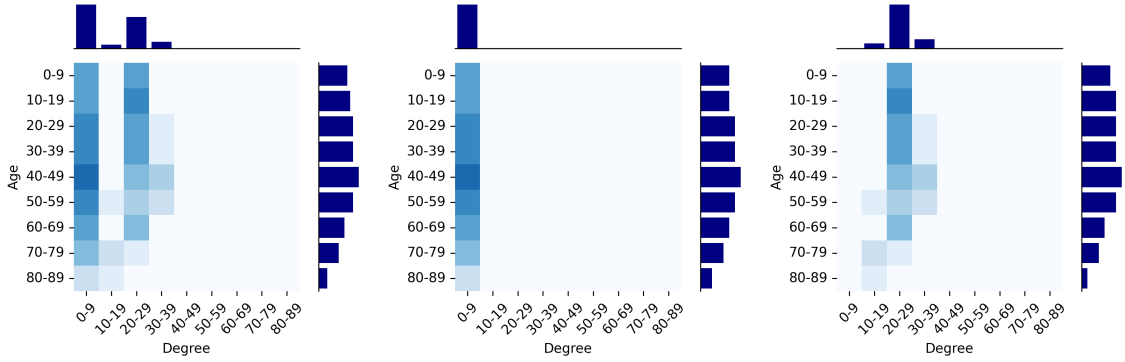


(e) Luxembourg

(f) Poland

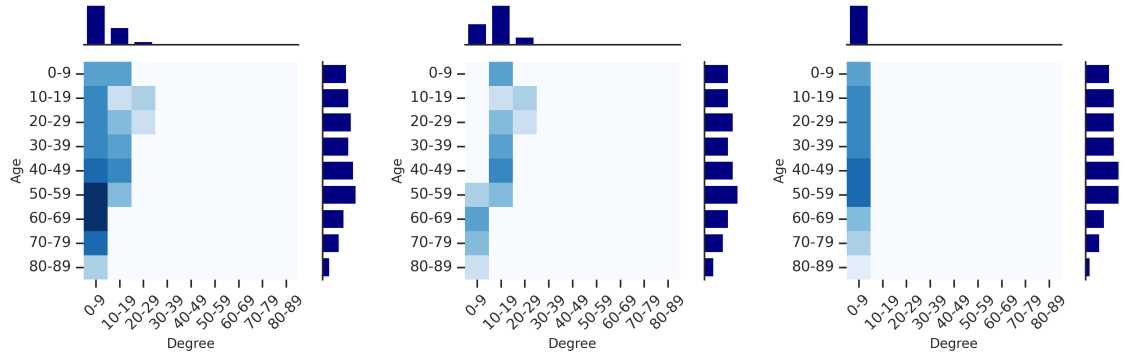
Figure 5.26: The social contact matrices recreated with the $HN - A_f - S$ model built on calibrated fuzzy sets for (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland.

features or preferred similarity, have larger node degrees. For example, in Poland, people around the age of $[10-30]$ have higher node degrees due to people's strong preference for age differences around 0 and young ages around 17 (Fig. 5.24 for $\Delta Age_f 1$ and $Age_f 2$).



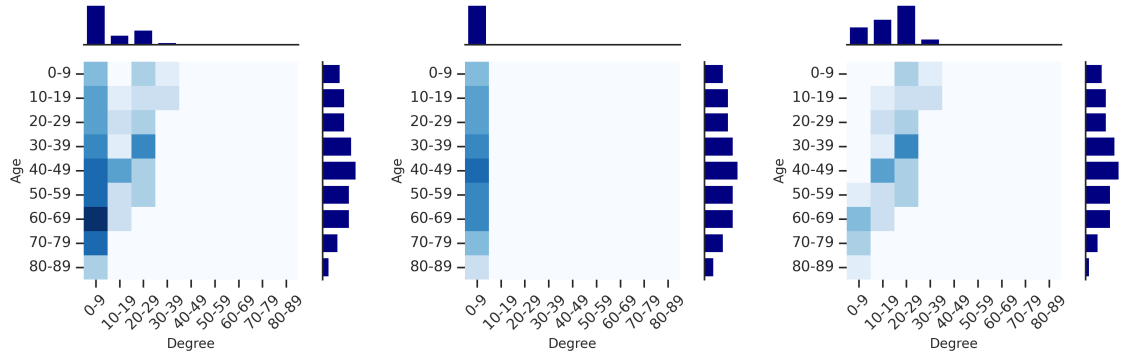
(a) Males&Females (b) Females (c) Males

Figure 5.27: The age and degree distributions in Belgium for (a) females and males, (b) females and (c) males.



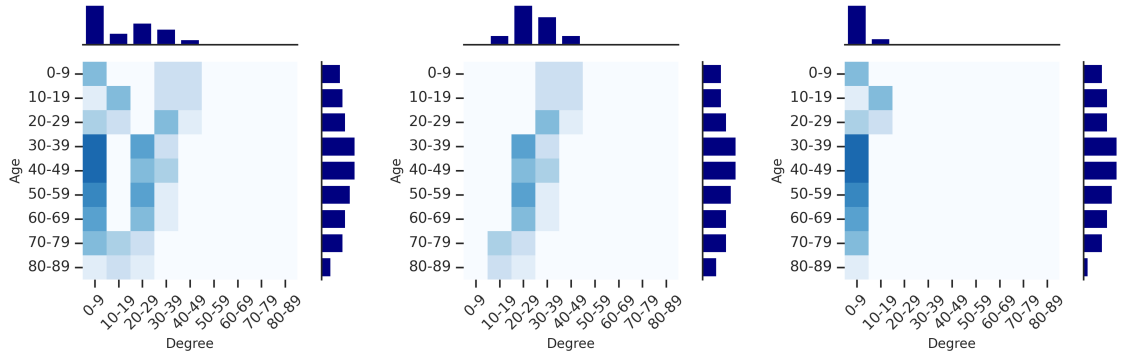
(a) Males&Females (b) Females (c) Males

Figure 5.28: The age and degree distributions in Finland for (a) females and males, (b) females and (c) males.



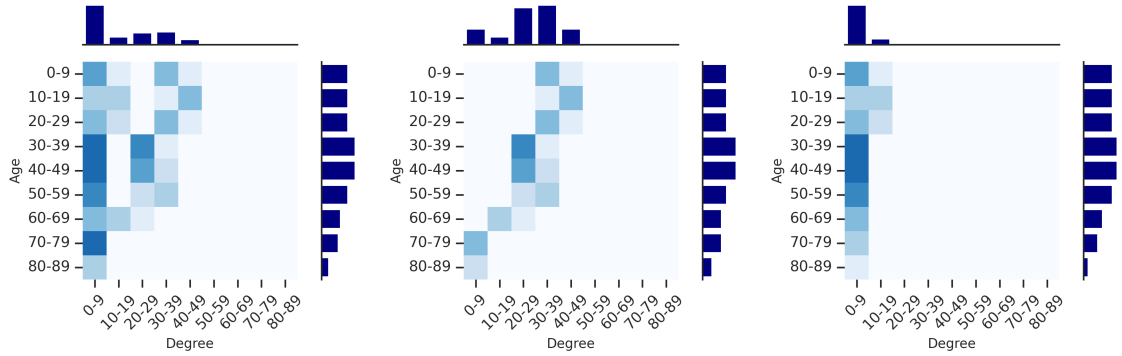
(a) Males&Females (b) Females (c) Males

Figure 5.29: The age and degree distributions in Germany for (a) females and males, (b) females and (c) males.



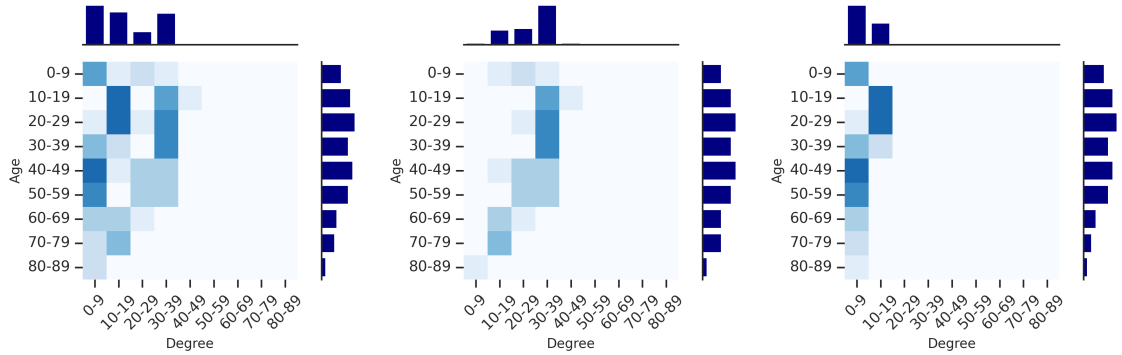
(a) Males&Females (b) Females (c) Males

Figure 5.30: The age and degree distributions in Italy for (a) females and males, (b) females and (c) males.



(a) Males&Females (b) Females (c) Males

Figure 5.31: The age and degree distributions in Luxembourg for (a) females and males, (b) females and (c) males.

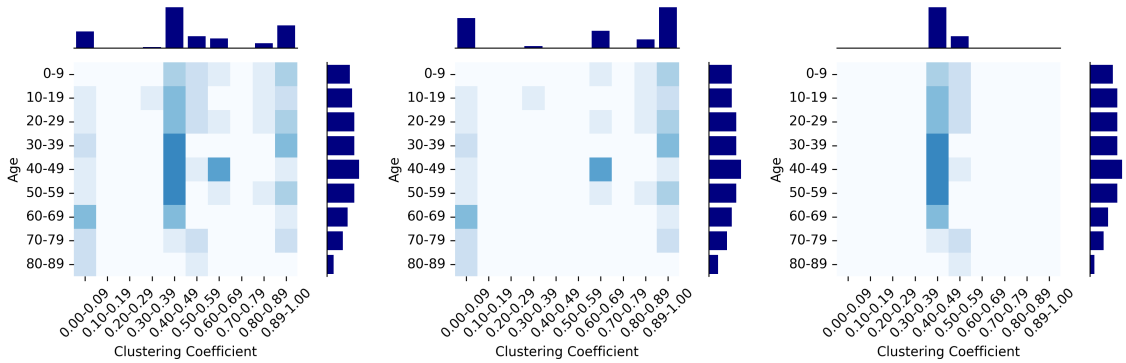


(a) Males&Females (b) Females (c) Males

Figure 5.32: The age and degree distributions in Poland for (a) females and males, (b) females and (c) males.

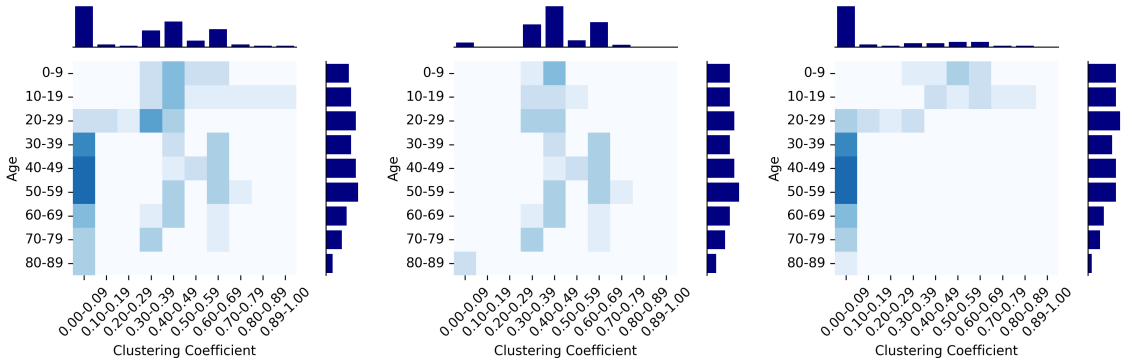
5.3.5.4 Clustering Coefficient Distribution

We explore the distributions of clustering coefficient for females, males and both females and males of different ages based on the network simulations of each country, including Belgium (Fig. 5.33), Finland (Fig. 5.34, Germany (Fig. 5.35), Italy (Fig. 5.36), Luxembourg (Fig. 5.37) and Poland (Fig. 5.38). The clustering coefficient is unevenly distributed, resulting from the heterogeneous preferences of nodes and the corresponding connections with various nodes. This challenges the interpretation of results. However, we can still find from the case of Belgium and Finland that the strongly preferred nodes have higher clustering coefficients than others. In Belgium, all males have a positive clustering coefficient due to the stronger preference for the crisp male features (Fig. 5.33); this contrasts with the clustering coefficient of females, where some of them have significantly lower clustering coefficients between 0 and 0.10. Similarly, in Finland, people around $[0-20]$ have a higher clustering coefficient than other nodes. This is because people in Finland have a strong interest and preference for age differences around 0 and 32. In this way, young people around the age of $[0-20]$ are preferred by a significant number of peers and people at their parents' age $[32-52]$ (Fig. 5.34).



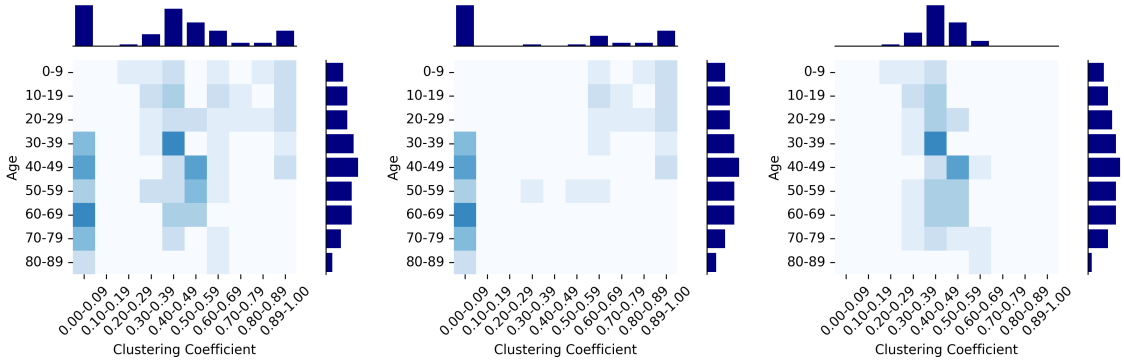
(a) Males&Females (b) Females (c) Males

Figure 5.33: The age and clustering coefficient distributions in Belgium for (a) females and males, (b) females and (c) males.



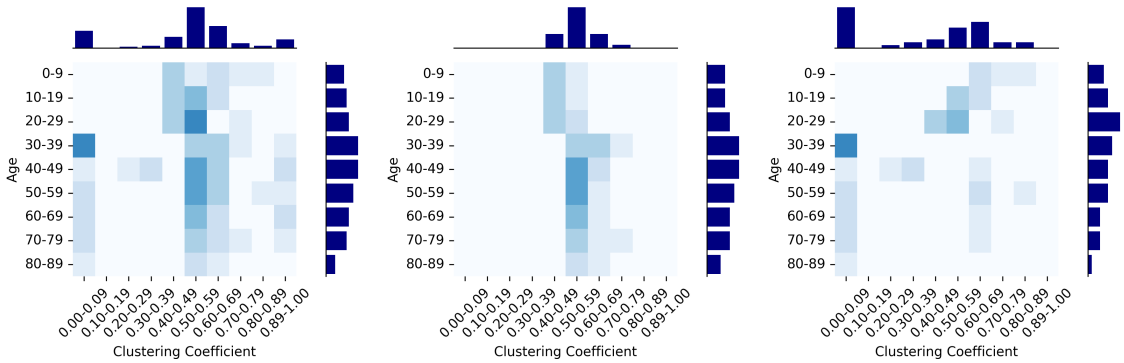
(a) Males&Females (b) Females (c) Males

Figure 5.34: The age and clustering coefficient distributions in Finland for (a) females and males, (b) females and (c) males.



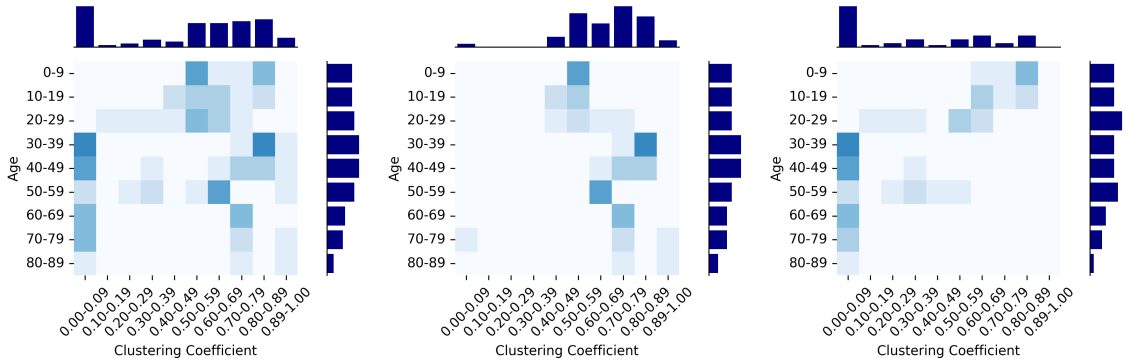
(a) Males&Females (b) Females (c) Males

Figure 5.35: The age and clustering coefficient distributions in Germany for (a) females and males, (b) females and (c) males.



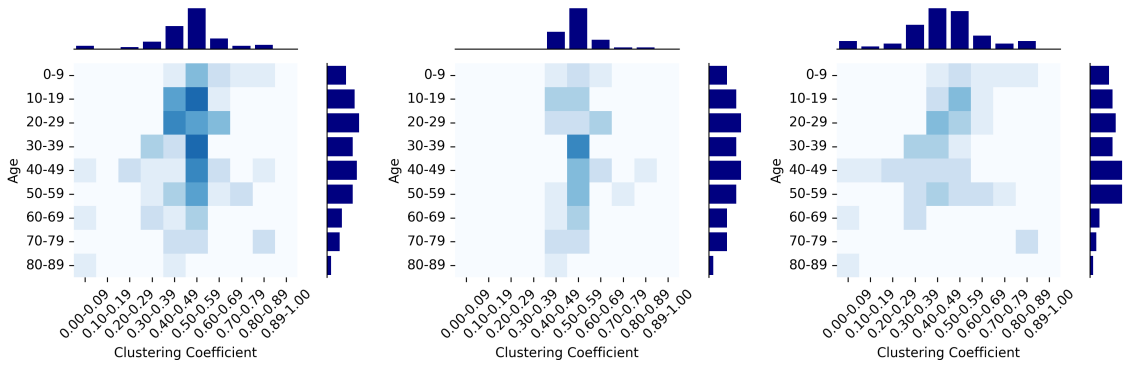
(a) Males&Females (b) Females (c) Males

Figure 5.36: The age and clustering coefficient distributions in Italy for (a) females and males, (b) females and (c) males.



(a) Males&Females (b) Females (c) Males

Figure 5.37: The age and clustering coefficient distributions in Luxembourg for (a) females and males, (b) females and (c) males.



(a) Males&Females (b) Females (c) Males

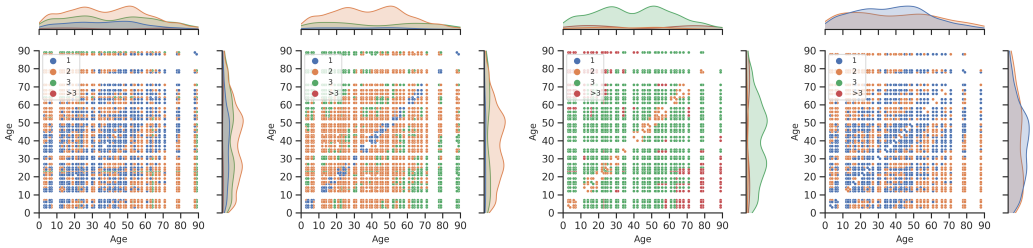
Figure 5.38: The age and clustering distributions in Belgium for (a) females and males, (b) females and (c) males.

5.3.5.5 Shortest Path Length Distribution

We explore the distributions of shortest path lengths between females and females, males and males, females and males, and all people based on the network simulations of each country, including Belgium (Fig. 5.39), Finland (Fig. 5.40), Germany (Fig. 5.41), Italy (Fig. 5.42), Luxembourg (Fig. 5.43) and Poland (Fig. 5.44). We identify different lengths of the shortest path with different colours (blue for 1, orange for 2, green for 3 and red for over 3; the fake (non-existing) paths are left vacant for clarity). We find that nodes with those unpopular features do not have a path between each other. For example, old females in Germany do not have direct or indirect connections to each other because they have a positive preference for males and dissimilar sex features (Fig. 5.41 and Fig. 5.21 (c) and (d)). This contrasts with the cases of Luxembourg and

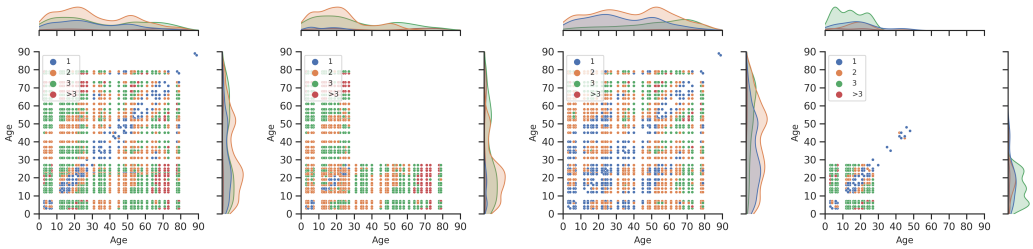
Finland, where old males over the age of 60 nearly do not have any direct or indirect interactions with others due to people's strong positive preference for young females and negative preferences for older males (Fig. 5.20 (c) and (d); Fig. 5.23 (c) and (d)). We also find that most of the direct interactions take place between people who have preferred sex features and preferred age differences. As an example, in Belgium and Germany, a significant number of direct interactions take place between males of similar ages or having an age difference of around 30 (Fig. 5.19 (c) and Fig. 5.21 (c) for preference information). This is different from the cases of Finland, Italy, Luxembourg and Poland, where most direct interactions take place between females around similar ages or having an age difference around 30 (Fig. 5.20 (c) and (d), Fig. 5.22 (c) and (d), Fig. 5.23 (c) and (d), and Fig. 5.24 (c) and (d)). In addition, we also find a limited number of direct interactions between the nodes with young and similar ages. For example, in Finland, Belgium, Germany, Italy, Luxembourg and Poland, we can find a few direct interactions between males and females at a similar young age despite their negative preference for different sex features. Moreover, in Belgium, there are more interactions between males and females over the age range of $[0-90]$ due to their preferences for different sex features and similar ages.

There is a significant number of indirect interactions, as indicated by a shortest path length of 2, distributed around the above-mentioned direct interactions. This takes place between the nodes and their neighbour's neighbours. For example, in Poland, females at the age of 50 prefer to be connected with females at the age of 20, and in the meantime, females around the age of 20 may be connected with people at the age around $[10-20]$ due to similar age differences. These two direct interactions lead to an indirect shortest path between the age around 50 and the age around 10 (Fig. 5.24 (b)). In all the countries, we can find a significant number of indirect interactions (with the shortest path length of 2) between female and male people and a longer length of shortest path between people with unpreferred sex features (e.g. the male-female contact in Fig. 5.23 (d)).



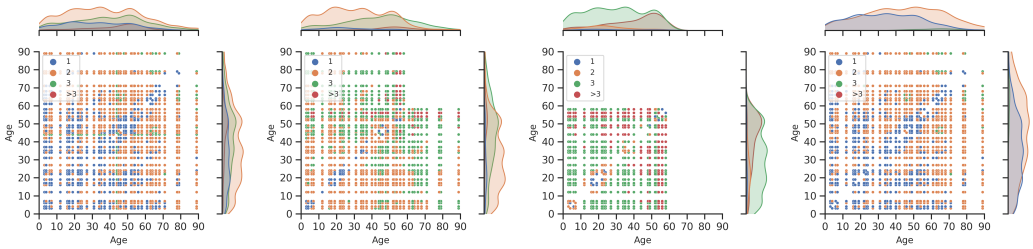
(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.39: The age and shortest path length distributions in Belgium for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.



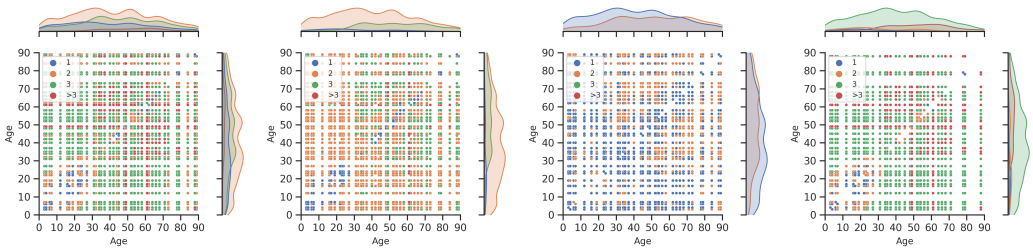
(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.40: The age and shortest path length distributions in Finland for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.



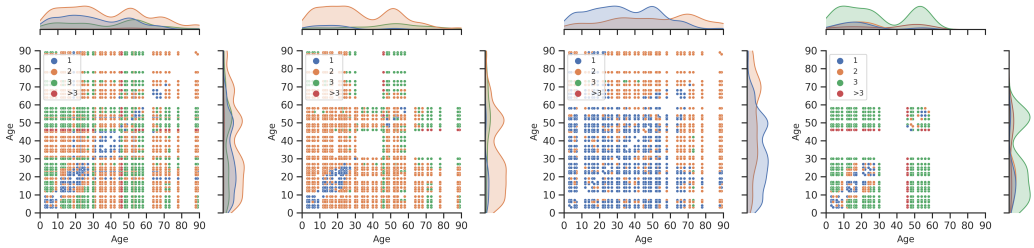
(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.41: The age and shortest path length distributions in Germany for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.



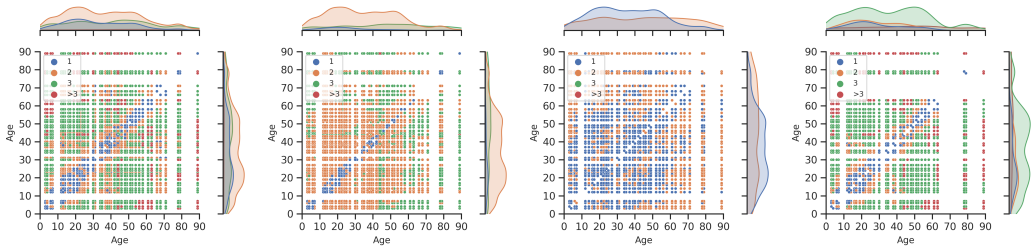
(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.42: The age and shortest path length distributions in Italy for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.



(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.43: The age and shortest path length distributions in Luxembourg for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.



(a) People-People (b) Male-Female (c) Female-Female (d) Male-Male

Figure 5.44: The age and shortest path length distributions in Poland for contact between (a) all people, (b) males and females, (c) females and females, and (d) males and males.

5.3.6 Disaster Resilience

We investigate the disaster resilience in by introducing a simulation-based epidemic spreading process in the recreated social networks. We assume that (i) the disease transmissibility given exposure is 0.80, without change over time, (ii) the epidemic spreads with a single seed selection in the initial stage based on the highest node degree, (iii) the epidemic spread on the network propagates one step (edge) away for each time step and (iv) the susceptible nodes get infected without recovery (SI process). The transmissibility at 0.8 is a high transmissibility which speeds up the epidemic spread and leads to a significant number of infections within 6 time steps. Under this transmissibility, the different network patterns of each country lead to heterogeneous infection patterns.

5.3.6.1 Seed Selection

The epidemic spreading process starts from the most popular node in each country (with the highest node degree), leading to the biggest number of infections within a limited time and space. The seed selection strategy involved in this study is mainly for illustration purposes. We can also introduce different seed selection strategies using the proposed model in future research. Tab. 5.7 shows the features of the most popular node in each country's social network. The most popular node is a female age around 20 in Finland, Italy, Luxembourg and Poland. In contrast, the most popular node is a male age 19 in Germany, and a 48-year-old male is the most popular node in Belgium's social networks. These differences result from different social networks in each country and lead to various spreads. More specifically, the differences in the sex features of the most popular node result from people's preferences. People in Belgium and Germany have strong positive preference for males (Fig. 5.19 (c) and Fig. 5.21 (c)), which contrasts the cases of Finland, Italy, Luxembourg and Poland (Fig. 5.20 (c), Fig. 5.22 (c), Fig. 5.23 (c), Fig. 5.23 (c) and Fig. 5.24 (c)). In addition, people in all the countries, except for Belgium, have a strong positive preference for young ages (or a significant negative preference for old ages), which makes young nodes popular. In Belgium, people have a greater interest and positive preference for similar ages (Fig. 5.19 (c)), and this results in a significant number of connections between nodes in a dense age group such as $[40-49]$.

Table 5.7: The starting node of the epidemic spreading process in each country.

Country	Belgium	Finland	Germany	Italy	Luxembourg	Poland
Age of the starting node	48	19	19	19	22	12
Sex of the starting node	Male	Female	Male	Female	Female	Female

5.3.6.2 Infection Occurrence

We simulate the epidemic spreading process for each country over time until all the connected nodes get infected.Fig. 5.45 shows the number of infections within specific distance to the seed node (Fig. 5.45(a)) and specific time steps of the epidemic transmission (Fig. 5.45(b)).

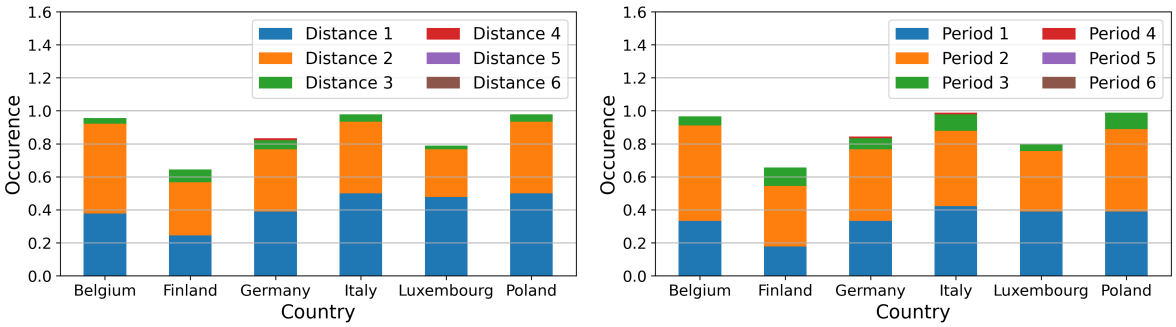


Figure 5.45: The proportion of infections to the overall population (90 nodes) within the specific distance (a) to the seeded node within specific simulation steps (b).

As shown in Fig. 5.45, generally, all the countries, except for Germany, obtain the most infections within a distance of around 3 and after 3 time steps. In contrast, Germany achieves the largest number of infections within the distance of 4 over 4 time steps, where one more infection takes place in the fourth time step and four edges away from the seed node. This can be caused by the existence of long shortest path lengths between the nodes and the uncertainty of epidemic transmission given multiple exposures. Compared with the other countries, Finland has the fewest infections in the first step (less than 30%) and overall infections (less than 70%, in contrasts with the other countries' infection occurrence over 80%.). This results from the isolated characteristics of people with unprefered features and the limited number of social contacts (the node degree distribution in Fig. 5.28 and the shortest path length dsitribution in Fig. 5.40). Moreover, Belgium and Germany have fewer infections (less than 40%) within the distance of 1 in the first step but more widespread infections afterwards (over 80%) than Italy, Luxembourg and Poland. This is because, in the social network simulations, people in Belgium and Germany have a smaller number of direct interactions and a larger number of indirect interactions than people in Italy, Luxembourg and Poland. This implies that isolation policies imposed on the seed node and its direct infections at the epidemic's beginning in countries like Belgium and Germany can effectively reduce the infection numbers.

5.3.6.3 People at Risk

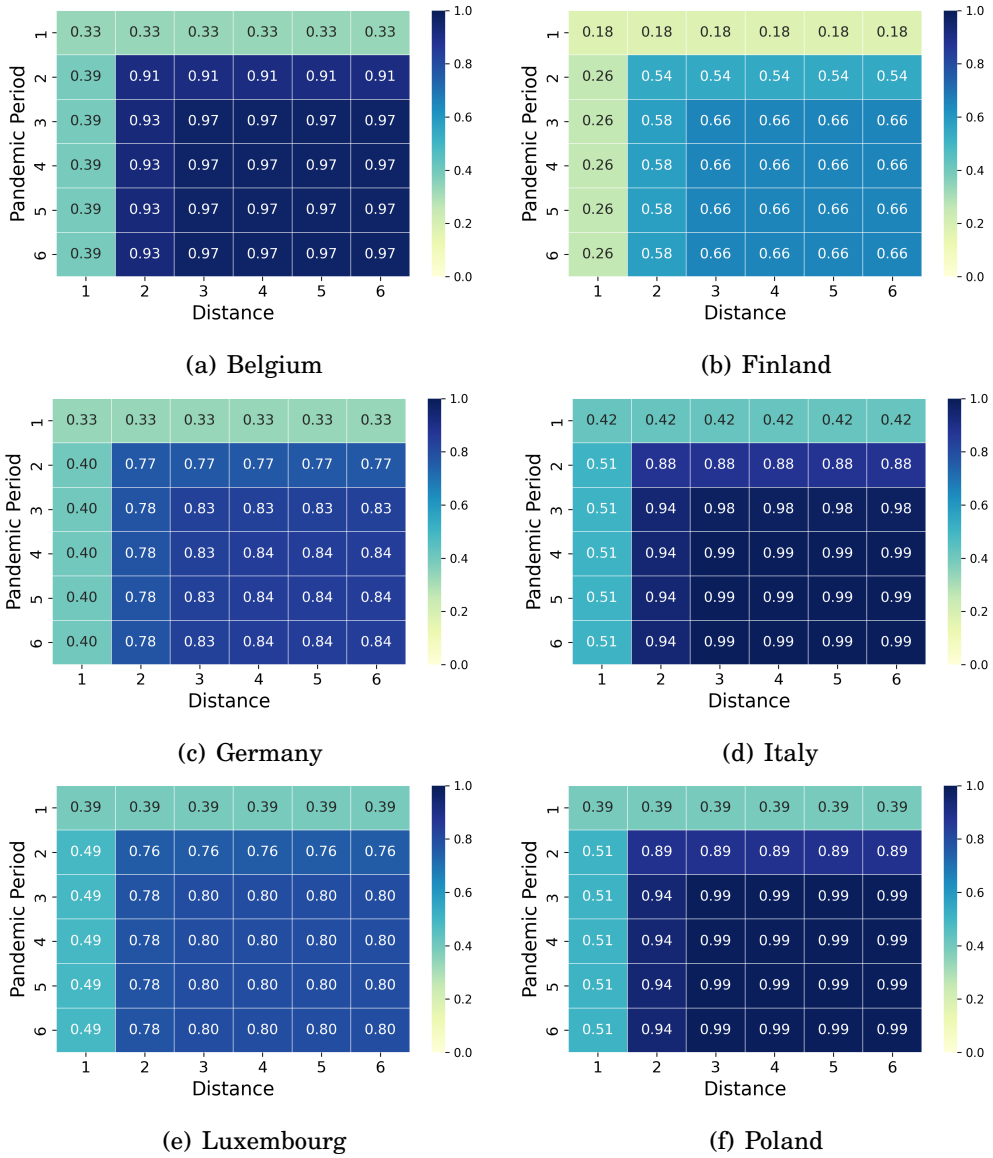
We describe the People at Risk ($PaR(T,D)$) of getting infected within the simulation steps of $[0,T]$ and the distance of D from the seed node based on the equation 5.15, as

proposed in our previous study [3].

$$PaR(T,D) = \frac{\sum_{v_{i,t} \in V_t} \mathbf{r}(v_{i,t}) \delta_{\mathrm{l}(v_{i,t},\mathbf{s}_t),D} \delta_{t,T}}{N}, \quad D \leq T \tag{5.15}$$

where $PaR(T,D)$ represents the proportion of infections within a distance D and T time steps since the epidemic outbreak from the seed node. $\delta_{t,T}$ and $\delta_{\mathrm{l}(v_{i,t},\mathbf{s}_t),D}$ each represents the Kronecker functions related to time t and the distance $\mathrm{l}(v_{i,t},\mathbf{s}_t)$ between node $v_{i,t}$ and seed $\mathbf{s}_t$. N represents the number of nodes.

As shown in Fig. 5.46, for each pandemic period, the PaR values increase with the distance and simulation time steps. Within a specific distance from the seed node, the PaR values increase until all the connected nodes are infected. Overall, the social networks in Belgium, Italy and Poland will, at the end of epidemic simulation, get higher PaR values (close to 100%) than in other countries if the epidemic keeps spreading. This can be caused by the higher number of direct and indirect interactions with the popular seed node and its popular neighbours. As shown in Fig. 5.33 (c), Fig. 5.36 (c) and Fig. 5.38 (c), males in Belgium, females in Italy and females Poland all have a positive clustering coefficient, which implies that the seed node and its neighbours with the same preferred sex features cluster densely and leads to extensive epidemic transmission. Moreover, there are less number of fake paths in cases of Belgium, Italy and Poland than that in other countries (Tab. 5.6, Fig. 5.39, Fig. 5.42 and Fig. 5.44), which leads to a densely connected network and higher number of overall infections. We also find that the $PaR(2,2)$ values of the social networks in Belgium, Italy, Germany, Luxembourg and Poland are significantly higher than their $PaR(1,1)$ values, which are over 75%. This implies the necessity of epidemic control within a distance of 2 – no later than the second time step. In the case of Finland, the $PaR(3,3)$ values reach the highest infection occurrence and keep lower than 70%, which is also significantly higher than $PaR(1,1)$ values but lower than that in other countries, which, considering the increase of infection occurrence, also requires epidemic control no later than the second time step.



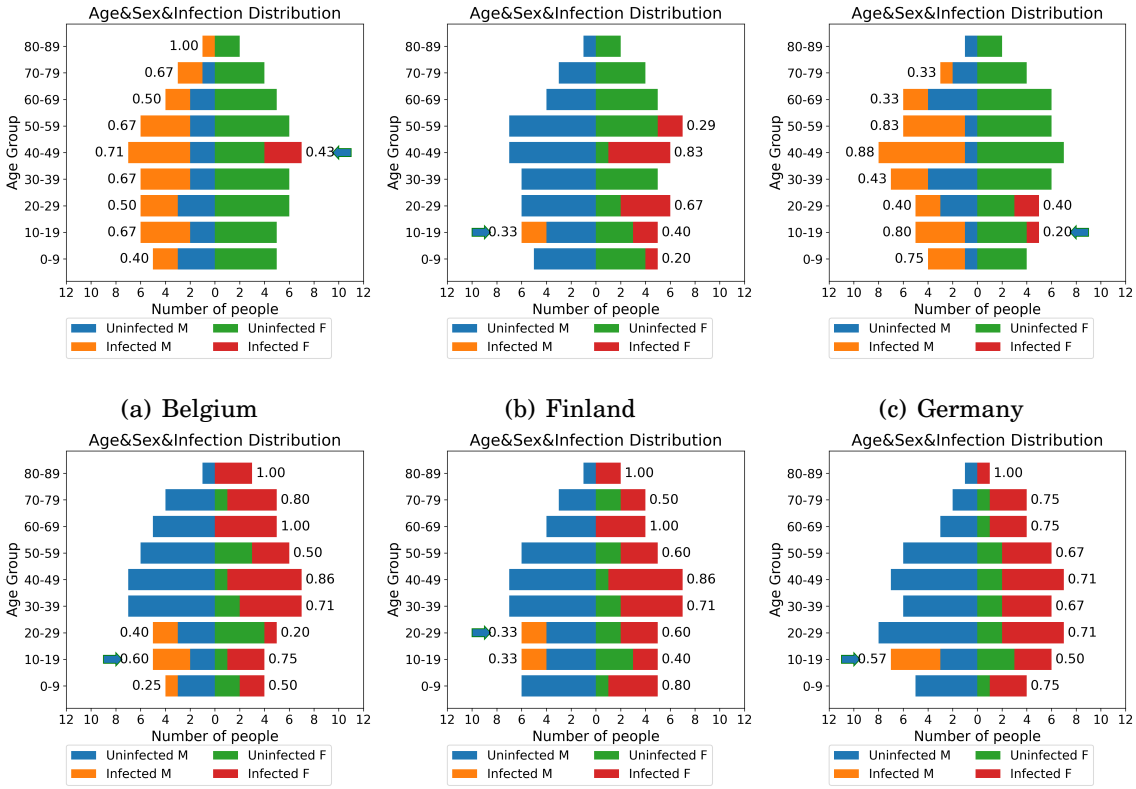
(a) Belgium (b) Finland

(c) Germany (d) Italy

(e) Luxembourg (f) Poland

Figure 5.46: The PaR within the specific distance to the seeded node over the specific time steps in (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland.

We calculate the $PaR(1,1)$ of each age & sex group to better understand the influence of age and sex features on the infection status in the first time step and one hop away from the seed node (the $PaR(1,1)$ values denoted for each age&sex bar in Fig. 5.47). Given a starting node from male/male group, people with the same sex features and people in a similar age have positive PaR values. For example, given a female starting node whose age ranges between 12 and 22 (Fig. 5.47 and Tab. 5.7), females in Finland, Italy, Luxembourg and Poland tend to have higher $PaR(1,1)$ values than males, and vice



(a) Belgium (b) Finland (c) Germany

(d) Italy (e) Luxembourg (f) Poland

Figure 5.47: The simulated age & sex & infection-distributions in 2005 of countries including (a) Belgium, (b) Finland, (c) Germany, (d) Italy, (e) Luxembourg and (f) Poland. The arrow points to the specific age & sex group that includes the seed node for epidemic spread (Tab. 5.7). We also denote each age&sex group with an infection ratio in the first step of epidemic spread to indicate the infections risk $PaR(1,1)$.

versa for Belgium and Germany. The younger age groups and bigger age groups (age groups allocated with more nodes) generally have higher $PaR(1,1)$ values due to their popularity resulting from their age & sex features and the social contacts developed based on homophily effect (preference for similar age and sex features). Specifically, in Finland, Italy, Luxembourg and Poland, males, in a similar age to the seed node, have positive $PaR(1,1)$ values due to people's preference for social contact with a similar age and young age. These positive $PaR(1,1)$ values indicate the existence of infection risks to the respective age & sex group. And for the same reason, all males and females in the young age groups in Belgium and Germany also have positive $PaR(1,1)$ values given the male starting node. In addition, an uneven distribution of $PaR(1,1)$ values covers the male age groups when the epidemic spreads from a male, and we can also find similar uneven distributions for the $PaR(1,1)$ of females if the epidemic spreads from a

female. This results from the different connections developed by the seed node and the randomness of epidemic spread given a transmissibility at 0.80.

Overall, the heterogeneous features lead to complex social contact patterns and epidemic spreads, which also need to be addressed with heterogeneous and flexible policies. For example, within the first time step and the distance one step away from the epidemic outbreak, we can create different epidemic control policies based on the various $PaR(1,1)$ values for people in each age & sex group. A higher $PaR(1,1)$ value indicates a higher infection risk under the influence of social network structure, which requires a more extensive quarantine policy in the respective age & sex group. More specifically, people with preferred features tend to cluster and thus have higher infection risks within the shortest time and distance to the seed node. For example, males in a similar age or with an age difference around 30 are popular in Belgium, which given significant number of males age around $[20-60]$, leads to more interactions and the infection risks from multiple exposure to the infection spread. People in Finland have strong preference for young females, which lead to isolated old males and correspondingly, less infection risks to those unconnected people. People with more interactions but higher interaction risks have decisive impact on the infection occurrence. Therefore, epidemic control policies such as suggestions, compulsory vaccination or isolation can be targeted at those popular and risky nodes with preferred features in the first instance.

5.4 Conclusions

In this chapter, we propose DT-CNS models based on Fuzzy Sets. More specifically, we improve the expressive power of the DT-CNS models proposed in the chapter 4 by introducing heterogeneous feature representation principles, including the crisp and fuzzy value representations. We investigate the impact of various features and feature representation principles and select and validate the best-performing model for a faithful network representation.

In our empirical analysis, we build DT-CNS models to recreate social contact networks for countries including Belgium, Finland, Germany, Italy, Luxembourg and Poland based on their fuzzy/crisp representation of the real age & sex distributions and their corresponding preferences for a given age & sex. We evaluate the outcomes and select the most promising model from the created ones considering the similarity between the target and the recreated social contact matrices. The model selection results suggest that (i) models built on age and sex features outperform the models without these features or with a

single age feature, (ii) models built on fuzzy representation principles generally perform better than others and (iii) network models built on sex and fuzzy representation of age ($HN - A_f - S$) are good candidates for the network representation. This suggests that increasing the modelling complexity and using more heterogeneous input information improve the modelling accuracy. In addition, we validate the initialised fuzzy representation principles by introducing a flexible number of fuzzy sets and partial parameter optimisation based on a global sensitivity analysis. The model validation results indicate that (i) increasing flexibility of fuzzy representation principles and (ii) optimisation of partial parameters, even with a small change, improve model performance.

Given the selected network models and fuzzy representation principles, we create realistic social contact networks and investigate their respective disaster resilience when the epidemic spreads from the most popular node. People at risk (PaR), as measured by the proportion of infection to the population within a specific period and distance, changes with time and distance and varies depending on the country. This results from the heterogeneous features and the heterogeneous principles of representation and preference, which require case-dependent and changeable mitigation policies. In addition, a higher PaR value indicates a higher infection risk under the influence of social network structure, and requires a more extensive quarantine policy.

To conclude, the structural and dynamic complexity increases when incorporating more features and introducing more flexible representation principles. Such an increase in complexity can potentially improve the faithfulness of network representation by preserving necessary heterogeneous information observed in reality. The feature heterogeneity influences the network structure and induces the epidemic outbreak among specific people. And this requires heterogeneous mitigation policies targeted at heterogeneous populations.

Overall, this chapter answers **RQ4** and contributes to the research topic on DT-CNS models from three perspectives:

- **RQ4**: How to represent uncertain features with DT-CNS models?

    1. **RO4.1**: Develop flexible rules of feature representation for the DT-CNS models.

       RC4.1: Feature representation with fuzzy sets for uncertain features.

    2. **RO4.2**: Validate the influence of different ways of feature representation on DT-CNS models.

       RC4.2: DT-CNS models built on different feature representation principles.

3. **RO4.3**: Evaluate the proposed DT-CNS models built on different feature representation principles.

   RC4.3: A case study on DT-CNS models built with increasingly flexible rules of feature representation.

The DT-CNS modelling framework currently focuses on static networks and dyanmic processes with a fixed transmissibility rate, which falls in the generation 1 category and is extendable to higher complexity levels. Our next research efforts focus on DT-CNS model extension that includes three aspects: (i) preference mutation mechanism that drives network evolution, (ii) transmission mechanisms dependent on node features, and (iii) interrelations between the dynamics of the network and the process. In the next chapter, we aim to develop DT-CNS models with more complex dynamics and answer the **RQ5**, concerning the building and evaluation of evolutionary DT-CNS models driven by mutation of feature preferences.

CHAPTER 6

NODE FEATURES AND THE MUTATION OF FEATURE PREFERENCES FOR DT-CNS MODEL.

6.1 Introduction

Only a few studies investigate the influence of an epidemic outbreak on network evolution [69, 70]. For example, [69] models the endogenous social distancing phenomenon in evolving social networks by incorporating the interaction risks from an epidemic outbreak into the nodes' decisions on connecting with others. Despite this study, an undeveloped research space remains where nodes make heterogeneous decisions on network evolution in response to the connectivity patterns and the infection occurrences for positive feedback. The interrelated changes of networks and processes, motivated by the nodes' purposeful behavioural changes, add complexity to the CNS dynamics and pose a challenge to the DT-CNS modelling framework.

Some studies have accommodated the individual motivations of social interactions into the individual decision-making processes on connection preferences. They employ the social capital concept to represent people's investment in social connections for expected returns [455, 456]. For example, in the knowledge-sharing process on social networks, nodes, motivated by reputation and altruism, conduct knowledge-sharing behaviours based on their social capital (such as trust, identification and reciprocity, which are embedded in social relations) [455].

Our previous work in chapter 4 and chapter 5 proposed the DT-CNS modelling

framework based on heterogeneous node features and related preferences, including preferential attachment and homophily [3]. The framework proposed there models CNS composed of static networks and a dynamic process on the networks. This chapter extends this framework by modelling the interrelated network evolution and dynamic process on the evolving networks, which naturally progress towards evolutionary DT-CNS models in generation 2 and 3 by introducing changes of connection preferences and points to the research question **RQ5** to be addressed in this chapter.

- **RQ5**: How to build and evaluate the DT-CNS model evolution driven by mutation of feature preferences?
    1. **RO5.1**: Build an evolutionary DT-CNS model with changeable feature preferences and interrelated DT-CNS components
    2. **RO5.2**: Build an evolving DT-CNS model based on the changes of feature preferences
    3. **RO5.3**: Evaluate the proposed evolutionary DT-CNS models.

In this chapter, we build an evolutionary DT-CNS modelling framework featuring heterogeneous node features and changeable connection preferences, which combine the effects of preferential attachment, homophily and transitivity. We propose to measure the nodes' interaction utilities based on the reward (social stimuli originates from and motivates the interactions) and epidemic risks from their interactions. We create heterogeneous preference mutation mechanisms to characterise nodes' adaptive decisions on changing the connection preferences in response to the changes of interaction utilities. In addition, we introduce the nodes' maximum investment in social connections, referred as social capital constraints, to the nodes' preference mutation process and propose to model nodes' preference changes as an optimisation process for higher interaction utilities with limited social connections. Under the social capital limit, nodes can follow heterogeneous preference mutation styles, including the (i) inactive style that keeps zero social preferences, (ii) ignorant style that randomly mutates preferences, (iii) egocentric style that optimises individual interaction utility, (iv) cooperative style that optimises the total interaction utilities by group decisions and (v) collaborative style that further allows the cooperative nodes to transfer social capital.

Through conducting simulation-based experiments on the evolutionary DT-CNS models, we find that (i) increasing social capital enables higher interactions and triadic closures but higher infection risks and uncertainty in decision-making; (ii) group decisions outperform individual decisions by avoiding the unawareness of the decisions of

other nodes; (iii) the collaborative strategy, given a strict social capital limit, outperforms the inactive, ignorant, egocentric and cooperative strategies by achieving more interactions and triadic closures, lower infection numbers and consequently, higher interaction utilities.

This chapter innovates in the following aspects

- proposal of an evolutionary DT-CNS modelling framework related to heterogeneous node features and the changeable preferences for connecting with others (**RC5.1**);
- introduction of social capital limits to nodes' adaptive decisions on preferences changes for interaction utility maximisation (**RC5.1**);
- creation of heterogeneous preference mutation mechanisms that drive nodes' individual and group behaviours under the impact of dynamic process on networks (**RC5.2**);
- suggestion of strict social capital limit on social connections and collaboration between nodes in an epidemic outbreak (**RC5.3**);

The rest of this chapter is structured as follows: section 6.2 presents the methodology of building an evolutionary DT-CNS model. Following this, section 6.3 builds and evaluates the evolutionary DT-CNS models. Finally, we conclude with section 6.4.

6.2 Modelling Framework

In this section, we build an evolutionary DT-CNS model based on (i) heterogeneous features of nodes, (ii) nodes' preferences for connecting with others and (iii) heterogeneous preference mutation styles that are driven by the utility and the risk of interactions. We present the components of DT-CNS models related to nodes, their interactions driven by features and related preferences, and their conditions in an epidemic spread in section 6.2.1 and propose ways to calculate the interaction utilities and risks for individuals. In section 6.2.2, we model the dynamics of DT-CNS model that governs the nodes' interactions and interaction preference mutation based on their utilities and risks.

6.2.1 Complex Networked System Components

CNS models are composed of the (i) nodes that possess features and preferences for connecting with others and (ii) the edges that represent interactions between pairs of

nodes. These nodes and edges compose the CNS models' network dimension, and in the CNS models' process dimension, the process (e.g. epidemic) spreads via the edges and results in changes in the nodes' status (e.g. nodes' health conditions). The network and process dimensions constitute the CNS model.

We represent the network at time t as $\mathbf{G}_t = \{\mathbf{V}_t, \mathbf{E}_t\}$, based on the dynamic network components: nodes $\mathbf{V}_t = \{v_{1,t}, \cdots, v_{N,t}\}$ and edges $\mathbf{E}_t = \{e_{ij,t} | v_{i,t}, v_{i,t} \in \mathbf{V}_t\}$. The node $v_{i,t}$ at time t is defined as:

$$v_{i,t}(\mathbf{f}_{i,t}, sDNA_{i,t}, \beta_{i,t}, \mathrm{U}_{i,t}, \mathrm{R}_{i,t}, \$_{i,t}, \hat{\$}_{i,t}) \tag{6.1}$$

where we represent the node's features and the related preferences with $\mathbf{f}_{i,t}$ and $sDNA_{i,t}$. $\beta_{i,t}$ represents the node's health condition with 0 and 1, each representing the node's healthy and infected condition. The node $v_{i,t}$ obtains the utility $\mathrm{U}_{i,t}$ from interacting with others at time t and suffers the interaction risk $\mathrm{R}_{i,t}$ given an epidemic outbreak. The social capital embedded in nodes' interactions is denoted by $\$_{i,t}$ and constrained by the social capital limit $\hat{\$}_{i,t}$. Nodes interact under the social capital limit $\hat{\$}_{i,t}$ for objectives related to interaction utilities.

The node's preferences $sDNA_{i,t}$ for the related features $\mathbf{f}_{i,t}$ is determined by

$$sDNA_{i,t} = (\mathbf{p}_{i,t}, \mathbf{w}^p_{i,t}, \mathbf{h}_{i,t}, \mathbf{w}^h_{i,t}, \mathbf{c}_{i,t}, \mathrm{w}^c_{i,t}) \tag{6.2}$$

where $\mathbf{p}_{i,t}$ and $\mathbf{h}$ are vectors that determine the negative, zero, and positive preferences for features and feature differences by values -1, 0 and 1. These two vectors are followed with the same length weighting vectors $\mathbf{w}^p$ and $\mathbf{w}^h$, indicating the weight of preference and valuing within $(0, 1]$. Similarly, $\mathbf{c}$ determines the negative, zero and positive preference for connecting with nodes with common friends by -1, 0 and 1, followed by the weight of preference $\mathrm{w}^c_{i,t}$.

The edge $e_{ij,t}$ represents the interactions between the node pair: $v_{i,t}$ and $v_{j,t}$, which is defined as:

$$e_{ij,t}(I_{ij,t}, w^I_{ij,t}) \tag{6.3}$$

where $I_{ij,t}$ represents no interaction and an interaction between a node pair with 0 and 1 respectively. It is followed by an interaction intensity factor $w^I_{ij,t}$, ranging beteween 0 and 1.

6.2.1.1 Interaction reward

The interaction reward generally refers to social stimuli that positively reinforce the frequency or intensity of a behaviour pattern [457, 458]. In the context of social networks,

the interaction reward $\mathbf{R}_{i,t}$ for node $v_{i,t}$ represents the social stimuli such as cultural goods, money, positive emotional expressions and social reputation, which originates from and further motivates the interactions between node $v_{i,t}$ and others.

In this chapter, we assume that a node's interaction reward depends on the scores and frequency of its interactions. It is calculated by aggregating the reward $r_{i,t}(i,j)$ from any pairwise interactions.

$$\mathbf{R}_{i,t} = \sum_{v_{j,t} \in V_t} r_{i,t}(i,j) \tag{6.4}$$

The reward for node $v_{i,t}$ from the interaction with node $v_{j,t}$ depends on the interaction score $\Pi_{ij,t}$ and the individual interaction cost $c_{i,t}$.

$$r_{i,t}(i,j) = \epsilon \Pi_{ij,t} - \eta_{i,t}^{\beta} c_{i,t} \tag{6.5}$$

where $\epsilon > 0$ indicates that the reward correlates directly with the interaction score $\Pi_{ij,t}$. $\eta \in (1,2]$ is a tuning parameter for the node's average interaction cost $c_{i,t}$, depending on the node's infection status $\beta_{i,t}$. The infected nodes have a higher interaction cost and low-score interactions.

The interaction score $\Pi_{ij,t}$ of the interaction between the node pair $v_{i,t}$ and $v_{j,t}$ depends on their mutual evaluation based on features and related preferences.

$$\Pi_{ij,t} = \frac{1}{2}(\Pi_{i \to j,t} + \Pi_{j \to i,t}) \tag{6.6}$$

where $\Pi_{i \to j}$ and $\Pi_{j \to i}$ each represents the unilateral evaluation of node $v_{i,t}$ and node $v + j, t$ on their interaction.

The score $\Pi_{i \to j}$ that node $v_{i,t}$ assigns to its interaction with $v_{j,t}$ is calculated based on the preferential attachment, homophily and closure effects, each represented as $\pi^h_{i \to j,t}$, $\pi^p_{i \to j,t}$ and $\pi^c_{i \to j,t}$.

$$\Pi_{i \to j,t} = \begin{cases} \frac{1}{3}(\pi^h_{i \to j,t} + \pi^p_{i \to j,t} + \pi^c_{i \to j,t}) + \epsilon_{i \to j,t} & \text{if } v_{i,t} \text{ and } v_{j,t} \text{ encounters.} \\ 0 & \text{if else.} \end{cases} \tag{6.7}$$

where $v_{i,t}$ evaluates the pairwise interaction with others by averaging the homophily score $\pi^h_{i \to j,t}$, preferential attachment score $\pi^p_{i \to j,t}$ and the closure effect $\pi^c_{i \to j,t}$ based on their encounters, ranging between $[0,1]$. Moreover, without an encounter, the subjective node assigns 0 to this unilateral evaluation of the pairwise interactions. $\epsilon_{i \to j,t} \sim N(0, 0.01^2)$ is a random interference that follows a normal distribution $N(0, 0.01^2)$. It characterises the unobservable and uninterpretable interference in nodes' evaluation on

interactions between pairs of nodes, which results in different node pair scores even for the node pairs attributed with the same features.

The homophily effect $\pi^h_{i\to j,t}$ describes the node's preference for connecting with similar or dissimilar others. It is calculated based on the feature differences and the preference vectors $\mathbf{p}_{i,t}$ and $\mathbf{w}^p_{i,t}$:

$$\pi^h_{i\to j,t} = \frac{1}{2l}|\mathbf{f}_i - \mathbf{f}_j|^\tau(\mathbf{p}_{i,t} \odot \mathbf{w}^p_{i,t}) + \frac{1}{2} \tag{6.8}$$

where the feature vector $\mathbf{f}_j$ includes the standardised feature values from a range $[0,1]$, and the aggregated homophily effects from each feature difference are averaged and then standardised based on the feature-length l, ranging between $[0,1]$.

The preferential attachment effect $\pi^p_{i\to j,t}$ describes the node's preference for connecting with other nodes with specific features:

$$\pi^p_{i\to j,t} = \frac{1}{2l}\mathbf{f}_j^\tau(\mathbf{p}_{i,t} \odot \mathbf{w}^p_{i,t}) + \frac{1}{2} \tag{6.9}$$

where, similar to the homophily effect, each feature's aggregated preferential attachment effects are also standardised with the feature-length l, ranging between $[0,1]$.

The closure effect $\pi^c_{i\to j,t}$ describes the node's preference for connecting with nodes with or without common friends.

$$\pi^c_{i\to j,t} = \mathrm{c}_{i,t}\mathrm{w}^c_{i,t}\frac{\sum\limits_{v_k\in\mathbf{V}} I_{ik,t}I_{jk,t}}{\sum\limits_{v_k\in\mathbf{V}} I_{ik,t}} \tag{6.10}$$

where we calculate the number of common friends between node $v_{i,t}$ and node $v_{j,t}$ with $\sum\limits_{v_k\in\mathbf{V}} I_{ik,t}I_{jk,t}$. The proportion of common friends with node $v_{j,t}$ from the perspective of node $v_{i,t}$ is calculated as $\frac{\sum\limits_{v_k\in\mathbf{V}} I_{ik,t}I_{jk,t}}{\sum\limits_{v_k\in\mathbf{V}} I_{ik,t}}$.

6.2.1.2 Interaction risk

Given an epidemic outbreak, the node $v_{i,t}$, either healthy or sick, takes the risk of getting infected or infecting the neighbours via interactions. We describe this epidemic spreading process with the "susceptible-infected-recovered" (SIR) model and assume an epidemic spreads from a single node and propagates with a transmissibility rate at ζ^s_t and a recovery rate at ζ^r_t every time an interaction occurs. The node $v_{i,t}$ gets infected at a probability of $\mathrm{pr}^s_t = \zeta^s_t$ given a single interaction with an infected node and, once infected, recovers at a probability of $\mathrm{pr}^h_t \equiv \zeta^r_t$ before another interaction occurs.

We measure the infection risk $\mathrm{R}^H_{i,t}$ of a healthy node $v_{i,t}$ at time t with its probability of getting infected based on the available information about interactions and node infections.

$$\mathrm{R}^H_{i,t} = 1 - \prod_{v_{j,t} \in V_t} (1 - pr^s)^{I_{ij,t}\beta_{j,t}} \tag{6.11}$$

where $\prod_{v_{j,t} \in V_t} (1 - pr^s)^{I_{ij,t}\beta_{j,t}}$ represents the probability of keeping healthy given the interactions $I_{ij,t}$ with the nodes neighbour and the neighbour's infected condition $\beta_{j,t} = 1$.

Similarly, we measure the spreading risk $\mathrm{R}^I_{i,t}$ of an infected node $v_{i,t}$ at time t with its probability of infecting others via interactions.

$$\mathrm{R}^I_{i,t} = 1 - \prod_{v_{j,t} \in V_t} (1 - pr^s)^{I_{ij,t}(1-\beta_{j,t})} \tag{6.12}$$

where $\prod_{v_{j,t} \in V_t} (1 - pr^s)^{I_{ij,t}(1-\beta_{j,t})}$ represents the probability of keeping healthy given the interactions $I_{ij,t}$ with the node's neighbour and the neighbour's healthy condition $\beta_{j,t} = 0$.

6.2.1.3 Interaction utility

Referring to [69], we calculate the interaction utility $\mathrm{U}_{i,t}$ of the node $v_{i,t}$ based on the interaction reward and the risk avoidance.

$$\mathscr{U}_{i,t} = \begin{cases} \mathscr{R}_{i,t}(1 - \mathrm{R}^H_{i,t}) & \text{if } \beta_{i,t} = 1. \\ \delta\mathscr{R}_{i,t}(1 - \mathrm{R}^I_{i,t}) & \text{if } \beta_{i,t} = 0. \end{cases} \tag{6.13}$$

where the utility for node $v_{i,t}$, given its healthy condition $\beta_{i,t} = 1$, is dependent on the infection risk $\mathrm{R}^H_{i,t}$ and the interaction reward $\mathscr{R}_{i,t}$. In contrast, in the infected condition, the node $v_{i,t}$'s utility is directly correlated with interaction reward $\mathscr{R}_{i,t}$, excluding the spreading risk $\mathrm{R}^I_{i,t}$ and depreciated by δ due to the infected condition.

The expected interaction utility $\mathrm{EU}_{i,t}$ depends on the interaction probability, recovery probability and the presumed behavioural changes at t. We denotes the interaction utility based on the node's health status, $\mathrm{U}^H_{i,t}$ for a healthy node and $\mathrm{U}^I_{i,t}$ for an infected node. The expected interaction utility is calculated as

$$\mathrm{EU}_{i,t} = \begin{cases} \mathrm{U}^H_{i,t}(1 - \mathrm{R}^H_{i,t-1}) + \mathrm{U}^I_{i,t}(\mathrm{R}^H_{i,t-1}) & \text{if } \beta_{i,t-1} = 1. \\ \mathrm{U}^I_{i,t}(1 - \mathrm{pr}^h) + \mathrm{U}^H_{i,t}\mathrm{pr}^h & \text{if } \beta_{i,t-1} = 0. \end{cases} \tag{6.14}$$

where $\mathrm{R}^H_{i,t-1}$ and pr^h each describes the probability of this node getting infected and recover over time, given the interaction structures at $t-1$. $\mathrm{U}^H_{i,t}$ and $\mathrm{U}^I_{i,t}$ each represents the individual utility calculated by each node based on their changes of preferences and the available information of the other nodes.

6.2.2 CNS Dynamics

We model the network dynamics based on the interaction scores driven by the features and related preferences. The interaction scores evolve by the preference mutation in response to an epidemic outbreak towards positive feedback related to the individual utility $U_{i,t}$. We present the network growth and its evolution process in section 6.2.2.1 and section 6.2.2.2.

6.2.2.1 Network formation

The unconnected nodes encounter and interact based on the expectations of interaction reward and interaction risks under social capital constraints. The social capital constraint refers to people's maximum investment, due to personal capacity or social restrictions, in social interactions for expected returns [455, 456]. The interaction between a node pair is determined by

$$I_{ij,t} = \begin{cases} 1 & \text{if } r_{i,t}(i,j), r_{j,t}(i,j) \geq 0,\ \$_{i,t} < \hat{\$}_{i,t} \text{ and } \$_{j,t} < \hat{\$}_{j,t}. \\ 0 & \text{if else.} \end{cases} \tag{6.15}$$

where nodes $v_{i,t}$ and $v_{j,t}$ interact when they benefit from their interactions. They tend to use social capital best and select the most promisingly beneficial interactions characterised by higher interaction rewards.The interaction intensity of this node pair also depends on the interaction score:

$$w^{I}_{ij,t} = \begin{cases} \Pi_{ij,t} & \text{if } I_{ij,t} = 1. \\ 0 & \text{if } I_{ij,t} = 0. \end{cases} \tag{6.16}$$

6.2.2.2 Network evolution

The networks evolve with the changes in social relations and people's preferences related to features. In this section, we introduce a preference mutation mechanism while incorporating five mutation styles: (i) inactive mutation where nodes keep initial preferences, (ii) ignorant mutation where nodes randomly mutate preferences, (iii) egocentric mutation that maximises the respective individual utility while unaware of others' mutation actions, (iv) cooperative mutation where bonded nodes cooperate to maximise their overall utilities and (v) collaborative mutation that enables the nodes to transfer social capital and cooperate.

Inactive style describes the nodes' inaction to the changes of interaction utilities and risks. These nodes keep their initial social preferences and make social contact by

calculating the interaction scores for node pairs under their respective social capital constraints.

Ignorant style describes the active nodes who are unaware of the individual utilities and thus randomly select their social preferences related to preferential attachment, homophily and closure effects between positive, zero and negative.

Egocentric nodes optimise their interaction utilities by social preference mutation shortly after interactions while aware of others' social preferences and, at the moment, assuming others' inactiveness in preference mutation. Each egocentric node can evaluate γ sets of random preference changes in what-if scenarios related to the respective preference change and selects the best-performing one to achieve the highest interaction utility. The individual optimisation process is determined as follows:

$$\begin{aligned} &\max_{\mathbf{sDNA}_{i,t}} \ \mathrm{EU}_{i,t} \\ &s.t. \quad 0 \leq \$_{i,t} \leq \hat{\$}_{i,t} \end{aligned} \tag{6.17}$$

Cooperative nodes cooperate and make a group decision on preference changes to maximise their total interaction utilities. These cooperative nodes consider γ attempts in preference mutation, and to evaluate these attempts, conduct what-if analyses under the social capital constraints of each node. The decision making process is shown as follows:

$$\begin{aligned} &\max_{\{\mathbf{sDNA}_{i,t}\}} \ \sum_{v_{i,t} \in \mathbf{V}_t} \mathrm{EU}_{i,t} \\ &s.t. \quad 0 \leq \$_{i,t} \leq \hat{\$}_{i,t} \end{aligned} \tag{6.18}$$

Collaborative nodes collaborate in a group decision on their preference changes. In contrast with the decision-making of the cooperative nodes under the respective individual social capital constraints, collaborative nodes can transfer their social capital between each other, where they make a group decision under the overall social capital constraints. The decision-making process of collaborative nodes is shown as follows:

$$\begin{aligned} &\min_{\{\mathbf{sDNA}_{i,t}\}} \ \sum_{v_{i,t} \in \mathbf{V}_t} \mathrm{EU}_{i,t} \\ &s.t. \quad 0 \leq \sum_{v_{i,t} \in \mathbf{V}_t} \$_{i,t} \leq \sum_{v_{i,t} \in \mathbf{V}_t} \hat{\$}_{i,t} \end{aligned} \tag{6.19}$$

6.2.2.3 Epidemic Transmission with Recovery

The epidemic spreads from the seed nodes $\mathbf{S}_t = \{\cdots, v_{i,t}, \cdots | v_{i,t} \in V_t\}$, propagates with the transmissibility rate at ζ_t^s and a recovery rate at ζ_t^r. As presented in section 6.2.1.2, a healthy node, given a connection with an infected node, gets infected at the probability of $\text{pr}^s = \zeta_t^s$. Correspondingly, the probability of a healthy node v_i ($\beta_{i,t-1} = 0$) keeping healthy ($\beta_{i,t} = 0$) can be calculated as

$$pr(\beta_{i,t} = 1 | \beta_{i,t-1} = 0) = \prod_{v_{j,t} \in V_t} (1 - pr_t^s)^{I_{ij,t}\beta_{j,t}} \tag{6.20}$$

which value depends on the epidemic transmissibility $\text{pr}_t^s = \zeta_t^s$, the interactions $I_{ij,t}$ with each neighbour and the neighbour's infected condition $\beta_{j,t} = 1$.

In contrast, the recovery rate ζ^r denotes the probability of the recovery of any infected node despite of the connections with infected others. The probability of an infected node v_i ($\beta_{i,t-1} = 1$) recovering to be healthy ($\beta_{i,t} = 1$) can be calculated as

$$pr(\beta_{i,t} = 0 | \beta_{i,t-1} = 1) \equiv \zeta^r \tag{6.21}$$

where the recovery probability keeps equal to the recovery rate at ζ^r.

6.3 Experiment Design, Results and Analysis

In this section, we conduct simulation-based experiments with the proposed DT-CNS model and investigate the impact of different social capital constraints and heterogeneous preference mutation styles on the evolving interaction and infection patterns. Section 6.3.1 illustrates the initialised parameters of the DT-CNS models. Section 6.3.3 and section 6.3.2 present the interaction and infection patterns in different interaction scenarios. We calculate and compare the infection utilities in section 6.3.4 based on the interaction and infection patterns.

6.3.1 Initial Parameters

We simulate a 30-node network interacting in an epidemic outbreak for 100 iterations based on their interaction utilities related to preferences for their respective features and epidemic risks. Tab. 6.3.1 shows the initial parameters related to the nodes' features, preferences, and epidemic transmissibility.

As shown in Tab. 6.3.1, the 30 nodes are each denoted by integer values ranging from 0 to 29. They have a single feature simulated by a random uniform distribution within

Table 6.1: The initialised parameters for the DT-CNS models.

CNS Component	Parameter	Meaning	Value
Network	i	Node identity	$[0,1,2,\cdots,29]$
	$\mathbf{f}$	Feature	$\mathscr{U}[0,1]$
	$\mathbf{p}$	Preference for features	[0]
	$\mathbf{w}^p$	Weights of preference for features	[1]
	$\mathbf{h}$	Preference for feature differences	[0]
	$\mathbf{w}^h$	Weights	[1]
	$\mathbf{c}$	Preference for transitivity	[0]
	$\mathbf{w}^c$	Weights of preference for transitivity	[1]
	η	Encounter rate	1.00
	u^0	Initial interaction cost	0.40
	$\eta_{i,t}$	Infection penalty	1.5
	γ	Length of action space per node	2
	δ	Interaction utility discount	0.01
	ϵ	Scaling parameter for interaction utility	1.00
Process	η^s	Infection rate	0.20
	η^r	Recover rate	0.20

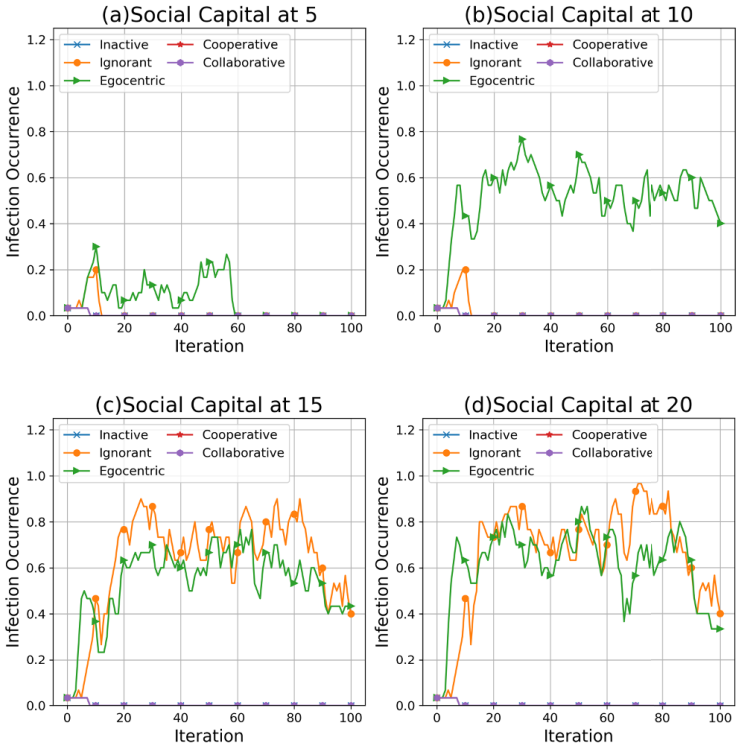
$[0,1]$. In the initialisation stage, these nodes are unconnected and have zero preferences regarding preferential attachment, homophily and closure effects. The corresponding preference weights keep fixed at 1.

Based on these features and preferences, the nodes encounter each other given an encounter rate of $\eta = 1$ and evaluate the node pairs based on the interaction scores. The interaction cost initialised for the healthy nodes keeps at 0.4, which turns into 0.6 under the infection penalty of $\eta = 1.5$ when the nodes get infected. The epidemic spread from the node denoted as 0, and under its influence, the networks evolve by preference mutation and re-evaluation of the pairwise interactions.

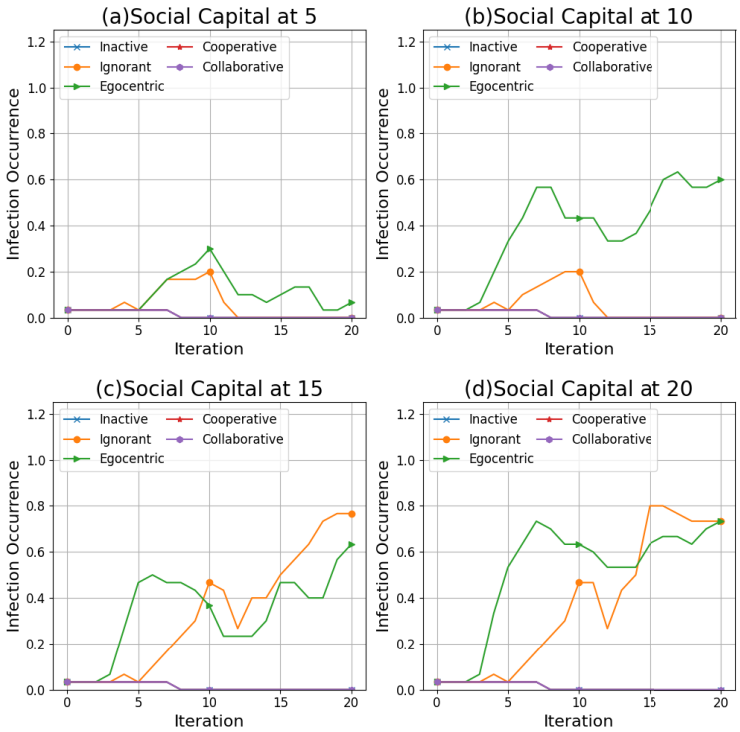
6.3.2 Infection Patterns

We investigate the infection patterns with a mutual influence on the evolving social networks according to the infection occurrence (proportion of infected nodes to the overall 30 nodes) in the social networks (Fig. 6.1).

As shown in Fig. 6.1, with the increased social capital limit, more infections occur in the epidemic spreading process, where some nodes get infected and some recover. Compared with the other node types, the ignorant and the egocentric nodes are more likely to be infected. Their infection occurrences, given the social capital limit at 15 and 20, fluctuate around 0.70 from the twentieth iteration. This indicates that around 70% of the nodes tend to be infected as the epidemic propagates for more than 20 iterations.This can be caused by the ignorance of group utilities and the unawareness



(a) Infection occurrences in 100 iterations



(b) Infection occurrences in first 20 iterations

Figure 6.1: The infection occurrences in 100 iterations (a) and first 20 iterations (b) in the evolving social networks that are respectively driven by inactive, ignorant, egocentric, cooperative and collaborative mutation styles.

of others' decisions in individual decision-making processes. In contrast, the inactive, cooperative and collaborative nodes have insignificant infection occurrences. These nodes keep healthy since the tenth iteration (Fig. 6.1 (b)). To better understand the interrelations between the abovementioned infection patterns and the nodes' interactions, we present and compare the nodes' interaction patterns driven by different preference mutation styles in section 6.3.3.

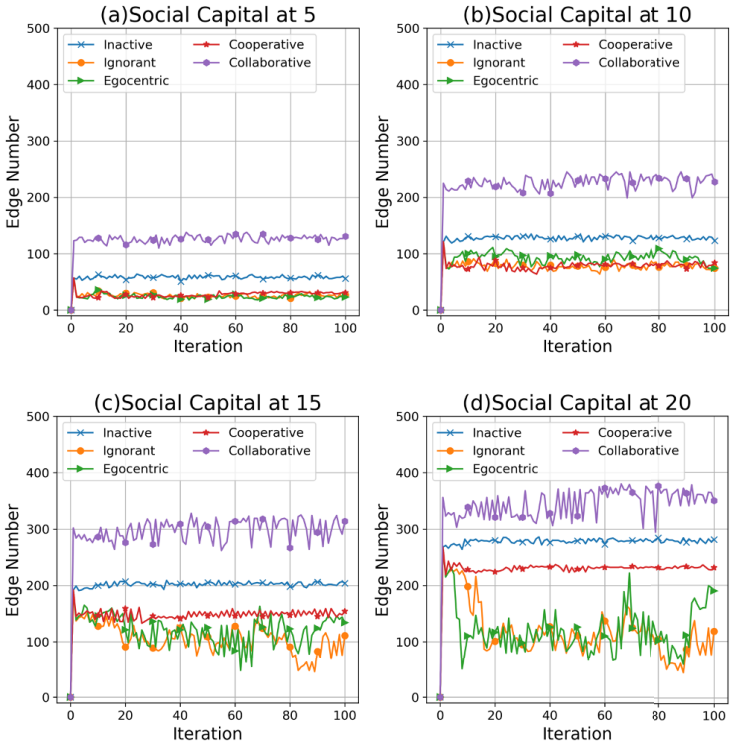
6.3.3 Interaction Patterns

This section investigates the interaction patterns in an epidemic outbreak simulated over 100 iterations. Interactions in the generated networks are driven by different preference mutation styles, including inactive, ignorant, egocentric, cooperative, and collaborative styles. We present the visualisations of the evolving social networks in first 20 iterations in Appendix D.1 and take the most typical ones as examples in the discussion of interaction structures (section 6.3.3.1). In addition, we compare the social networks driven by different mutation styles by analysing the nodes' degree values (section 6.3.3.2) and clustering coefficients (section 6.3.3.3), and node pairs' shortest path lengths (section 6.3.3.4). To better understand the differences between the interaction structures of the evolving social networks driven by different preference mutation styles, we also calculate the Euclidean differences between these network statistics.

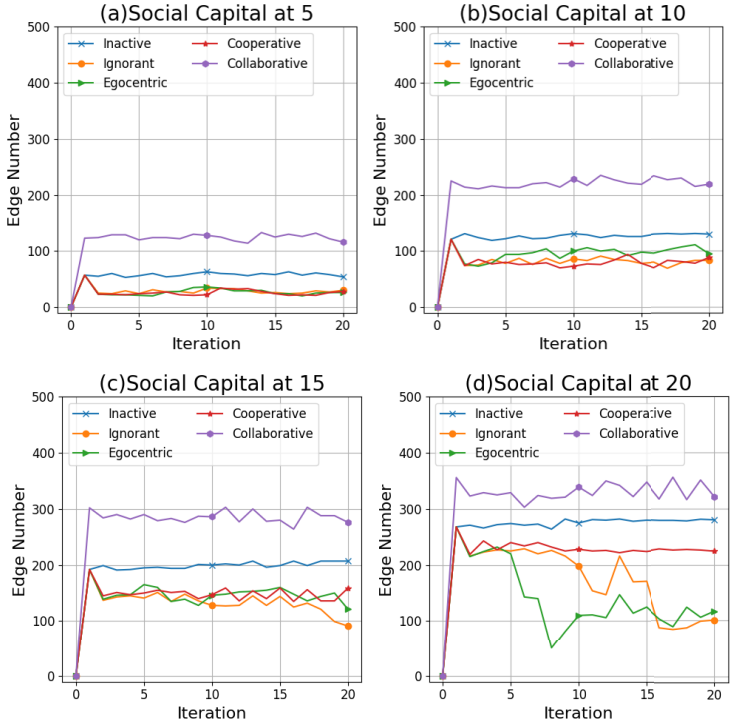
6.3.3.1 Interactions

Fig. 6.2 (a) and Fig. 6.2 (b) show impact on the interactions of different social capital constraints and preference mutation styles in 100 iterations and first 20 iterations. To illustrate the phenomenon presented in Fig. 6.2, we respectively discuss these two aspects based on typical examples of social network simulations presented in Fig. 6.5, Fig. 6.4 and Fig. 6.6.

When the social capital limit increases, the number of interactions generally increases as nodes can interact more with others (Fig. 6.2). Considering different preference mutation styles, collaborative nodes have the highest interaction numbers over the iterations. This can be caused by the flexible social capital transfer between the collaborative nodes, which enables the creation of more edges with the preferred nodes. In addition, the interaction numbers for each simulated network increase significantly in the first iterations and then fluctuate around a fixed value afterwards, except for the networks driven by the ignorant and the egocentric preference mutation styles under the social capital limit at



(a) Interaction number in 100 iterations



(b) Interaction Number in first 20 iterations

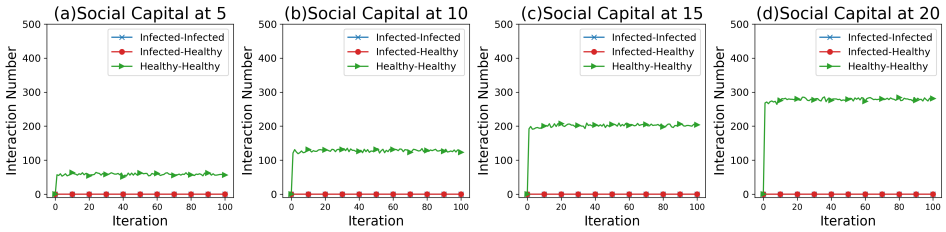
Figure 6.2: The number of interactions in the 100 simulations steps (a) and first 20 iterations (b) in the evolving social networks that are respectively driven by inactive, ignorant, egocentric, cooperative and collaborative mutation styles.

15 and 20. This can result from the significant number of infected nodes (Fig. 6.1 (a)) in the ignorant and the egocentric network simulations, which increases the corresponding interaction cost and risk, reducing the interactions with infected nodes. In contrast, due to insignificant infection occurrence and low infection risks, the interaction numbers in networks driven by the inactive, cooperative, and collaborative mutation styles remain steady without significant fluctuations.

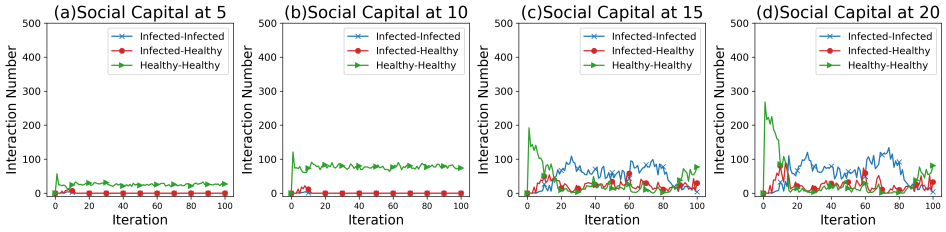
To better understand the interrelations between the interaction numbers and the infection patterns, we present the interaction numbers between nodes with different health statuses in the evolving social networks driven by different mutation styles (Fig.6.3). In networks driven by the inactive, the cooperative and the collaborative preference mutation styles (Fig.6.3 (a), Fig.6.3 (d) and Fig.6.3 (e)), social interactions generally take place between the healthy nodes. In contrast, in social network simulations driven by the ignorant and the egocentric preference mutation styles (Fig.6.3 (b) and Fig.6.3 (c)), there are interactions with the infected nodes over the iterations and some of them take place between healthy and infected nodes. These interactions increase significantly as the social capital limit increases from 10 to 20, resulting from and, in turn, reinforcing a high infection occurrence (Fig. 6.1).

Since the social networks present different interaction numbers and infection occurrences in the first 20 iterations and afterwards they stay stable with small fluctuations, we present the social network simulations of the ignorant nodes (Fig. 6.5) and the collaborative nodes (Fig. 6.4) in the first 20 iterations under different social capital limits as an example to further explore the network structures and epidemic propagation. As shown in Fig. 6.4, the networks get denser with more interactions given the increased social capital limits. The nodes collaborate to avoid contact with the infected seed node and interact most by reaching their respective social capital limits. In contrast, in Fig. 6.5, the networks in the first ten iterations get denser with more interactions given the increased social capital limits. However, under the social capital limits at 15 and 20, the networks get sparser since the tenth iteration with less interactions and more infections. The decreased interaction number can be caused by the increase in infection occurrence (Fig. 6.5 (c) and (d)). In addition, as shown in Fig. 6.5 (a) and (b), under the social capital limit at 5 and 10, there are fewer infections due to a limited number of interactions and exposures to the epidemic spread. This indicates that a lower social capital limit enables to avoid specific interaction risks and limit the impact of epidemic spread on interaction patterns.

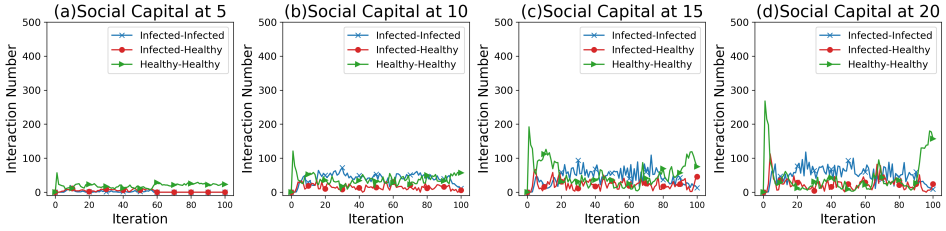
We also find from Fig. 6.2 that under the same social capital limit, nodes in the



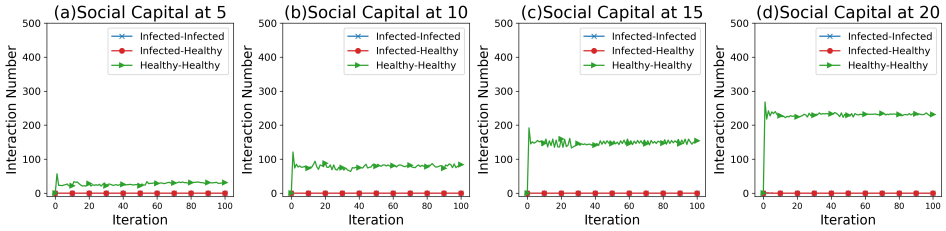
(a) Inactive



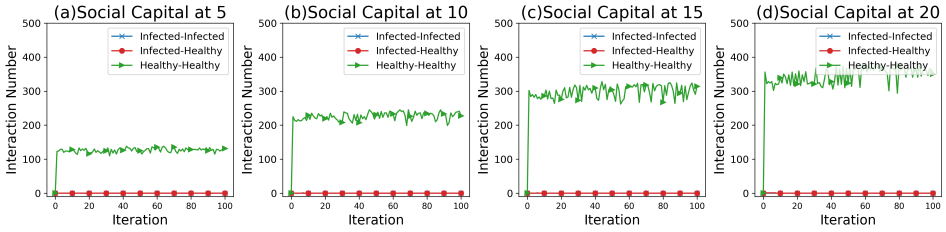
(b) Ignorant



(c) Egocentric

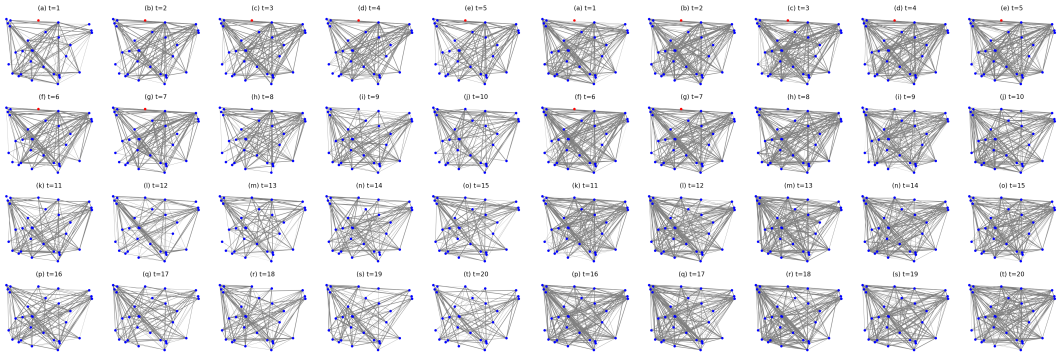


(d) Cooperative



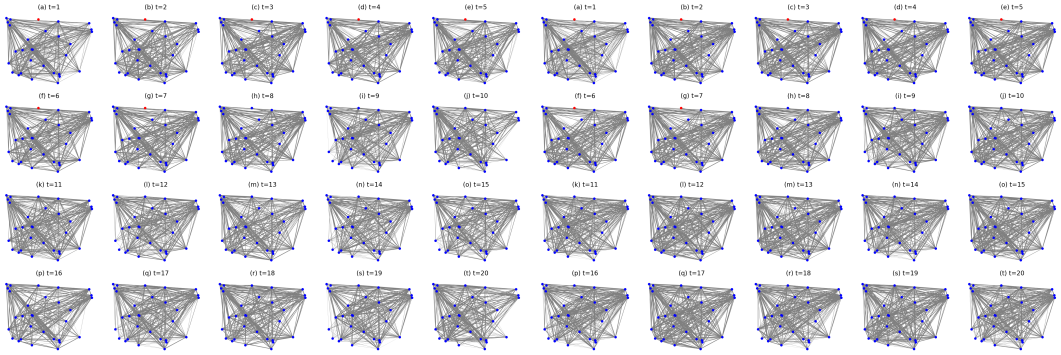
(e) Collaborative

Figure 6.3: The interaction numbers' changes between nodes in each health statuses in evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.



(a) Social Capital Limit at 5

(b) Social Capital Limit at 10

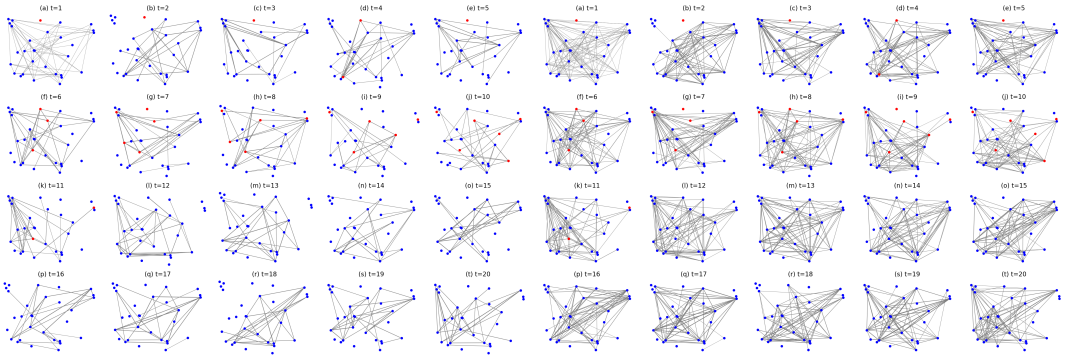


(c) Social Capital Limit at 15

(d) Social Capital Limit at 20

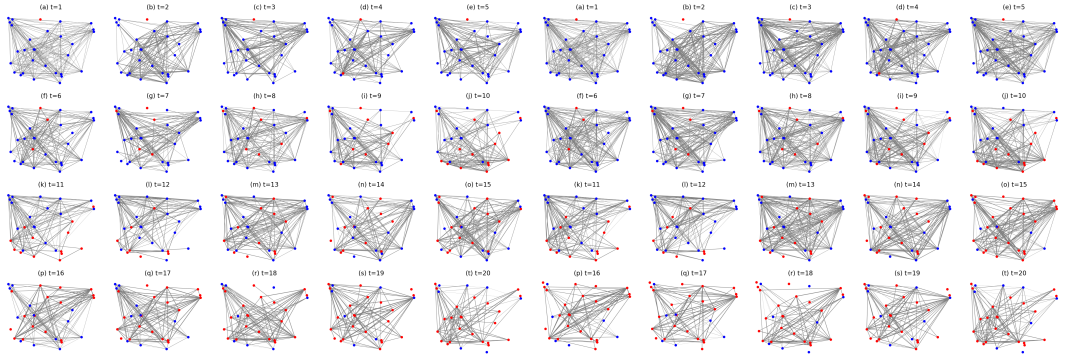
Figure 6.4: The evolving social networks driven by collaborative preference mutation styles under the impact of different social capital limits.

inactive, ignorant and cooperative mutation styles have different number of interactions. To better illustrate this phenomenon, we take the social network simulations under the social capital limit of 10 and driven by different mutation styles as an example (Fig. 6.6). In Fig. 6.6 (e), collaborative nodes interact more than other node types. This is because there are few infections, and the collaborative nodes can transfer social capital to enable more interactions. The other types of nodes engage in a similar number of interactions but have different interaction structures and infection patterns. In Fig. 6.6 (c), the egocentric nodes have the highest infection occurrence and a significant number of interactions with infected nodes. In contrast, there are fewer infections and limited interactions with the infected nodes in the networks driven by ignorant preference mutation styles. Moreover, there is no social contact with the infected nodes in the networks driven by the inactive and the cooperative preference mutation styles. The abovementioned phenomenon indicates that the egocentric nodes maximise their individual utilities by expanding social contact even though they are infected, detrimental to the utilities of others.



(a) Social Capital Limit at 5

(b) Social Capital Limit at 10



(c) Social Capital Limit at 15

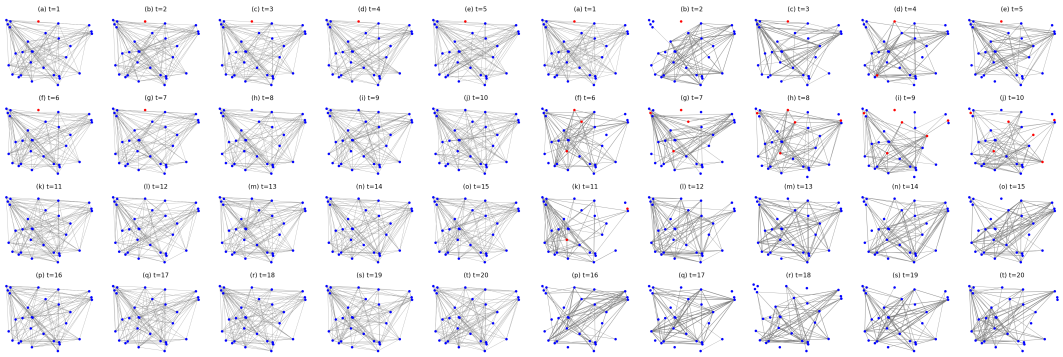
(d) Social Capital Limit at 20

Figure 6.5: The evolving social networks driven by ignorant preference mutation styles under the impact of different social capital limits.

6.3.3.2 Degree Distribution

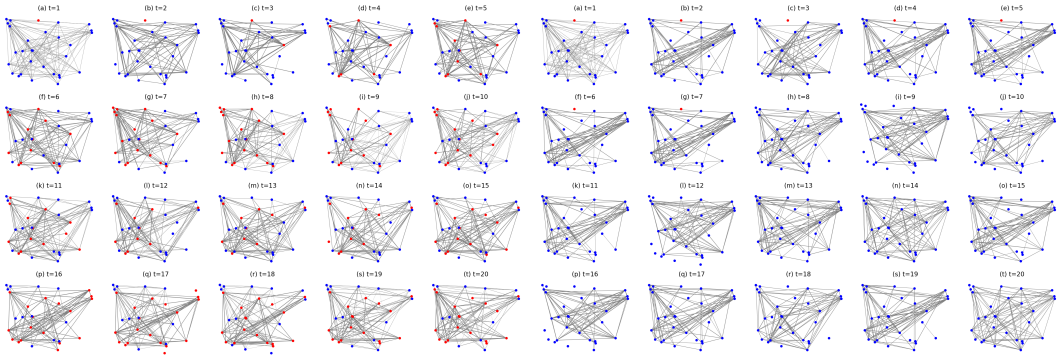
We compare the node degrees of the evolving social networks driven by different preference mutation styles over iterations, including the node degree distributions (Fig. 6.7), average node degrees of the infected nodes and the healthy nodes (Fig. 6.8) and the Euclidean distance values between node degrees in different social networks (Fig. 6.9).

As shown in Fig. 6.7, the node degree values increase with the increasing social capital limit. We take the social networks driven by ignorant nodes as an example (Fig. 6.7 (a)). Given a social capital limit of 5, a significant number of ignorant nodes' degrees are less than 5 and less than 10 nodes' have a degree value at 5, reaching the social capital limit. The corresponding node degree distribution for ignorant nodes changes as the social capital increases, with more observations of node degrees over 5. In addition, compared with other node types, the collaborative nodes have a generally higher node degree values (Fig. 6.7 (e)). The collaborative nodes' degrees can be higher than the social capital limit as they can transfer social capital flexibly, as nodes with higher social capital can have more social interactions and higher node degrees. The



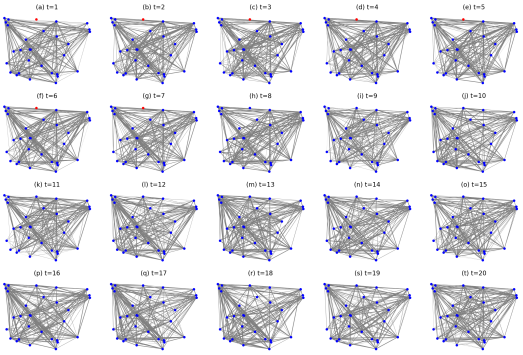
(a) Inactive

(b) Ignorant



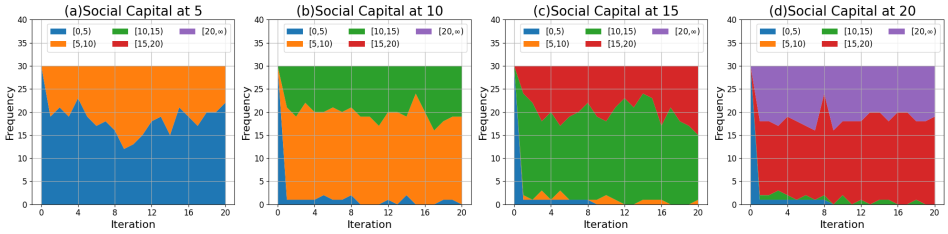
(c) Egocentric

(d) Cooperative

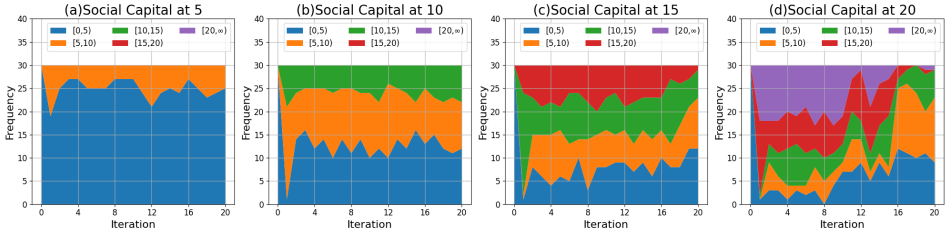


(e) Collaborative

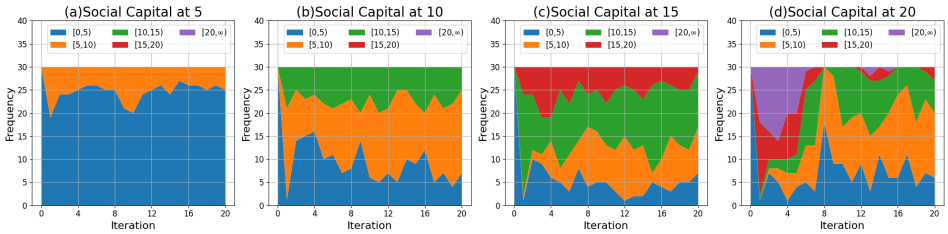
Figure 6.6: The evolving social networks driven by different mutation styles under the social capital limit at 10. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.



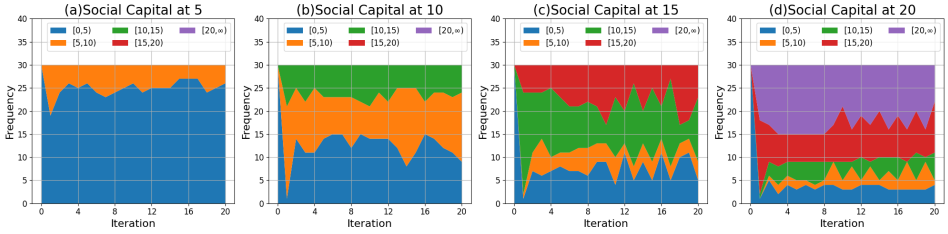
(a) Inactive



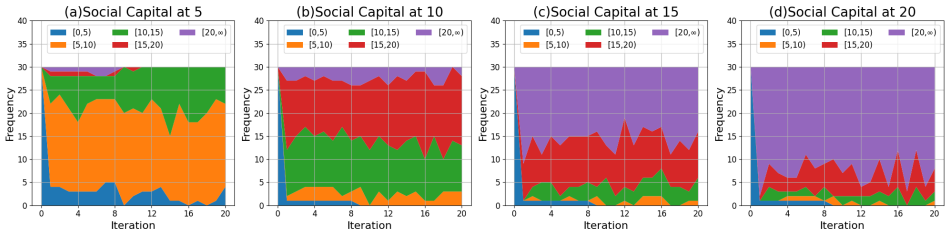
(b) Ignorant



(c) Egocentric



(d) Cooperative



(e) Collaborative

Figure 6.7: The degree distribution of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.

collaborative nodes generally have more interactions and correspondingly, higher node degrees.

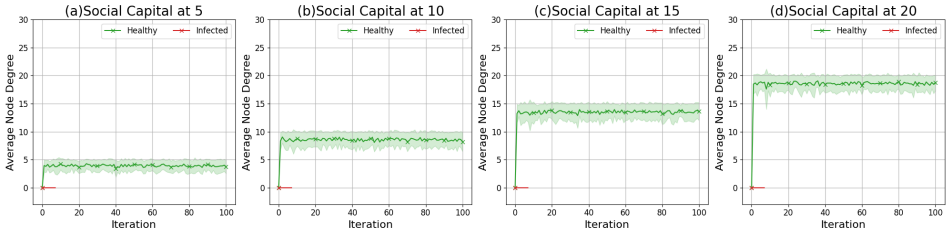
To further analyse the impact of infection patterns on the node degrees, we calculate the average values and standard deviations for the infected and healthy nodes (Fig. 6.8). The standard deviations are presented as the width of the shade around average node degrees. In social networks driven by the inactive, cooperative and collaborative preference mutation styles, generally, all the nodes keep healthy over 100 iterations. Their average node degrees keep steady with small fluctuations since the second iteration. In contrast, the average node degrees for healthy ignorant nodes and healthy egocentric nodes increase over the first iteration and afterwards gradually decrease. This can be caused by the decreasing interactions between healthy nodes (Fig. 6.3) and the increasing interaction occurrences in the epidemic outbreak (Fig. 6.1). In the meantime, the average node degrees for the infected ignorant nodes and the infected egocentric nodes fluctuate around the similar average node degree values with that of healthy nodes (Fig. 6.8 (b) and Fig. 6.8 (c)). We also find that the standard deviations of infected nodes are higher than those of healthy nodes when there are more infected nodes in social networks. Moreover, the standard deviations of node degrees for inactive nodes are significantly smaller than that of others. This is because each inactive node shares a similar connection pattern, where they randomly interact without specific connection preferences, in contrast to other node types.

As shown in Fig. 6.9, we find that the differences between the degree distributions increase significantly when the social capital limit increases (Fig. 6.9). This can be caused by increasing social interactions where nodes can connect with more people and make different choices. Within each subplot, there are more visible differences between the degree distributions of collaborative social networks and other social networks. This is because of the flexible selection of social connections given social capital transfer in group decisions when it comes to the collaborative nodes.

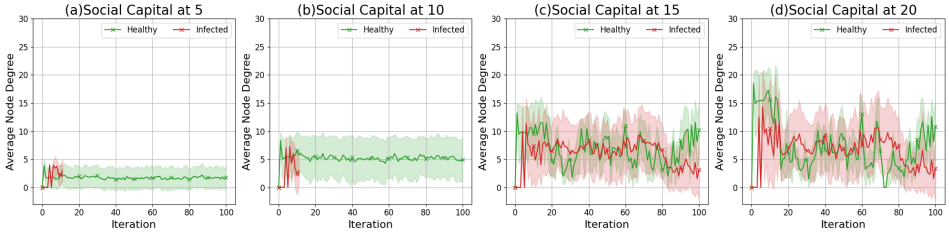
6.3.3.3 Clustering Coefficient Distribution

We compare the node clustering coefficients of the evolving social networks driven by different preference mutation styles over iterations, including the node clustering coefficient distributions (Fig. 6.10), average node clustering coefficients of the infected nodes and the healthy nodes (Fig. 6.11) and the Euclidean distance values between the node clustering coefficients in different social networks (Fig. 6.12).

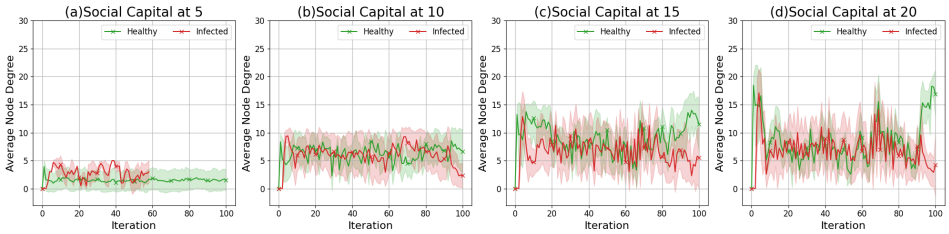
As shown in Fig. 6.10, the node clustering coefficient increases as the social capital



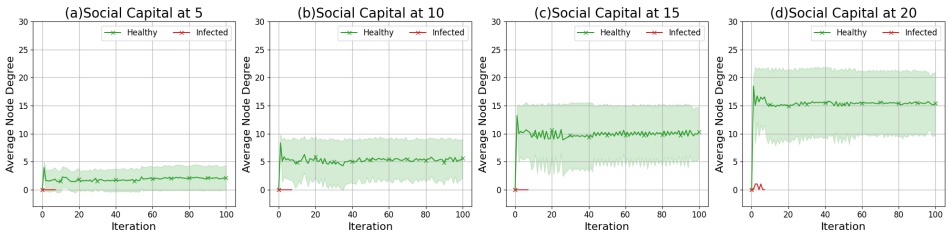
(a) Inactive



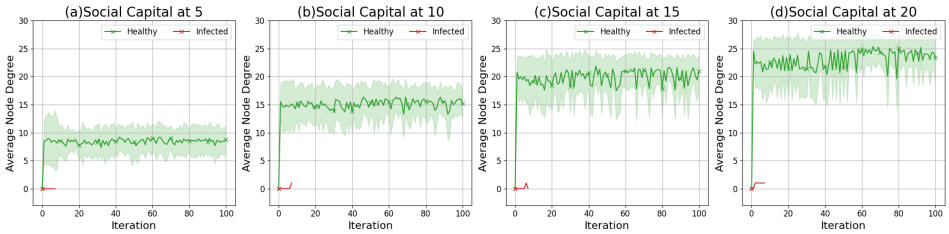
(b) Ignorant



(c) Egocentric

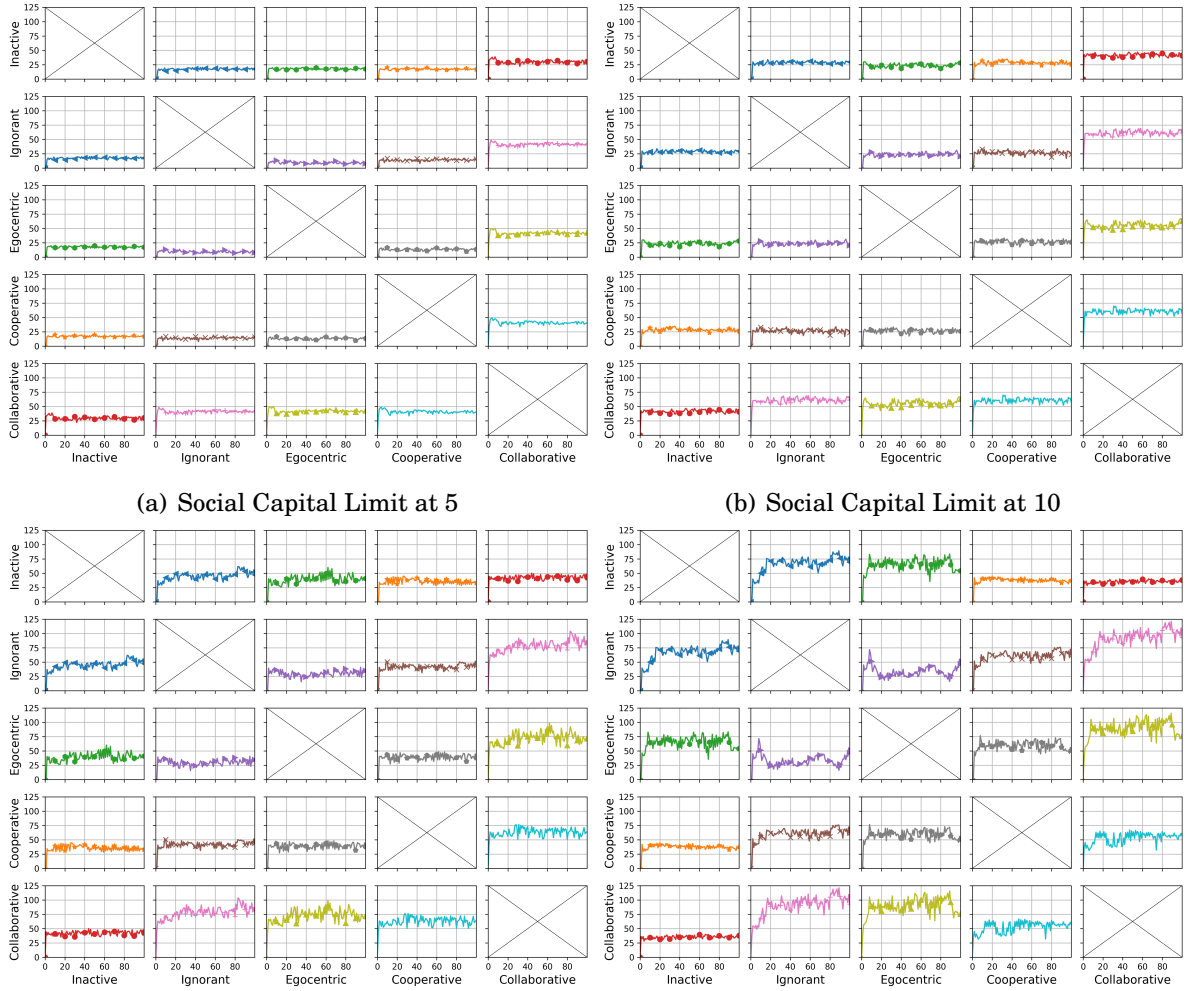


(d) Cooperative



(e) Collaborative

Figure 6.8: The average node degree for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.



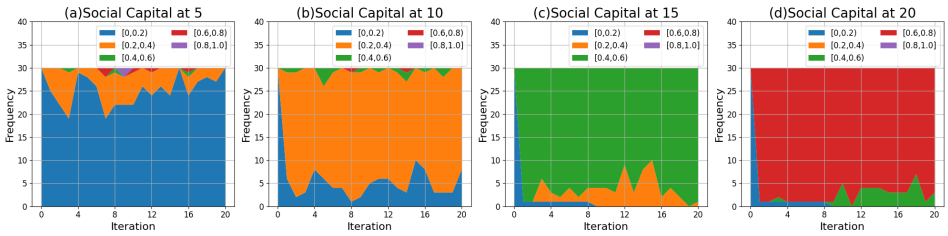
(a) Social Capital Limit at 5

(b) Social Capital Limit at 10

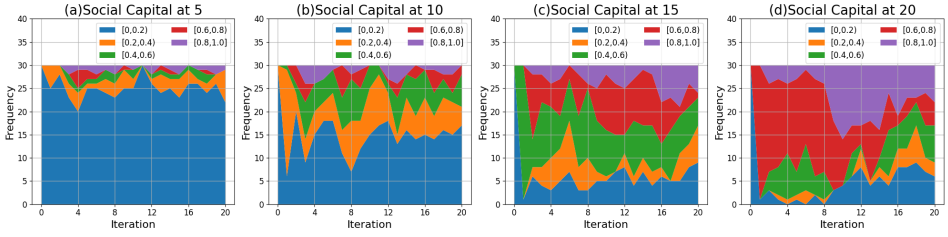
(c) Social Capital Limit at 15

(d) Social Capital Limit at 20

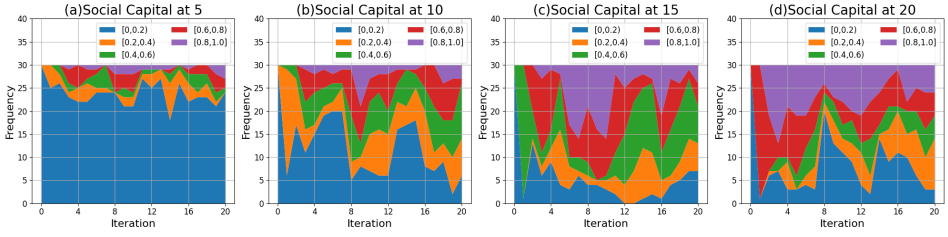
Figure 6.9: The Euclidean distances between the degree distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance.



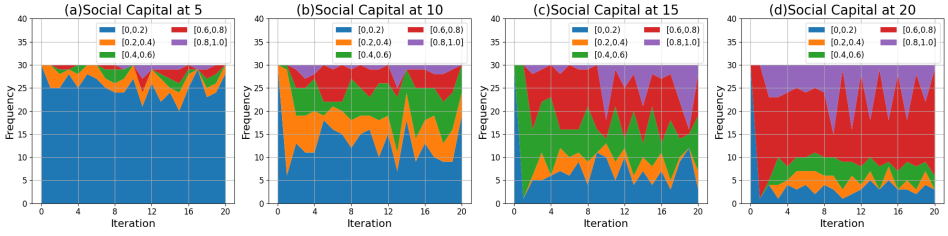
(a) Inactive



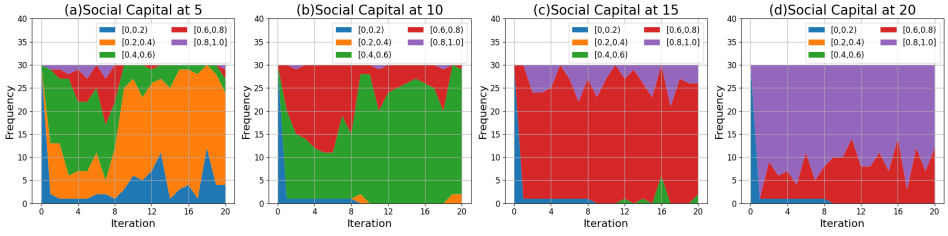
(b) Ignorant



(c) Egocentric



(d) Cooperative



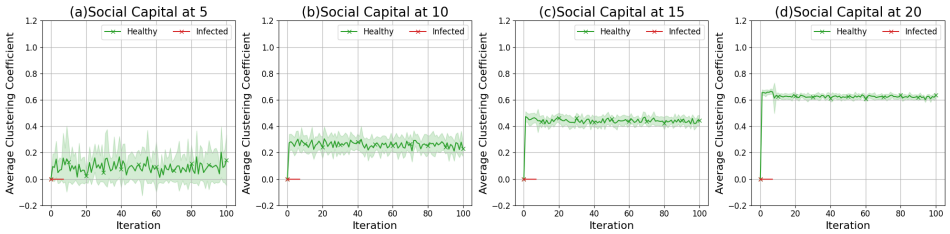
(e) Collaborative

Figure 6.10: The clustering coefficient distribution of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.

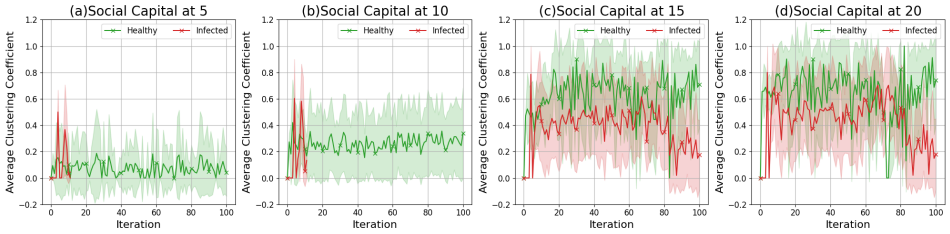
limit increases. We take the social networks driven by cooperative nodes as an example (Fig. 6.10 (d)). Given a social capital limit of 5, more than 20 cooperative nodes' clustering coefficients are less than 0.20. As the social capital limit increases, more node clustering coefficients take a value larger than 0.20, and some range between $[0.80, 1.00]$. In contrast, the inactive nodes' clustering coefficient keep valuing lower than 0.8. Compared with other node types, the collaborative nodes have a generally higher node clustering coefficient values (Fig. 6.7 (e)). The abovementioned phenomenon implies that the social capital limits significantly impact the node clustering coefficient. Compared with other node types, collaborative nodes are more likely to cluster around the preferred nodes as they can transfer social capital to others.

In Fig. 6.11, we analyse the impact of infection patterns on the node clustering coefficients by presenting their average values and standard deviations for the infected and healthy nodes. In social networks driven by the inactive and collaborative preference mutation styles (Fig. 6.11 (a) and Fig. 6.11 (e)), generally, all the nodes keep healthy since the sixth iteration, where their node clustering coefficients have a steady average value and a small standard deviation than that of other node types. Compared with the collaborative nodes, the cooperative nodes have smaller average node clustering coefficients but larger standard deviations (Fig. 6.11 (d)). This implies the diversity of node clustering coefficients for cooperative nodes, which can also be seen in Fig. 6.10 (d). In contrast to the inactive, cooperative and the collaborative nodes, there is a significant number of infections among the ignorant and the egocentric nodes (Fig. 6.1). The average node clustering coefficients for healthy ignorant nodes and healthy egocentric nodes are generally lower than that of their infected peers (Fig. 6.11 (b) and Fig. 6.11 (c)). This indicates that the ignorant and egocentric nodes are more likely to cluster around the healthy nodes than the infected ones. In turn, it also implies that healthy ignorant and healthy egocentric nodes are more likely to interact with others. This is because the healthy nodes have lower interaction costs than the infected nodes, which influence the interactions of each node type despite their differences in preference styles.

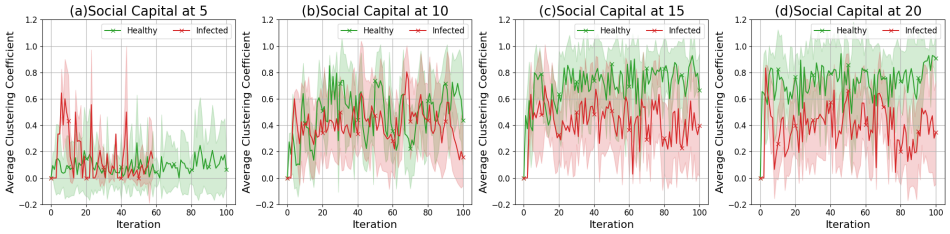
When it comes to the comparison between clustering coefficient distributions (Fig. 6.12), we find that the differences between them are significant in the first interactions and then slightly fluctuate over time. We find no significant changes in differences between Euclidean Distances when the social capital limit increases. This indicates that increasing interactions can not significantly influence the differences in clustering patterns. In addition, we find that the Euclidean Distances between the collaborative nodes and the egocentric/ignorant nodes are significantly higher than that between other evolving



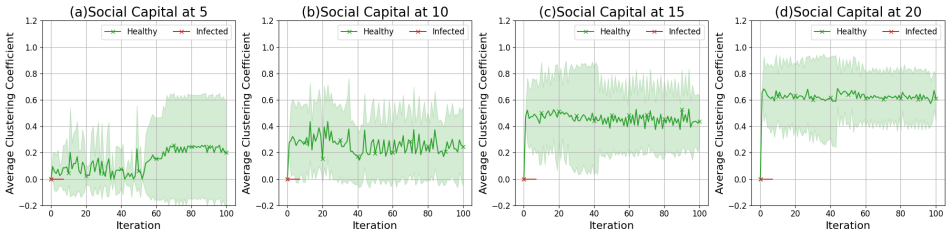
(a) Inactive



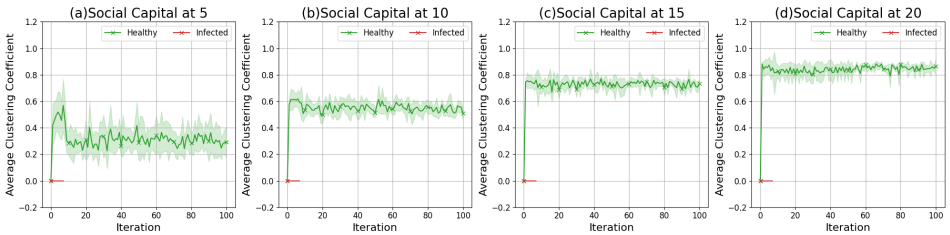
(b) Ignorant



(c) Egocentric

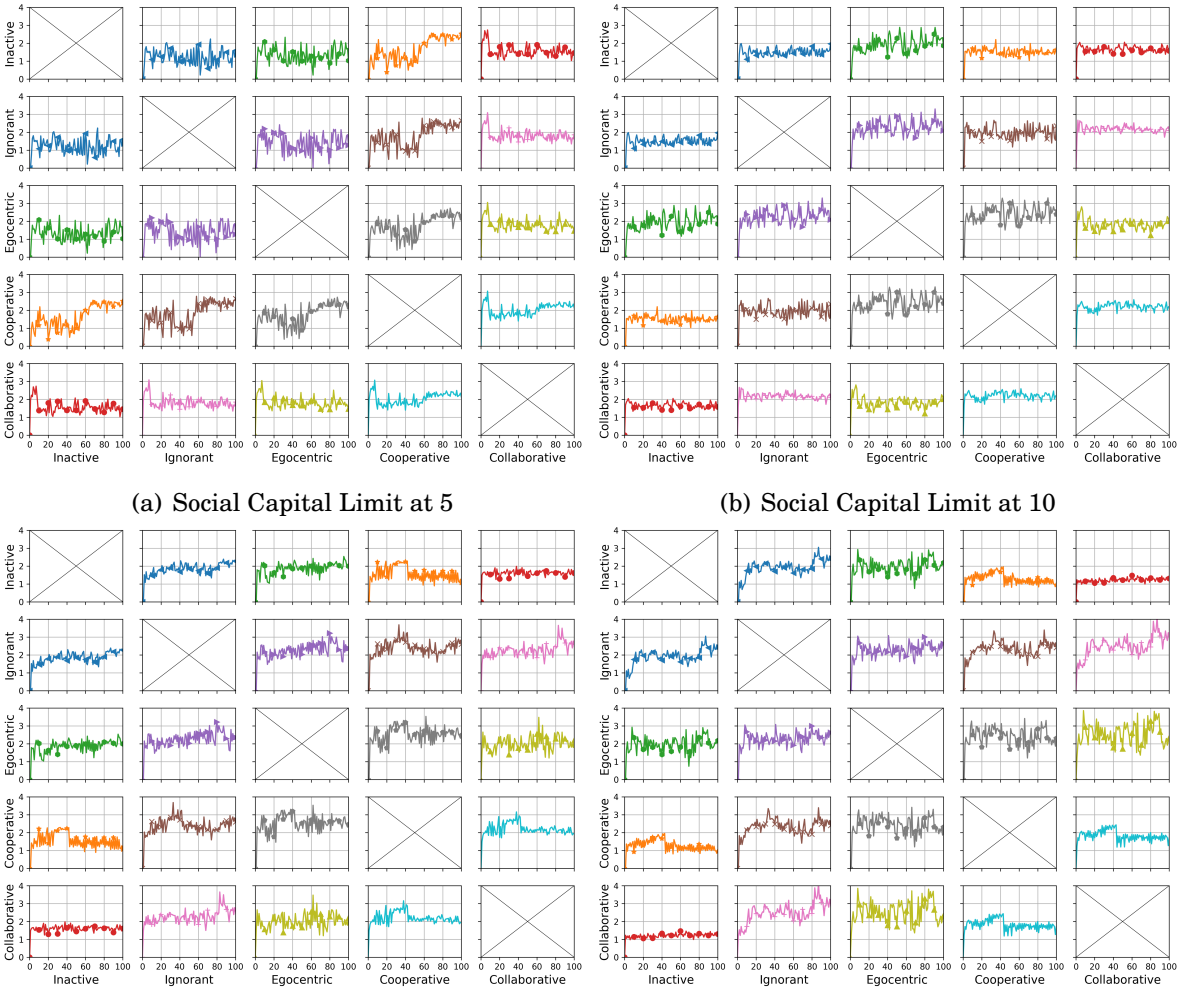


(d) Cooperative



(e) Collaborative

Figure 6.11: The average node clustering coefficient for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.



(a) Social Capital Limit at 5

(b) Social Capital Limit at 10

(c) Social Capital Limit at 15

(d) Social Capital Limit at 20

Figure 6.12: The Euclidean distances between the clustering coefficient distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance.

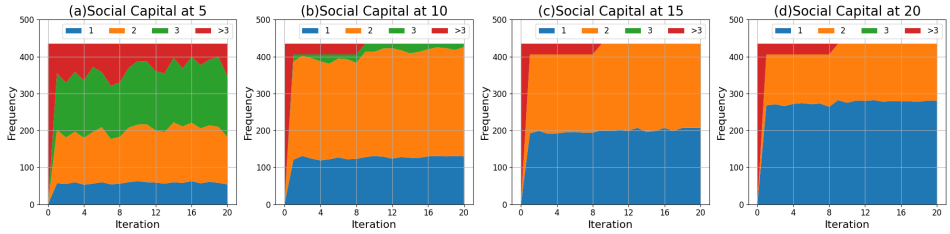
social networks. The collaborative nodes generally keep healthy over the iterations and can transfer social capital to others. This leads to node clusters around the nodes with preferred features, different from the cases of ignorant nodes and egocentric nodes.

6.3.3.4 Shortest Path Length Distribution

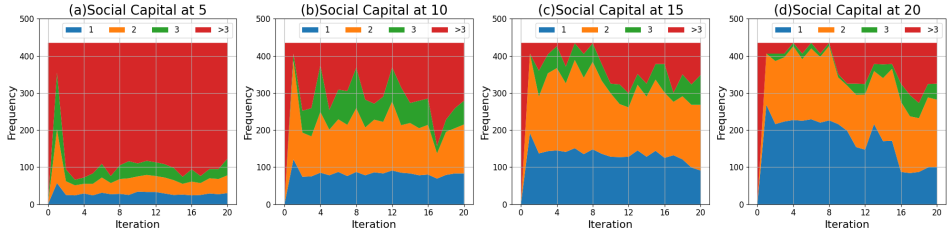
We compare the shortest path lengths of the evolving social networks driven by different preference mutation styles over iterations, including the shortest path length distributions (Fig. 6.13), average shortest path lengths of the node pairs considering their health statuses (Fig. 6.14) and the Euclidean distance values between shortest path lengths in different social networks (Fig. 6.15).

As shown in Fig. 6.13, the shortest path length generally decreases as the social capital limit increases. We take the social networks driven by inactive nodes as an example (Fig. 6.13 (a)). Given a social capital limit of 5, the shortest path lengths between half of all inactive node pairs are 3 or over 3. As the social capital limit increases, the average shortest path lengths get shorter, and most ignorant node pairs have shortest path lengths below 3. Given a social capital limit at 20, around 280 node pairs are directly connected, with an average shortest path length at 1. In addition, compared with other node types, the collaborative nodes have a generally lower shortest path length (Fig. 6.13 (e)). This implies that the social capital limits also significantly impact the shortest path lengths. Moreover, collaborative nodes are more likely to be more tightly connected than the other node types.

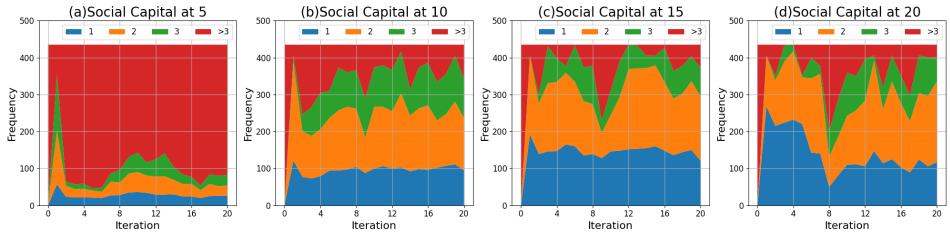
In Fig. 6.14, we analyse the impact of infection patterns on the shortest path lengths by presenting their average values for the node pairs considering their health statuses. In social networks driven by the inactive, the cooperative and the collaborative preference mutation styles (Fig. 6.14 (a), Fig. 6.14 (d) and Fig. 6.14 (e)), generally, all the nodes keep healthy since the sixth iteration. Thus, we mainly present the average shortest path lengths between healthy nodes, in contrast to ignorant and egocentric nodes. Among different node types, the average shortest path lengths for inactive nodes and collaborative nodes are lower than 5 given any social capital limits, much smaller than that of other node types. This implies that there are fewer unconnected node pairs, and the inactive/collaborative nodes tend to be tightly connected despite the changes in social capital limit. In contrast, the average shortest path lengths between the cooperative nodes decrease with an increased social capital limit, indicating that more unconnected pairs get connected. Regarding the ignorant and egocentric nodes, the average shortest path lengths between the node pairs in respective health statuses fluctuate significantly.



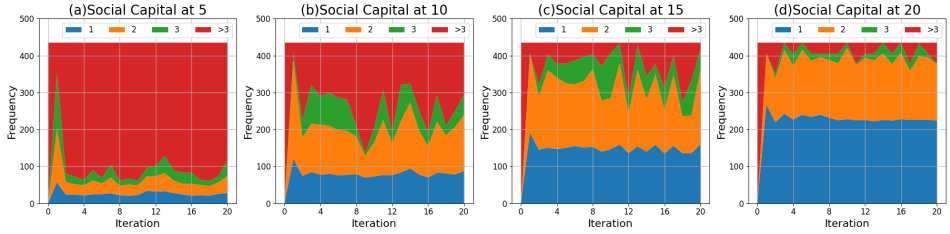
(a) Inactive



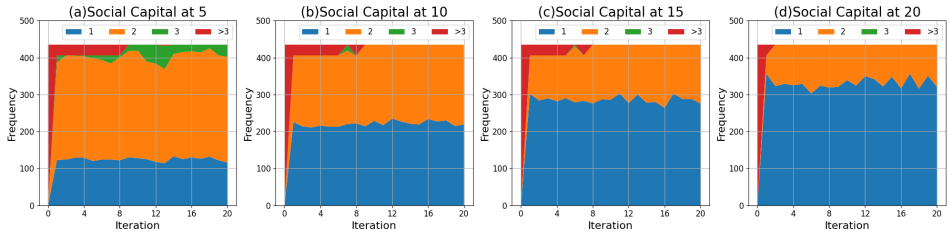
(b) Ignorant



(c) Egocentric

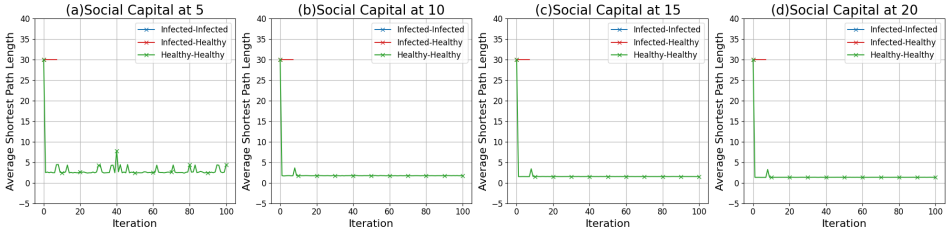


(d) Cooperative

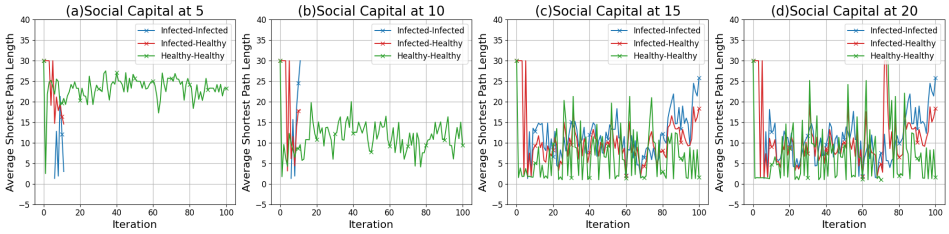


(e) Collaborative

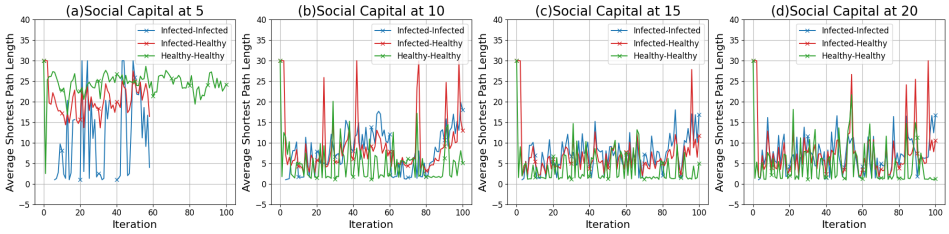
Figure 6.13: The shortest path length distributions of evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.



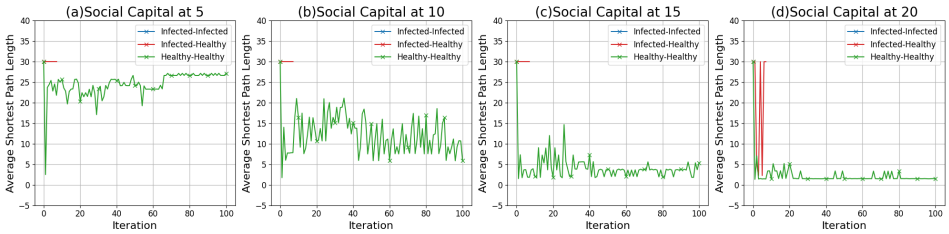
(a) Inactive



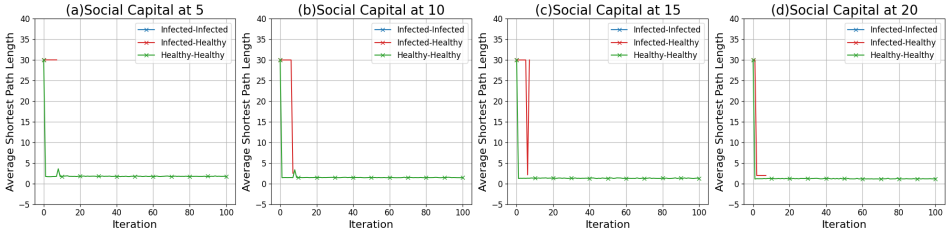
(b) Ignorant



(c) Egocentric



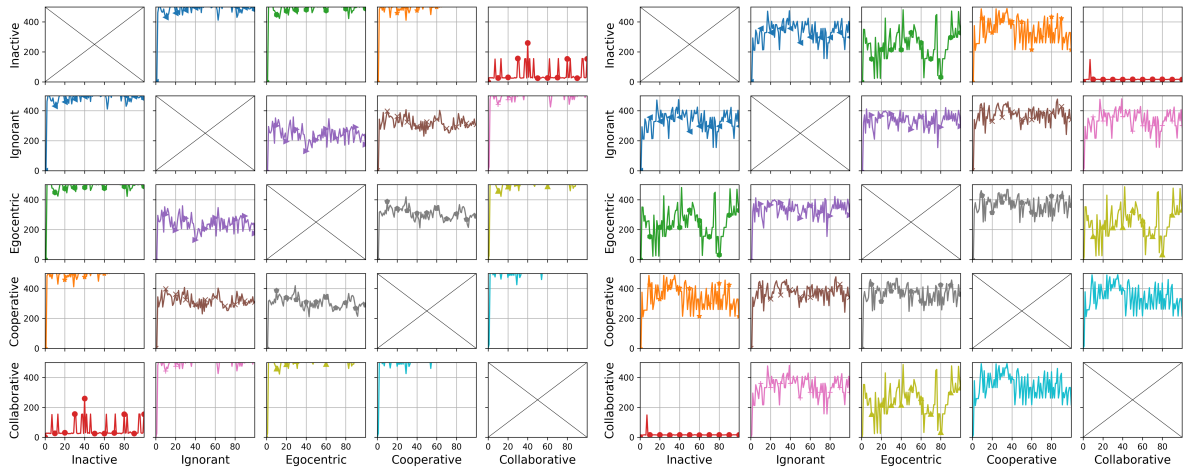
(d) Cooperative



(e) Collaborative

Figure 6.14: The average shortest path length for infected nodes and healthy nodes in the evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.

This can result from the fluctuating infection occurrence (Fig. 6.1). In addition, given a social capital limit over 5, the average shortest path lengths between the healthy nodes fluctuate significantly and sometimes reach a lower value of around 1 when compared with the shortest path lengths between the infected and healthy nodes. This implies that the ignorant and the egocentric nodes can avoid the interaction risks by reducing tight connections with the infected nodes. This is because the infected nodes have higher interaction costs, suppressing the corresponding social interactions.



(a) Social Capital Limit at 5

(b) Social Capital Limit at 10

(c) Social Capital Limit at 15

(d) Social Capital Limit at 20

Figure 6.15: The Euclidean distances between the shortest path length distributions of the evolving interaction networks under different social capital limits, driven by different mutation styles including inactive, ignorant, egocentric, cooperative and collaborative mutation styles. In each subplot, the x-axis represents the number of interactions ranging from 0 to 20, and the y-axis represents the Euclidean distance.

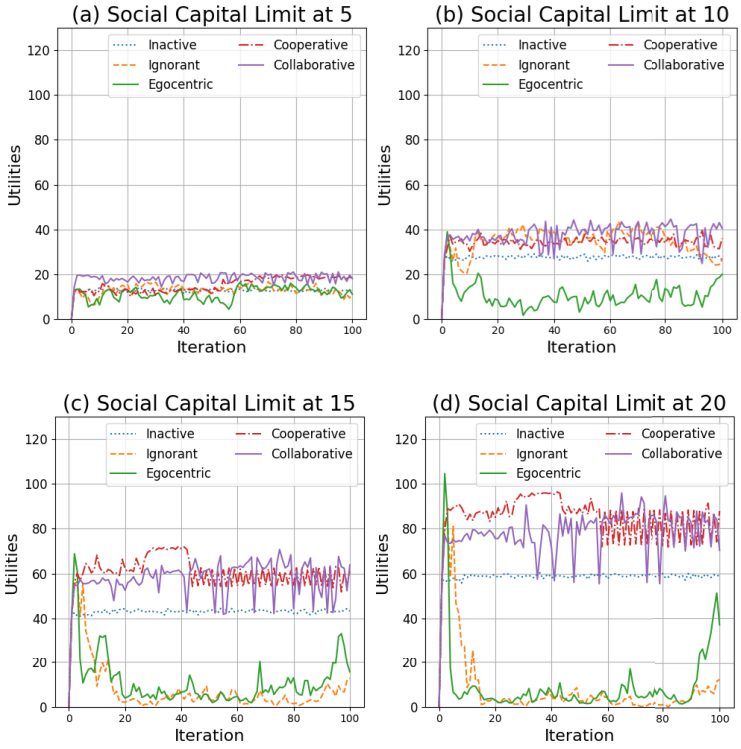
Regarding the comparison between shortest path length distributions (Fig. 6.15),

we find that the differences between them are significant when the social capital limit is 5. This implies that decreasing interactions can lead to larger differences in the shortest path length distributions. In addition, we find that the Euclidean Distances between the collaborative nodes and the inactive nodes are significantly lower than that between other evolving social networks. This is because both the inactive nodes and the collaborative nodes are adequately connected in social networks, as presented in the average shortest path lengths (Fig. 6.14 (a) and Fig. 6.14 (d)).

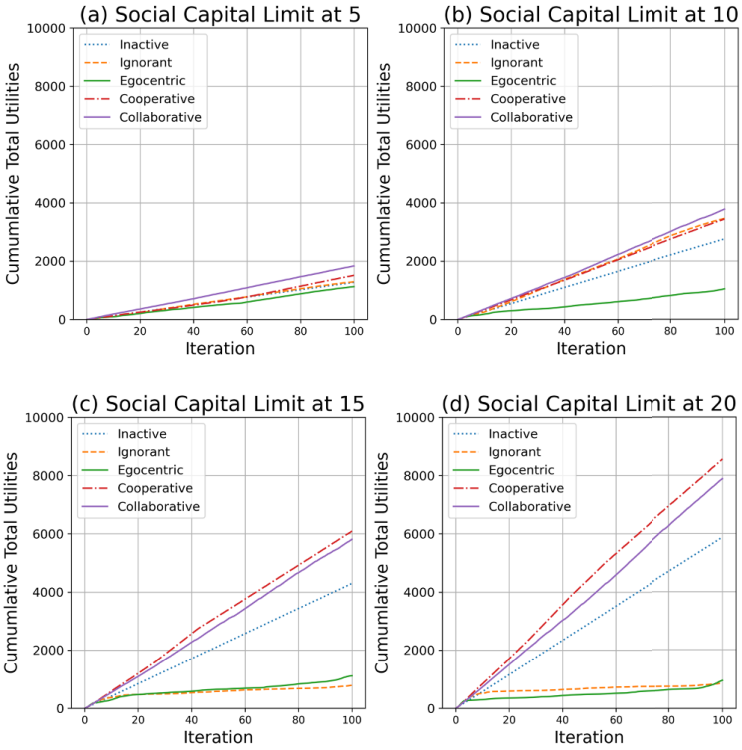
6.3.4 Interaction Utilities

We further investigate the interaction utilities of the nodes in the evolving social networks. Fig. 6.3.4 shows the interaction total utilities of nodes in evolving social networks and their cumulative values over time.

As shown in Fig. 6.16 (a), the interaction utilities for the inactive, cooperative and collaborative nodes generally increase in the first iterations and keep steady with minor fluctuations. This is because most nodes interact without exposure to the infections. In contrast, the interaction utilities of ignorant and egocentric nodes fluctuate greatly due to many infections (Fig. 6.1). Compared with the other node types and given any social capital limit, the social interaction utilities of cooperative nodes and collaborative nodes are generally higher.Fig. 6.16 (b) shows the cumulative interaction utilities of all nodes in the evolving social networks. We find that the egocentric nodes have the lowest cumulative utilities regardless the social capital limit. This is because the egocentric nodes make inefficient individual decisions without considering the others' behavioural changes. The collaborative and cooperative nodes can achieve the highest cumulative interaction utilities through group decisions. In addition, as shown in Fig. 6.16 (b), the collaborative nodes outperform the cooperative nodes under the social capital limit of 5 and 10 by more interactions and similar infection risks. In contrast, the cooperative nodes outperform the collaborative nodes under the social capital limit of 15 and 20 by fewer interactions and lower interaction risks. This indicates that the interaction risks and the infection occurrences greatly influence the utility maximisation process. A lower social capital limit makes it easier to avoid infection risks under the influence of random interference, ensuring the accuracy of group decisions on social capital transfer and group utility maximisation for collaborative nodes. In contrast, given a social capital limit at 10 or 15, the collaborative nodes have more flexible social capital transfer between the nodes and correspondingly develop more interactions, including the low-intensity ones which are significantly impacted by the random interference. The random interference



(a) Total interaction utility



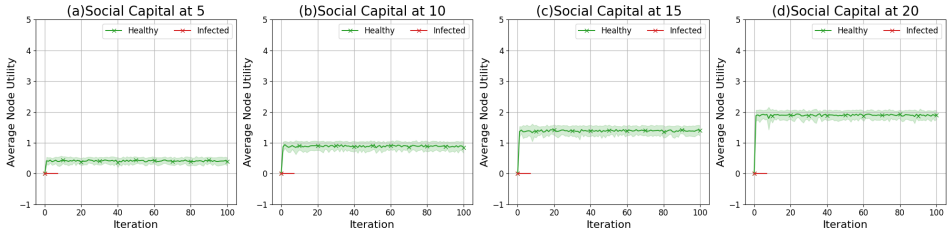
(b) Cumulative values of total interaction utility

Figure 6.16: The total interaction utilities of nodes in the evolving social networks (a) and their cumulative values over time (b).

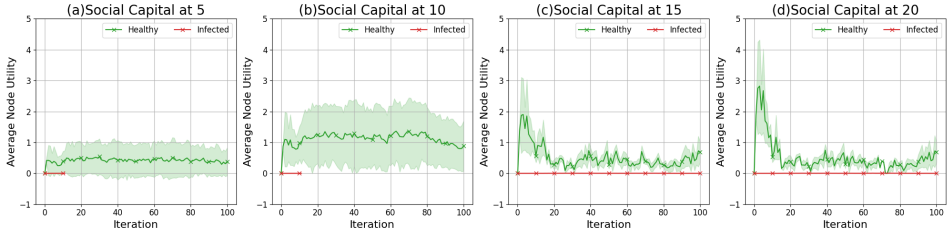
leads to inaccurate group decision making process and harms the efficiency of interaction utility maximisation.

We present the average values and standard deviations of node utilities for the infected and healthy nodes in Fig. 6.17. The average node utilities of healthy nodes are much higher than those of infected nodes. This is because the infected nodes tend to have higher interaction costs, and their interaction utilities are discounted due to their health statuses. It also indicates that, apart from the social capital limit, the utility discount (0.01 of the aggregated interaction utilities) and the interaction cost (0.60 for infected nodes and 0.40 for healthy nodes) can also suppress the interactions of infected nodes, and thus leading to the infection avoidance effect. Compared with the other node types, the average node utilities for the healthy ignorant nodes and the healthy egocentric nodes are much lower due to the interaction risks given a greater infection occurrence (Fig. 6.1). This also implies that the spread of the epidemic can decrease the utilities of both infected and healthy nodes.

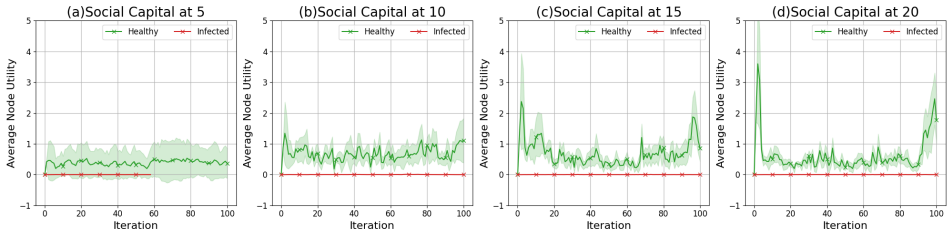
To summarise, first, regarding social capital, increasing social capital enables higher interactions but in the same time results in higher infection risks and uncertainty in decision-making. For example, the inactive, cooperative and collaborative nodes have more interactions when the social capital increases. The ignorant and egocentric nodes make more interactions when there are limited number of infections as the social capital limit increases from 5 to 10. In contrast, the ignorant and egocentric nodes reduce interactions when infection occurrence increases significantly under the social capital limit at 15 and 20. Second, among all the mutation styles, cooperative and collaborative mutation styles achieve higher total utilities than the other mutation styles. This implies that group decisions outperform individual decisions by avoiding the unawareness of the decisions of other nodes. Third, given an insignificant social capital limit that partially excludes the influence of random interference, the collaborative strategy outperforms the inactive, ignorant, egocentric and cooperative strategies by achieving higher interactions, lower infection occurrence and consequently, higher interaction utilities. This indicates that promoting collaborative interactions while imposing strict social capital limits in an epidemic outbreak can effectively reduce infection occurrence and preserve the overall benefit of people.



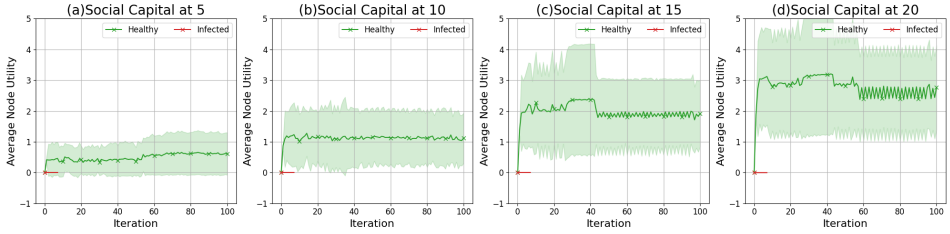
(a) Inactive



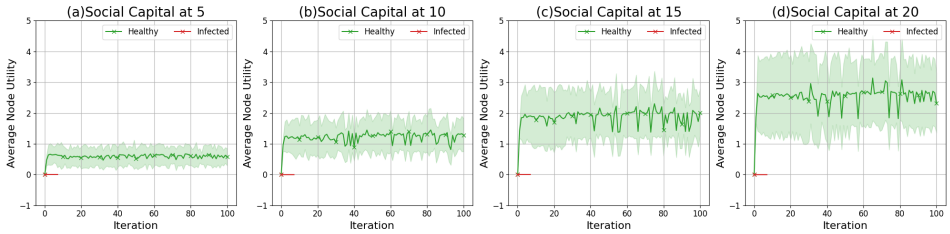
(b) Ignorant



(c) Egocentric



(d) Cooperative



(e) Collaborative

Figure 6.17: The average utility of infected nodes and healthy nodes in evolving social networks driven by different mutation styles under different social capital limits ranging within [5,10,15,20]. The mutation styles include the inactive (a), the ignorant (b), the egocentric (c), cooperative (d) and the collaborative (e) preference mutation styles.

6.4 Conclusion

This chapter extends our previous work on the DT-CNS modelling framework in chapter 4 and chapter 5 related to heterogeneous node features and connection preferences by introducing preference mutation mechanisms [3]. Our previous work enables to model CNSs composed of static networks and a dynamic process on the networks. This chapter progresses this framework as an extensive work by modelling the interrelated component changes in DT-CNS models.

More specifically, we propose an evolutionary generation 2 and 3 DT-CNS modelling framework concerning the nodes' adaptive decisions on preference mutation in response to the interaction patterns and epidemic risks. Under this framework, we consider nodes' heterogeneous features and changeable connection preferences, which combine the effects of preferential attachment, homophily and transitivity. We create heterogeneous preference mutation mechanisms to characterise nodes' adaptive decisions on preference mutation in response to the interaction patterns and epidemic risks. The preference mutation styles can be inactive, ignorant, egocentric, cooperative and collaborative dependent on the consideration of node utilities and the transferability of social capital between nodes in their decision-making processes.

To build the preference mutation mechanism while incorporating the feedback of interactions and infections, we introduce the social capital limit (which limits the number of possible relationships) to the nodes' preference mutation process and propose to model nodes' preference changes as an optimisation process for higher interaction utilities with limited social capital. Nodes can follow heterogeneous preference mutation styles, including the (i)inactive style that keeps zero social preferences, the (ii) ignorant style that randomly mutates preferences, the (iii) egocentric style that optimises individual interaction utility, (iv) cooperative style that optimises the total interaction utilities by group decisions and (v) collaborative style that further allows the cooperative nodes to transfer social capital.

We conduct simulation-based experiments on the evolutionary DT-CNS models. In the experiments, the nodes interact and mutate their preferences based on heterogeneous strategies, which lead to heterogeneous interaction and infection patterns. We find that (i) increasing social capital enables more interactions, higher node degrees, higher clustering coefficient, lower shortest path lengths, but higher infection risks and uncertainty in decision-making; (ii) group decisions outperform individual decisions by eliminating the unawareness of the decisions of other nodes; (iii) the collaborative strategy, given a

strict social capital limit, outperforms the inactive, ignorant, egocentric and cooperative strategies by achieving higher interactions, lower infection numbers and consequently, higher interaction utilities.

In summary, this chapter proposes an evolutionary DT-CNS modelling framework related to heterogeneous node features and the changeable preferences for connecting with others. This enables the answer to **RQ5** and contributes to the research topic on DT-CNS models from three perspectives:

- **RQ5**: How to build and evaluate the DT-CNS model evolution driven by mutation of feature preferences?

    1. **RC5.1**: New feature preference mutation mechanism considering heterogeneous mutation styles.
    2. **RC5.2**: Evolutionary DT-CNS models with evolving networks, dynamic processes and the interrelations between their dynamics.
    3. **RC5.3**: A case study on the evolutionary DT-CNS models in an epidemic outbreak based on different preference mutation styles.

The proposed modelling framework can model the interrelated CNS component changes by incorporating the impact of epidemic processes on the network evolution.

CHAPTER 7

DISCUSSION AND CONCLUSION

This chapter summarises the conclusions and contributions (see section 7.1) and provides an outlook on future research (see section 7.2).

7.1 Conclusion

This section summarises and concludes the research work conducted and presented in the previous chapters as well as discusses their respective contributions to this project in context of the research motivation and questions introduced in chapter 1.

To answer the **RQ1:** How do we compare modelling paradigms of Complex Networked Systems and measure their distances to a Digital Twin?, we conducted a comprehensive survey on CNS models in chapter 2 in which we presented the current state-of-the-art when it comes to modelling of CNSs, identify research gaps and challenges as well as discussed future research directions in DT-CNSs space. We contribute to this research topic by proposing four complexity dimensions for the network representation and five generations for the dynamics modelling (both of and on CNS, referring to Tab. 1.2). We also created unified assessment criteria to assess their respective capabilities of reaching the ultimate goal of building Digital Twin of a networked system. Using the proposed modelling framework, we assessed that, as shown in Fig. 7.2, generation 1 CNS models (dynamic process on static networks) and generation 2 CNS models (evolving dynamic process on static networks and dynamic process on evolving networks) have been relatively well developed in the current studies. However, few of the involved

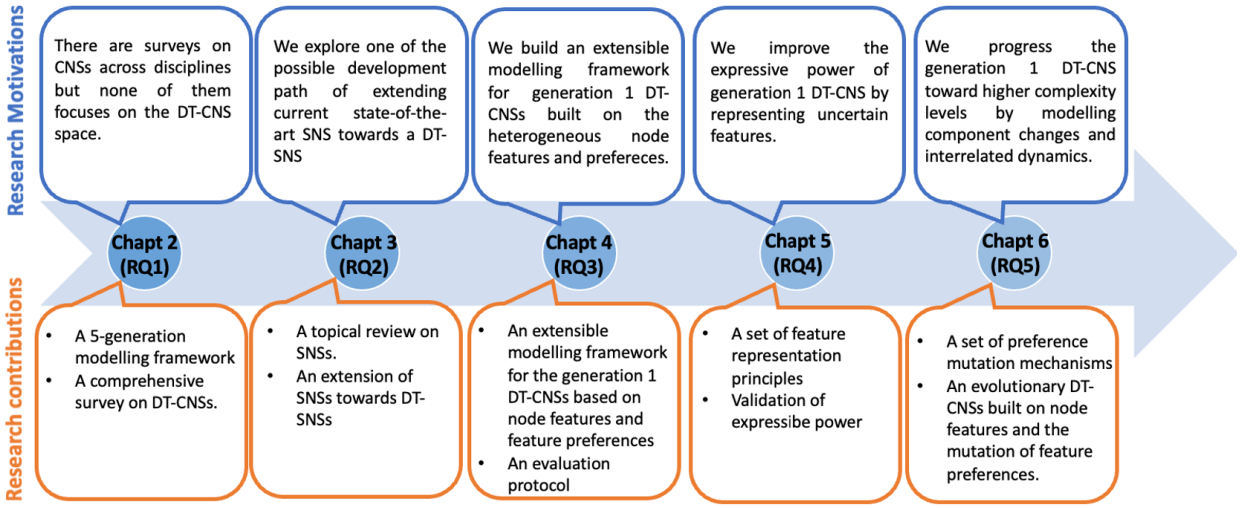


Figure 7.1: Discussion of the research motivations and the research contributions of each chapter

studies consider node attributes and the changes of preferences for node attributes [34]. The construction of DT-CNS models based on features and preference mutation remain a research gap to be further explored. Generation 3 CNS models (evolving dynamic processes on evolving networks with interrelations between them) have been partially developed with still many areas for further extensions concerning the network evolution driven by preference mutation in response to the dynamic process, transmissibility changes in process dimension referring to network patterns, as well as the co-evolution of network and process with interrelated dynamics. In contrast, generation 4 and 5 CNS models, characterised with temporal and interrelated dynamics, have not been sufficiently studied and progress in this space is very limited.

In order to explore one of the possible development pathways, we created Digital Twin-Oriented Social Network Simulator (DT-SNS) underpinned by the conceptual modelling framework proposed in chapter 2. First, we performed a topical review on the current state-of-the-art Social Network Simulators (SNSs) and conducted an experiment on developing the SNS towards DT-SNSs (chapter 3). Through the critical review of SNSs, we find that most existing SNSs focus on generating purely simulated networks. At the same time, only a small proportion of SNSs aim to model real nodes and node attributes with/without observable information about edges. Current SNSs generally focus on static networks or dynamic networks (belonging to the generation 1/2 DT-SNSs). Few SNSs consider jointly networks, the dynamic process on the network and their

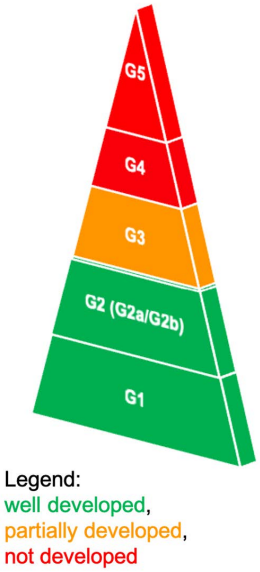


Figure 7.2: The research directions and research gaps.

interrelations (generation 3 DT-SNSs). The interrelated social networks and dynamic processes (generation 4 and 5 DT-SNSs), with the continuous acquisition of real-time information and feedback, also remain a research gap and require further study. In the experiments, we extended one of the most promising simulators from the evaluated ones [34] towards a Digital Twin Oriented SNS (DT Oriented SNS) by introducing heterogeneous features of nodes. We also proposed an approach to assess the similarity between real and simulated networks with the composite performance indexes based on global and local structural measures while taking the runtime of the simulator as an indicator of its efficiency. Using a well-known Karate Club network as an example, we found that the similarity levels between the target and simulated networks change with the introduced features and vary with the discussed network characteristics. Work presented in chapter 3 enabled us to answer **RQ2**: How do we build and evaluate DT-SNS that accommodates heterogeneous features of nodes? At the same time, this work revealed the research challenge of developing and evaluating DT-CNS model in future studies, which is to build an extensible modelling framework based on heterogeneous node features and the interaction rules related to feature preferences.

In chapter 4, we answered **RQ3**: How do we build and evaluate DT-CNS model considering the feature preferences? By that we addressed the research challenges related to CNS model extension and evaluation revealed in chapter 3. This is done by proposing an extensible DT-CNS modelling framework for generation 1 DT-CNS models, which is com-

posed of static networks and spreading processes on these networks without considering their interrelations. We introduced this modelling framework with novel elements that were not considered before: heterogeneous node features and interaction rules. These rules characterise nodes' preferences for connecting with others, covering two widely employed principles for network formation and growth: preferential attachment and homophily. In addition, we conducted experiments on simulation-based DT-CNS models that incorporate various age features and rules about network growth and different transmissibilities related to an epidemic spread on these networks. We presented a case study on disaster resilience of social networks given an epidemic outbreak by investigating the infection occurrence within specific time and social distance. The experiment results showed how different age features and interaction rules influence network growth and epidemic spread. They also indicated that the real-world scenarios can be even more complex and more realistic information related to node features and related rules, such as uncertain deductions on age, need to be included to simulate social networks in more faithful way. The analysis revealed that, to achieve maximum disaster resilience, mitigation policies should be targeted at nodes with preferred features as they have higher infection risks and can be the focus of epidemic control. In general, this study showed that the growing complexity of the modelling approach results in better performance of the DT-CNS models.

To build on the results obtained in chapter 4 and to introduce even more realistic node features to the DT-CNS models with improved expressive power, in chapter 5, we modelled uncertain features with heterogeneous feature representation principles and with that we answered **RQ4**: How do we represent uncertain features with DT-CNS models? We introduce fuzzy sets to describe the uncertain features and feature differences. In addition, to investigate the influence of fuzzy sets on expressive power of the proposed DT-CNS models based on uncertain features, we conducted sensitivity checks and model validation on the DT-CNS models by introducing a different number of fuzzy sets and optimising the respective parameters. In the experiments, we built generation 1 DT-CNS models to recreate realistic physical contact networks in different countries from real age and gender feature distributions based on various representation principles and an optimised feature preference. We also investigated their respective disaster resilience to an epidemic outbreak from the most popular node. The results suggest that the increased flexibility of feature representation with fuzzy sets improves expressive power. In addition, the heterogeneous features influence the network structures and induce the epidemic outbreak. The nodes with preferred features are more susceptible to the

epidemic and should be targeted with mitigation policies in an early stage. This further showed that increasing modelling complexity is beneficial for the performance of the model.

As the next step, in chapter 6, we progress our modelling framework towards higher complexity level (generation 2 and 3) where dynamic processes take place on evolving social networks. This framework extension and conducted experiments enabled us to answer **RQ5**: How do we build and evaluate the DT-CNS model evolution driven by mutation of feature preferences? We proposed an evolutionary DT-CNS model framework related to heterogeneous node features, changeable connection preferences and their interrelation with the epidemic risks. We also proposed heterogeneous preference mutation mechanisms to characterise nodes' adaptive decisions on preference mutation for higher interaction utilities related to interaction patterns and epidemic risks under the social capital constraint. The nodes' heterogeneous preference mutation styles include the (i) inactive style that keeps initial social preferences, (ii) ignorant style that randomly mutates preferences, (iii) egocentric style that optimises individual interaction utility, (iv) cooperative style that optimises the total interaction utilities by group decisions and (v) collaborative style that further allows the cooperative nodes to transfer social capital. The evolutionary DT-CNS models fall in generation 2 or 3, dependent on whether they consider the interrelations between networks and processes by including epidemic risks in the preference mutation. Our simulation experiments on evolutionary DT-CNS models revealed that heterogeneous preference mutation styles lead to various interaction and infection patterns. The collaborative nodes under a strict social capital limit can promote interactions, reduce infection risks and achieve higher overall interaction utilities. In general, this study showed that the increasing complexity in DT-CNS models enables to simulate different what-if scenarios with highly flexible experiment set-ups, mimicking and extending the reality.

In summary, this thesis contributes to the topic Digital Twin-Oriented Complex Networked Systems in the following aspects:

- A comprehensive literature review on Complex Networked Systems and the proposal of a conceptual modelling framework for DT-CNSs that includes four complexity dimensions and five generations of dynamics modelling (chapter 2).

- A critical review and evaluation of SNSs under the proposed conceptual modelling framework for DT-CNSs and an extension of one of the most promising SNSs towards a DT-SNS (chapter 3).

- The proposal of the novel DT-CNS modelling framework for static networks and dynamic process on these networks based on heterogeneous node features and interaction rules that characterise nodes' preferences for connecting with others (chapter 4).
- The introduction of fuzzy feature representation principles to the DT-CNS modelling framework with the improved expressive power of the uncertain features (chapter 5).
- The proposal of an evolutionary DT-CNS modelling framework based on heterogeneous features and the corresponding preference changes for the network evolution under the influence of dynamic processes (chapter 5).

Research work conducted in this thesis is not the end of the road for the Digital Twin oriented Complex Networked Systems and the ultimate goal of generation 5 will be the focus of the future work (section 7.2).

7.2 Future Work

This section outlines our future research path based on the current research output. Our future study will focus on the extension of this novel DT-CNS framework with continuous changes of dynamics of both network and process dimensions. We will also progress our work towards the ultimate goal of Digital Twin by introducing more realistic and complex dynamics and information to the DT-CNSs. The DT-CNSs modelling framework can also be employed to multidisciplinary application scenarios.

Regarding the extendability of the DT-CNS modelling framework, the DT-CNS models can be promisingly extended to higher complexity generations by introducing changes to features and related rules, including feature representation and feature preferences. We provide one of the possible development paths to extend the DT-CNSs towards the DT of a real-world networked system though introducing (co-)evolutionary and real-time updatable interaction rules (feature representation and preferences) and transmission rules (seed selection and transmissibility) [3]. This development path progresses over five DT-CNS model generations and gradually enables the evolvability in dynamics, interconnections between evolutionary dynamics and the interplay between the DT-CNSs and the reality.

Due to the flexibility of the DT-CNS modelling framework in introducing node features and the related preferences, this modelling framework can be further employed in

Table 7.1: Development path of DT-CNS model from generation 1 to generation 5 (the ultimate goal of a DT) with increasing complexity levels, concerning the interaction rules, transmission rules, and the real-world decision-making dynamics that are linked to the generation 5 DT-CNSs in a real-time feedback loop.

Generation	DT-CNSs		Reality
	Interaction Rule	Transmission Rule	Decision-Making
1	Non-Evolutionary	Non-Evolutionary	×
2	Non-Evolutionary	Evolutionary	×
	Evolutionary	Non-Evolutionary	×
3	Co-Evolutionary		×
4	Co-Evolutionary and Real-time Updatable		×
5	Co-Evolutionary and Real-time Updatable		

multiple disciplines by capturing various node characteristics and their connection preferences. The DT-CNS model can simulate the social networked systems, where nodes have social contact based on a set of interaction rules and, meanwhile, transmit the epidemic, opinion, influence or any other spread to neighbours. In addition, we would consider more realistic setups of the dynamic processes to recreate real-world scenarios. For example, in the DT-CNS models of a social network under an epidemic outbreak, we model the social network growth based on nodes' heterogeneous features (e.g. age, sex, height, popularity, etc.) and their interaction rules that characterise their preferences for nodes' features (e.g. preference for a similar age and a young age). Furthermore, the epidemic models can be introduced with more flexible setups of the transmission rules (e.g. the rules that govern the infection and the recovery processes) and seed selection strategies. The network might shrink in an epidemic outbreak as people are reluctant to contact each other, and this can be realised by introducing a flexible social capital limit to the DT-CNS framework. There are other possible application scenarios for the proposed DT-CNS modelling framework, such as the financial cooperation systems where financial risks can spread over the cooperators in a financial crisis. In a DT-CNS for a financial cooperation system, the nodes (firms) cooperate considering the firm characteristics (e.g. asset values, stock price, etc.) and spread financial risks through trade flows and cooperation.

As an extension of this thesis, we currently work on the generation 4 DT-CNS models that consider temporal changes of the interaction rules. In addition, we also introduce the DT-CNS with more realistic node features, such as language features, and employ large language models to equip nodes with the capability to analyse features and the tendency to make connections. The DT-CNS modelling framework has been and will

be under active development. In the future,w e will endeavour to further advance our research in this field and contribute to the ongoing evolution of the DT-CNS modelling framework.



APPENDIX

A.1 Extendable Social Network Simulators

In this appendix, we list some extendable social network simulators found through the review of literature and websites (http://caagt.ugent.be/CaGe/index.html; https://hog.grinvin.org/). The Tab. A.1 below includes the names, references and links to the code for these SNSs while briefly describing their functions.

Table A.1: SNSs in the current studies

SNS	Study	Link to the code	Description
VirtualSoc	[34]	https://github.com/AkandaAshraf/VirtualSoc	Dynamic Social Network Simulation Data with Ground Truth Labels and Features
Hashkat		docs.hashkat.org	A dynamic network simulor designed to model the growth of and information propagation through an online social network.
Minibaum	[390]	http://caagt.ugent.be/minibaum/	A generator for connected cubic graphs, but can be restricted to generating only graphs that have a fixed minimal girth or are bipartite
snarkhunter	[400]	http://caagt.ugent.be/cubic/	A generator for connected cubic graphs and snarks.
GenHypohamiltonian	[401]	http://caagt.ugent.be/hypoham/	A generator for hypohamiltonian and almost hypohamiltonian graphs
Genreg	[391]	http://www.mathe2.uni-bayreuth.de/markus/reggraphs.html	A very fast structure generator for regular graphs
MOLGEN	[392]	http://www.molgen.de/?src=documents/molgen4.html	A structure generator for molecular graphs.
alternating	[393]	https://github.com/nvcleemp/alternating	A generator for alternating planar graphs
CographGeneration	[394]	https://github.com/atilaajones/CographGeneration	A generator for cographs
CriticalPfreeGraph	[395]	http://caagt.ugent.be/criticalpfree/	A generator for k-critical graphs without long induced paths
nauty and Traces	[402]	https://pallini.di.uniroma1.it/	A generator for automorphism groups of graphs and digraphs. It can also produce a canonical labelling.
plantri and fullgen	[403]	https://users.cecs.anu.edu.au/ bdm/plantri/	A generator for planar graph.
Buckygen	[396]	http://caagt.ugent.be/buckygen/	A generator for all nonisomorphic fullerenes
perihamiltonian	[397]	https://github.com/nvcleemp/perihamiltonian	A generator for perihamiltonian graphs with a given connectivity.
GenerateUHG	[398]	http://caagt.ugent.be/uhg/	A generator for graphs with a given number k of hamiltonian cycles (which is especially efficient for small values of k)

A.2 Optimised Parameters for DT-SNSs

In this appendix, we list the best solution for zero feature–based SNS, real feature–based SNS, simulated feature–based SNS and hybrid feature–based SNS. Tab. A.2 below includes the names and values of the optimised parameters for SNSs.

Table A.2: Best solution for the optimisation of zero feature–based SNS considering similarity and efficiency.

Parameter	Meaning	Parameter Value			
		Zero feature SNS	Real feature SNS	Simulated feature SNS	Hybrid feature SNS
p_k	Preference vector	–	[-1]	[1]	[1,1]
w_k	Preference weighting vector	–	[0.81]	[0.17]	[0.40,0.68]
q_k	Weights of feature–based score	–	0.28	0.40	0.42
r_k	Weights of popularity–based score	0.36	0.54	0.71	0.33
m_k	Preference attachment parameter	0.04	0.06	0.25	0.46
c_k	Weights of shortest path-based score	0.30	0.83	0.88	0.80
γ^1	Weights of preference for path-length at 1	0.91	0.03	0.94	0.79
γ^2	Weights of preference for path-length at 2	0.09	0.12	0.12	0.14
γ^3	Weights of preference for path-length at 3	0.09	0.81	0.94	0.30

APPENDIX B

APPENDIX

B.1 Parameter set-ups for preferences and weights of preferences of the DT-CNSs in this study

In our experiments, we assume an encounter rate at 0.80 and a random interference that follows a random normal distribution $\mathbf{N}(0, 0.005^2)$. We optimise other parameters of the social network simulators, including preferences and the weights of preferences (See Table B.1).

Table B.1: The parameters of the DT-CNSs.

Parameters	p	w^p	h	w^h
$DT{-}CNS_U^{P+}$	1	1.00	1	0.00
$DT{-}CNS_U^{P-}$	-1	1.00	1	0.00
$DT{-}CNS_U^{H+}$	1	0.00	1	1.00
$DT{-}CNS_U^{H-}$	1	0.00	-1	1.00
$DT{-}CNS_U^{PH}$	-1	0.05	1	0.08
$DT{-}CNS_B^{P+}$	1	1.00	1	0.00
$DT{-}CNS_B^{P-}$	-1	1.00	1	0.00
$DT{-}CNS_B^{H+}$	1	0.00	1	1.00
$DT{-}CNS_B^{H-}$	1	0.00	-1	1.00
$DT{-}CNS_B^{PH}$	-1	0.03	1	0.06
$DT{-}CNS_I^{P+}$	1	1.00	1	0.00
$DT{-}CNS_I^{P-}$	-1	1.00	1	0.00
$DT{-}CNS_I^{H+}$	1	0.00	1	1.00
$DT{-}CNS_I^{H-}$	1	0.00	-1	1.00
$DT{-}CNS_I^{PH}$	1	0.68	-1	0.73
$DT{-}CNS_L^{P+}$	1	1.00	1	0.00
$DT{-}CNS_L^{P-}$	-1	1.00	1	0.00
$DT{-}CNS_L^{H+}$	1	0.00	1	1.00
$DT{-}CNS_L^{H-}$	1	0.00	-1	1.00
$DT{-}CNS_L^{PH}$	1	0.02	-1	0.08
$DT{-}CNS_R^{P+}$	1	1.00	1	0.00
$DT{-}CNS_R^{P-}$	-1	1.00	1	0.00
$DT{-}CNS_R^{H+}$	1	0.00	1	1.00
$DT{-}CNS_R^{H-}$	1	0.00	-1	1.00
$DT{-}CNS_R^{PH}$	1	0.02	-1	0.06

B.2 Distance between the network patterns of simulated networks and the target network

In our experiments, we employ JS divergence to evaluate the differences between distribution-based network patterns, including degree distributions, clustering coefficient distributions and shortest path length distributions. Tab. B.2 presents the JS divergence values between different network patterns.

Table B.2: The JS divergence between the network patterns of simulated networks and the target network.

Network Patterns	Degree Distribution	Clustering Coefficient Distribution	Shortest Path Length Distribution
$DT-CNS_U^{P+}$	0.51	0.63	0.29
$DT-CNS_U^{P-}$	0.53	0.68	0.28
$DT-CNS_U^{H+}$	0.36	0.68	0.12
$DT-CNS_U^{H-}$	0.43	0.69	0.35
$DT-CNS_U^{PH}$	0.26	0.60	0.00
$DT-CNS_B^{P+}$	0.50	0.63	0.09
$DT-CNS_B^{P-}$	0.50	0.66	0.16
$DT-CNS_B^{H+}$	0.45	0.55	0.01
$DT-CNS_B^{H-}$	0.50	0.69	0.34
$DT-CNS_B^{PH}$	0.27	0.41	0.01
$DT-CNS_I^{P+}$	0.56	0.68	, 0.31
$DT-CNS_I^{P-}$	0.54	0.68	0.30
$DT-CNS_I^{H+}$	0.46	0.81	0.22
$DT-CNS_I^{H-}$	0.42	0.67	0.41
$DT-CNS_I^{PH}$	0.33	0.70	0.06
$DT-CNS_L^{P+}$	0.56	0.70	0.37
$DT-CNS_L^{P-}$	0.53	0.68	0.01
$DT-CNS_L^{H+}$	0.49	0.53	0.01
$DT-CNS_L^{H-}$	0.53	0.68	0.37
$DT-CNS_L^{PH}$	0.28	0.45	0.01
$DT-CNS_R^{P+}$	0.51	0.69	0.01
$DT-CNS_R^{P-}$	0.59	0.72	0.40
$DT-CNS_R^{H+}$	0.47	0.58	0.01
$DT-CNS_R^{H-}$	0.55	0.68	0.39
$DT-CNS_R^{PH}$	0.28	0.50	0.01

APPENDIX C

APPENDIX

C.1 Parameter bounds' set-ups in the sensitivity analysis and optimisation process

Table C.1: Parameter bounds' set-ups in the sensitivity analysis and optimisation process

Parameter	Belgium	Finland	Germany	Italy	Luxembourg	Poland
μ_1^h	[-45,45]	[-9,9]	[-9,9]	[-9,-9]	[-9,-9]	[-5.6250,5.6250]
μ_2^h	–	[9,27]	[9,27]	[9,27]	[9,27]	[5.6250,16.8750]
μ_3^h	–	[27,45]	[27,45]	[27,45]	[27,45]	[16.8750,28.1250]
μ_4^h	–	[45,63]	[45,63]	[45,63]	[45,63]	[20.1250,39.3750]
μ_5^h	–	[63,81]	[63,81]	[63,81]	[63,81]	[39.3750,50.6250]
μ_6^h	–	[81,99]	[81,99]	[81,99]	[81,99]	[50.6250,61.8750]
μ_7^h	–	–	–	–	–	[61.8750,73.1250]
μ_8^h	–	–	–	–	–	[73.1250,84.3750]
μ_9^h	–	–	–	–	–	[84.3750,95.6250]
μ_1^h	[-15,15]	[-15,15]	[-15,15]	[-22.5,22.5]	[-45,45]	[-15,15]
μ_2^h	[15,45]	[15,45]	[15,45]	[22.5,67.5]	[45,135]	[15,45]
μ_3^h	[45,75]	[45,75]	[45,75]	[67.5,102.5]	–	[45,75]
μ_4^h	[75,105]	[75,105]	[75,105]	–	–	[75,105]
σ_i^h	[1,9]	[1,9]	[1,9]	[1,9]	[1,9]	[1,9]
σ_i^p	[1,9]	[1,9]	[1,9]	[1,9]	[1,9]	[1,9]

APPENDIX D

APPENDIX

D.1 Networks and their statistics in the first 20 simulation steps

This section presents the dynamic social networks generated over twenty iterations considering different social DNA mutation styles.

D.1.1 Social Capital Limit at 5

D.1.1.1 Inactive

Table D.1: Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	57	3.8	1.3	5	0	0.09	0.1	0.33	0.00	29	4.34	6.92	30	1
2	1	29	55	3.67	1.27	5	0	0.08	0.1	0.33	0.00	29	4.45	6.89	30	1
3	1	29	60	4.0	1.1	5	0	0.15	0.14	0.5	0.00	29	4.32	6.92	30	1
4	1	29	53	3.53	1.2	5	0	0.05	0.08	0.33	0.00	29	4.44	6.89	30	1
5	1	29	56	3.73	1.24	5	0	0.06	0.09	0.33	0.00	29	4.3	6.92	30	1
6	1	29	60	4.0	1.24	5	0	0.07	0.08	0.2	0.00	29	4.31	6.92	30	1
7	2	28	54	3.6	1.43	5	0	0.15	0.19	0.67	0.00	57	6.14	9.3	30	1
8	1	29	56	3.73	1.46	5	0	0.11	0.13	0.5	0.00	29	4.46	6.9	30	1
9	0	30	60	4.0	1.39	5	1	0.15	0.25	1.00	0.00	0	2.55	0.94	5	1
10	0	30	63	4.2	1.01	5	2	0.13	0.15	0.67	0.00	0	2.47	0.88	5	1
11	0	30	60	4.0	1.21	5	1	0.06	0.09	0.33	0.00	0	2.48	0.88	5	1
12	0	30	59	3.93	1.15	5	1	0.1	0.14	0.67	0.00	0	2.66	1.1	6	1
13	1	29	56	3.73	1.26	5	0	0.07	0.1	0.33	0.00	29	4.34	6.91	30	1
14	0	30	60	4.0	1.21	5	1	0.1	0.11	0.33	0.00	0	2.44	0.85	5	1
15	0	30	58	3.87	1.12	5	1	0.01	0.04	0.17	0.00	0	2.55	0.94	5	1
16	0	30	63	4.2	0.7	5	3	0.11	0.16	0.67	0.00	0	2.43	0.83	4	1
17	0	30	57	3.8	1.28	5	1	0.05	0.09	0.33	0.00	0	2.54	0.92	5	1
18	0	30	61	4.07	0.85	5	2	0.07	0.1	0.33	0.00	0	2.47	0.86	5	1
19	0	30	58	3.87	0.96	5	2	0.07	0.1	0.33	0.00	0	2.47	0.83	4	1
20	0	30	54	3.6	1.08	5	1	0.02	0.05	0.17	0.00	0	2.69	1.00	6	1

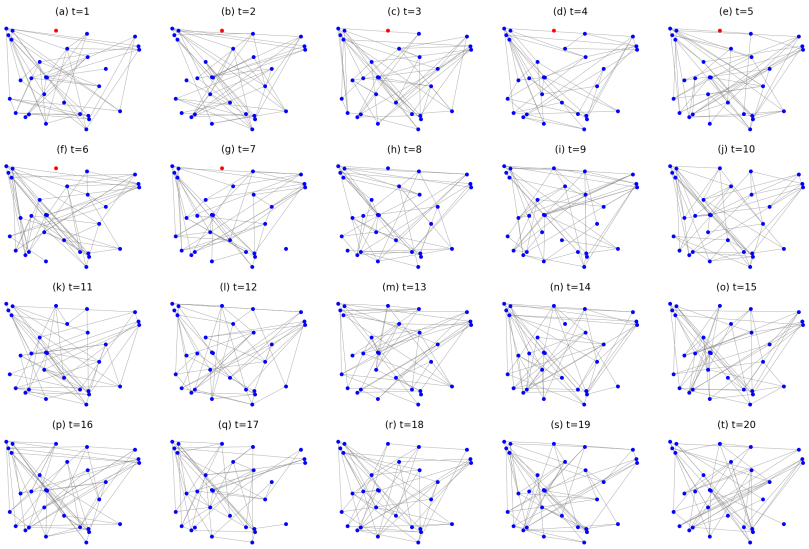


Figure D.1: The evolving social networks driven by inactive nodes in an epidemic outbreak.

D.1.1.2 Ignorant

Table D.2: Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	57	3.8	1.3	5	0	0.09	0.1	0.33	0.00	29	4.34	6.92	30	1
2	12	18	25	1.67	1.87	5	0	0.03	0.08	0.3	0.00	314	22.37	12.31	30	1
3	13	17	24	1.6	1.8	5	0	0.14	0.27	1.00	0.00	359	25.14	10.58	30	1
4	10	20	29	1.93	1.75	5	0	0.18	0.28	1.00	0.00	346	24.34	11.18	30	1
5	14	16	24	1.6	1.96	5	0	0.11	0.24	1.00	0.00	343	24.13	11.34	30	1
6	12	18	31	2.07	1.95	5	0	0.11	0.28	1.00	0.00	282	20.4	13.06	30	1
7	11	19	27	1.8	1.85	5	0	0.1	0.22	1.00	0.00	340	23.98	11.41	30	1
8	14	16	28	1.87	1.96	5	0	0.14	0.27	1.00	0.00	315	22.37	12.37	30	1
9	11	19	25	1.67	1.74	5	0	0.06	0.15	0.67	0.00	264	19.35	13.25	30	1
10	8	22	34	2.27	1.82	5	0	0.09	0.22	1.00	0.00	244	18.29	13.29	30	1
11	6	24	33	2.2	1.81	5	0	0.00	0.00	0	0	275	19.97	13.18	30	1
12	11	19	33	2.2	2.12	5	0	0.09	0.26	1.00	0.00	264	19.35	13.26	30	1
13	13	17	29	1.93	2.06	5	0	0.09	0.16	0.67	0.00	299	21.41	12.75	30	1
14	14	16	25	1.67	1.96	5	0	0.1	0.23	1.00	0.00	315	22.4	12.32	30	1
15	16	14	26	1.73	2.06	5	0	0.13	0.26	1.00	0.00	344	24.21	11.26	30	1
16	13	17	24	1.6	1.84	5	0	0.06	0.13	0.5	0.00	329	23.26	11.89	30	1
17	13	17	25	1.67	1.89	5	0	0.07	0.2	1.00	0.00	355	24.86	10.84	30	1
18	15	15	29	1.93	2.14	5	0	0.12	0.24	1.00	0.00	330	23.28	11.92	30	1
19	15	15	27	1.8	2.02	5	0	0.1	0.26	1.00	0.00	330	23.29	11.9	30	1
20	12	18	30	2.0	1.91	5	0	0.11	0.21	1.00	0.00	282	20.34	13.14	30	1

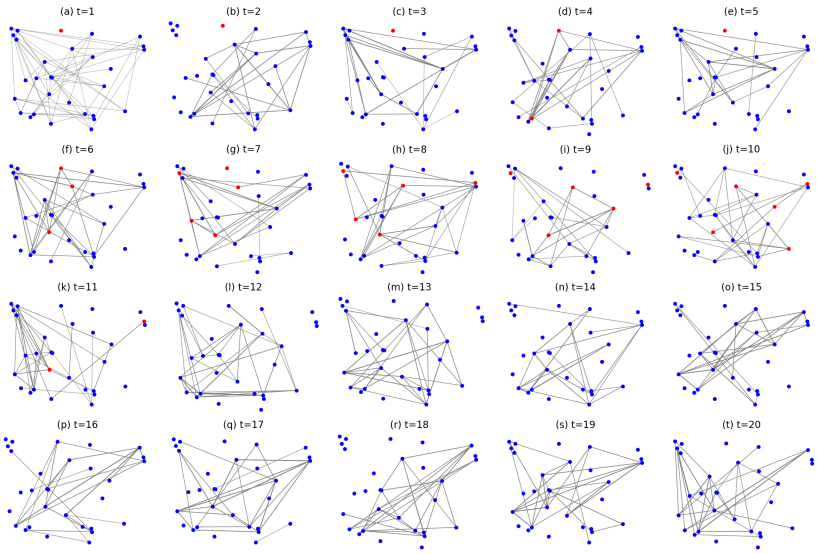


Figure D.2: The evolving social networks driven by ignorant nodes in an epidemic outbreak.

D.1.1.3 Egocentric

Table D.3: Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	57	3.8	1.3	5	0	0.09	0.1	0.33	0.00	29	4.34	6.92	30	1
2	16	14	23	1.53	2.03	5	0	0.06	0.14	0.5	0.00	368	25.67	10.15	30	1
3	16	14	22	1.47	2.0	5	0	0.13	0.26	1.00	0.00	368	25.72	10.04	30	1
4	16	14	22	1.47	1.96	5	0	0.15	0.26	0.7	0.00	368	25.71	10.07	30	1
5	16	14	21	1.4	1.84	5	0	0.13	0.25	1.00	0.00	389	27.0	8.71	30	1
6	18	12	20	1.33	1.89	5	0	0.11	0.23	0.83	0.00	369	25.83	9.87	30	1
7	15	15	27	1.8	2.02	5	0	0.1	0.18	0.5	0.00	330	23.33	11.83	30	1
8	13	17	28	1.87	1.94	5	0	0.15	0.32	1.00	0.00	299	21.52	12.6	30	1
9	12	18	35	2.33	2.13	5	0	0.19	0.31	1.00	0.00	282	20.28	13.21	30	1
10	11	19	36	2.4	2.2	5	0	0.19	0.31	1.00	0.00	264	19.17	13.47	30	1
11	12	18	34	2.27	2.05	5	0	0.08	0.2	1.00	0.00	282	20.37	13.1	30	1
12	9	21	29	1.93	1.75	5	0	0.08	0.2	1.00	0.00	263	19.24	13.33	30	1
13	9	21	29	1.93	1.75	5	0	0.06	0.19	1.00	0.00	225	16.94	13.54	30	1
14	12	18	30	2.0	2.0	5	0	0.18	0.25	1.00	0.00	282	20.46	12.98	30	1
15	15	15	24	1.6	1.93	5	0	0.05	0.12	0.5	0.00	330	23.35	11.8	30	1
16	16	14	24	1.6	1.96	5	0	0.14	0.23	1.00	0.00	344	24.2	11.28	30	1
17	16	14	20	1.33	1.8	5	0	0.12	0.22	0.67	0.00	368	25.74	9.99	30	1
18	14	16	25	1.67	1.96	5	0	0.11	0.2	0.6	0.00	315	22.46	12.23	30	1
19	13	17	26	1.73	1.86	5	0	0.21	0.32	1.00	0.00	299	21.64	12.44	30	1
20	13	17	26	1.73	1.9	5	0	0.14	0.28	0.8	0.00	299	21.63	12.45	30	1

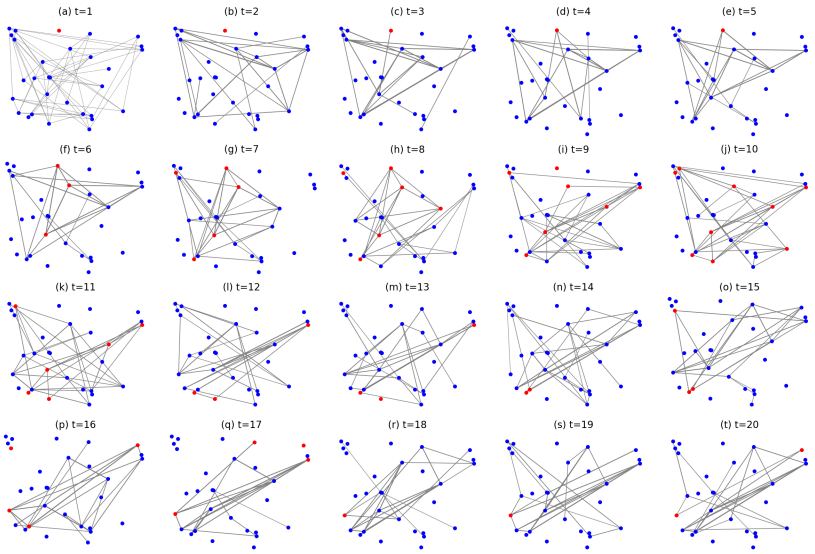


Figure D.3: The evolving social networks driven by egocentric nodes in an epidemic outbreak.

D.1.1.4 Cooperative

Table D.4: Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	57	3.8	1.3	5	0	0.09	0.1	0.33	0.00	29	4.34	6.92	30	1
2	14	16	23	1.53	1.94	5	0	0.06	0.15	0.67	0.00	343	24.14	11.33	30	1
3	14	16	23	1.53	1.86	5	0	0.04	0.13	0.67	0.00	354	24.82	10.83	30	1
4	18	12	22	1.47	2.0	5	0	0.07	0.15	0.5	0.00	369	25.74	10.07	30	1
5	15	15	24	1.6	1.91	5	0	0.02	0.07	0.33	0.00	330	23.33	11.84	30	1
6	15	15	25	1.67	2.01	5	0	0.04	0.09	0.33	0.00	356	24.94	10.75	30	1
7	14	16	27	1.8	2.06	5	0	0.08	0.17	0.67	0.00	315	22.37	12.38	30	1
8	16	14	22	1.47	2.0	5	0	0.08	0.17	0.6	0.00	368	25.69	10.1	30	1
9	17	13	21	1.4	1.89	5	0	0.09	0.22	1.00	0.00	357	25.03	10.65	30	1
10	16	14	22	1.47	1.91	5	0	0.05	0.1	0.33	0.00	368	25.69	10.11	30	1
11	10	20	34	2.27	1.86	5	0	0.2	0.34	1.00	0.00	336	23.62	11.77	30	1
12	9	21	32	2.13	1.84	5	0	0.08	0.2	1.00	0.00	323	22.83	12.18	30	1
13	7	23	33	2.2	1.78	5	0	0.12	0.23	1.00	0.00	272	19.71	13.31	30	1
14	10	20	28	1.87	1.82	5	0	0.12	0.25	1.00	0.00	320	22.74	12.12	30	1
15	13	17	24	1.6	1.8	5	0	0.17	0.27	1.00	0.00	329	23.32	11.79	30	1
16	15	15	21	1.4	1.76	5	0	0.06	0.15	0.67	0.00	330	23.38	11.75	30	1
17	16	14	22	1.47	1.82	5	0	0.04	0.18	1.00	0.00	368	25.68	10.12	30	1
18	18	12	21	1.4	1.99	5	0	0.13	0.25	1.00	0.00	369	25.76	10.04	30	1
19	14	16	26	1.73	1.98	5	0	0.1	0.2	0.6	0.00	354	24.8	10.88	30	1
20	12	18	28	1.87	1.91	5	0	0.02	0.06	0.2	0.00	282	20.37	13.09	30	1

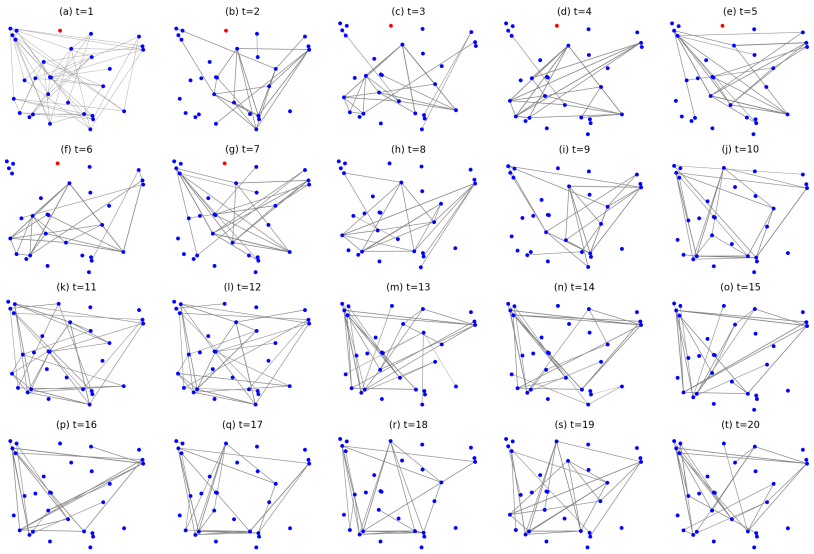


Figure D.4: The evolving social networks driven by cooperative nodes in an epidemic outbreak.

D.1.1.5 Collaborative

Table D.5: Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	123	8.2	3.96	21	0	0.41	0.17	1.00	0.00	29	3.63	7.07	30	1
2	1	29	124	8.27	4.57	27	0	0.44	0.16	1.00	0.00	29	3.58	7.07	30	1
3	1	29	129	8.6	4.74	27	0	0.48	0.15	0.8	0.00	29	3.57	7.08	30	1
4	1	29	129	8.6	4.85	27	0	0.5	0.18	1.00	0.00	29	3.57	7.08	30	1
5	1	29	120	8.0	4.6	27	0	0.48	0.19	1.00	0.00	29	3.61	7.07	30	1
6	1	29	124	8.27	4.71	26	0	0.44	0.17	0.7	0.00	29	3.61	7.07	30	1
7	1	29	124	8.27	5.54	26	0	0.55	0.22	1.00	0.00	29	3.63	7.07	30	1
8	1	29	122	8.13	4.65	26	0	0.46	0.16	0.7	0.00	29	3.6	7.07	30	1
9	0	30	130	8.67	2.52	14	5	0.3	0.09	0.5	0.1	0	1.74	0.52	3	1
10	0	30	128	8.53	2.83	16	4	0.28	0.09	0.47	0.1	0	1.74	0.51	3	1
11	0	30	125	8.33	2.87	13	2	0.3	0.13	0.5	0.00	0	1.82	0.6	3	1
12	0	30	118	7.87	2.73	13	3	0.27	0.12	0.6	0.00	0	1.85	0.6	3	1
13	0	30	114	7.6	2.8	12	1	0.25	0.12	0.43	0.00	0	1.9	0.65	4	1
14	0	30	133	8.87	2.51	14	2	0.31	0.09	0.5	0.00	0	1.75	0.55	3	1
15	0	30	125	8.33	1.94	13	4	0.27	0.07	0.43	0.14	0	1.76	0.52	3	1
16	0	30	130	8.67	1.85	12	5	0.29	0.08	0.5	0.14	0	1.74	0.52	3	1
17	0	30	126	8.4	2.2	13	3	0.29	0.07	0.5	0.19	0	1.76	0.53	3	1
18	0	30	132	8.8	1.38	11	6	0.23	0.07	0.39	0.00	0	1.72	0.5	3	1
19	0	30	122	8.13	2.08	13	4	0.28	0.09	0.48	0.1	0	1.79	0.55	3	1
20	0	30	116	7.73	2.69	14	4	0.31	0.15	0.8	0.14	0	1.81	0.55	3	1

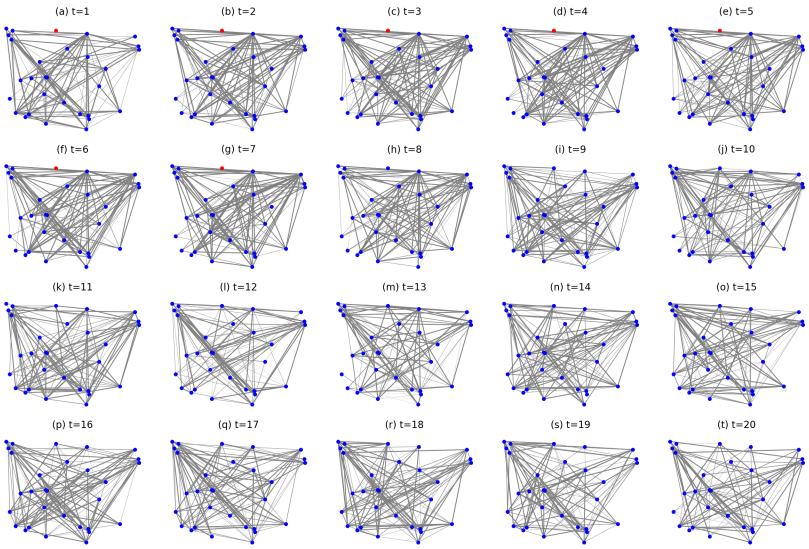


Figure D.5: The evolving social networks driven by collaborative nodes in an epidemic outbreak.

D.1.2 Social Capital Limit at 10

D.1.2.1 Inactive

Table D.6: Topological information of the network simulations driven by inactive mutation style under a social capital limit at 10.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	121	8.07	2.1	10	0	0.26	0.08	0.4	0.00	29	3.63	7.07	30	1
2	1	29	131	8.73	1.93	10	0	0.27	0.08	0.46	0.00	29	3.57	7.08	30	1
3	1	29	124	8.27	2.1	10	0	0.26	0.08	0.38	0.00	29	3.6	7.07	30	1
4	1	29	119	7.93	2.13	10	0	0.24	0.08	0.36	0.00	29	3.64	7.06	30	1
5	1	29	122	8.13	2.28	10	0	0.27	0.1	0.42	0.00	29	3.64	7.06	30	1
6	1	29	127	8.47	1.93	10	0	0.29	0.08	0.42	0.00	29	3.6	7.07	30	1
7	1	29	122	8.13	2.11	10	0	0.24	0.06	0.33	0.00	29	3.62	7.07	30	1
8	1	29	123	8.2	2.18	10	0	0.29	0.1	0.67	0.00	29	3.64	7.07	30	1
9	0	30	128	8.53	1.59	10	5	0.28	0.05	0.4	0.16	0	1.76	0.53	3	1
10	0	30	131	8.73	1.29	10	5	0.27	0.06	0.39	0.14	0	1.75	0.54	3	1
11	0	30	129	8.6	1.5	10	5	0.26	0.07	0.4	0.1	0	1.73	0.51	3	1
12	0	30	124	8.27	1.65	10	4	0.24	0.05	0.33	0.17	0	1.74	0.5	3	1
13	0	30	128	8.53	1.45	10	5	0.27	0.07	0.4	0.11	0	1.75	0.52	3	1
14	0	30	126	8.4	1.82	10	3	0.3	0.1	0.67	0.13	0	1.77	0.54	3	1
15	0	30	126	8.4	1.38	10	5	0.22	0.07	0.36	0.05	0	1.76	0.53	3	1
16	0	30	130	8.67	1.37	10	5	0.26	0.07	0.43	0.13	0	1.74	0.51	3	1
17	0	30	131	8.73	1.44	10	6	0.27	0.05	0.38	0.14	0	1.72	0.5	3	1
18	0	30	130	8.67	1.51	10	4	0.29	0.06	0.42	0.17	0	1.73	0.5	3	1
19	0	30	131	8.73	1.44	10	4	0.27	0.06	0.36	0.13	0	1.74	0.52	3	1
20	0	30	130	8.67	1.4	10	5	0.24	0.08	0.38	0.07	0	1.72	0.49	3	1

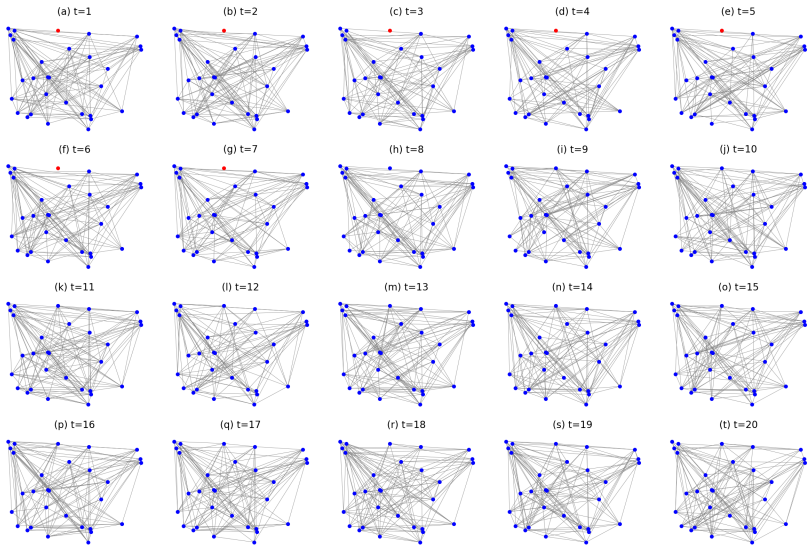


Figure D.6: The evolving social networks driven by inactive nodes in an epidemic outbreak.

D.1.2.2 Ignorant

Table D.7: Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 10.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	121	8.07	2.1	10	0	0.26	0.08	0.4	0.00	29	3.63	7.07	30	1
2	5	25	74	4.93	3.71	10	0	0.18	0.23	0.67	0.00	135	10.95	12.82	30	1
3	3	27	75	5.0	3.46	10	0	0.41	0.33	1.00	0.00	84	7.94	10.86	30	1
4	1	29	85	5.67	3.13	10	0	0.29	0.33	1.00	0.00	29	4.1	6.97	30	1
5	4	26	78	5.2	3.55	10	0	0.24	0.3	1.00	0.00	179	13.47	13.83	30	1
6	4	26	87	5.8	3.71	10	0	0.21	0.23	1.00	0.00	110	9.14	12.16	30	1
7	4	26	76	5.07	3.69	10	0	0.32	0.26	0.7	0.00	110	9.21	12.12	30	1
8	2	28	87	5.8	2.93	10	0	0.35	0.23	1.00	0.00	57	5.77	9.44	30	1
9	5	25	78	5.2	3.73	10	0	0.31	0.27	1.00	0.00	135	10.77	12.92	30	1
10	6	24	86	5.73	3.82	10	0	0.19	0.17	0.47	0.00	159	12.16	13.55	30	1
11	5	25	83	5.53	3.86	10	0	0.16	0.14	0.4	0.00	135	10.7	12.96	30	1
12	2	28	91	6.07	3.17	10	0	0.26	0.29	1.00	0.00	57	5.71	9.46	30	1
13	2	28	85	5.67	3.33	10	0	0.35	0.33	1.00	0.00	57	6.03	9.36	30	1
14	5	25	83	5.53	3.65	10	0	0.24	0.25	1.00	0.00	135	10.77	12.93	30	1
15	5	25	78	5.2	3.77	10	0	0.27	0.23	0.67	0.00	135	10.78	12.91	30	1
16	5	25	80	5.33	3.79	10	0	0.22	0.25	1.00	0.00	135	10.74	12.94	30	1
17	7	23	69	4.6	4.14	10	0	0.28	0.29	1.00	0.00	278	19.78	13.6	30	1
18	8	22	79	5.27	4.07	10	0	0.24	0.26	1.00	0.00	204	15.04	14.07	30	1
19	4	26	83	5.53	3.73	10	0	0.27	0.31	1.00	0.00	158	12.19	13.47	30	1
20	5	25	83	5.53	3.69	10	0	0.2	0.22	0.7	0.00	135	10.75	12.93	30	1

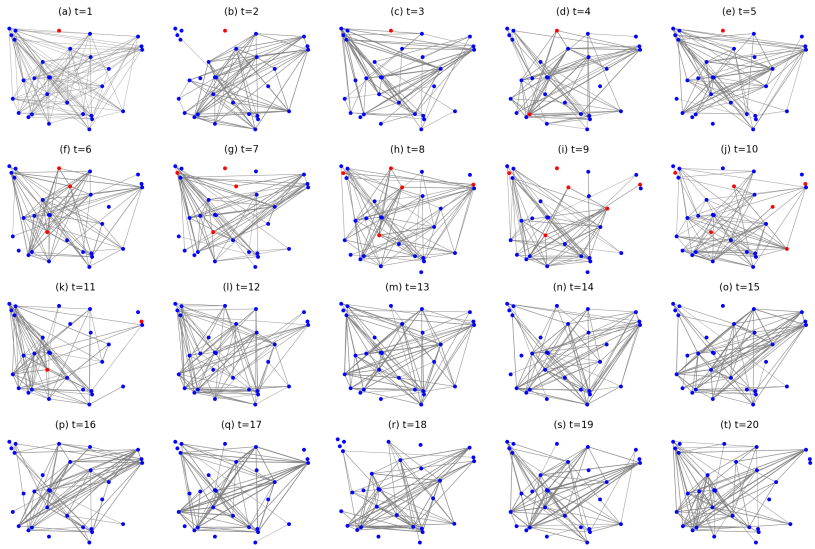


Figure D.7: The evolving social networks driven by ignorant nodes in an epidemic outbreak.

D.1.2.3 Egocentric

Table D.8: Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 10.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	121	8.07	2.1	10	0	0.26	0.08	0.4	0.00	29	3.63	7.07	30	1
2	7	23	77	5.13	3.7	10	0	0.17	0.16	0.51	0.00	182	13.67	13.86	30	1
3	3	27	73	4.87	3.59	10	0	0.34	0.29	1.00	0.00	134	10.81	12.83	30	1
4	2	28	79	5.27	3.43	10	0	0.31	0.32	1.00	0.00	57	6.13	9.34	30	1
5	3	27	94	6.27	3.41	10	0	0.22	0.25	0.8	0.00	84	7.58	11.01	30	1
6	1	29	94	6.27	3.38	10	0	0.21	0.3	1.00	0.00	29	4.08	6.98	30	1
7	2	28	97	6.47	3.15	10	0	0.19	0.27	1.00	0.00	57	5.74	9.45	30	1
8	2	28	104	6.93	3.17	10	0	0.46	0.26	0.93	0.00	57	5.72	9.46	30	1
9	1	29	87	5.8	3.61	10	0	0.5	0.33	1.00	0.00	83	7.69	10.88	30	1
10	1	29	100	6.67	2.74	10	0	0.42	0.27	1.00	0.00	29	4.03	6.99	30	1
11	1	29	106	7.07	2.9	10	0	0.39	0.21	0.8	0.00	29	4.0	7.0	30	1
12	1	29	100	6.67	3.08	10	0	0.41	0.26	0.9	0.00	29	4.07	6.99	30	1
13	0	30	103	6.87	2.42	10	2	0.25	0.25	1.00	0.00	0	2.11	0.81	5	1
14	4	26	92	6.13	3.39	10	0	0.25	0.25	1.00	0.00	110	9.09	12.19	30	1
15	1	29	98	6.53	3.24	10	0	0.19	0.24	1.00	0.00	29	4.05	6.99	30	1
16	0	30	96	6.4	2.92	10	3	0.33	0.22	0.69	0.00	0	2.27	0.93	4	1
17	3	27	102	6.8	2.9	10	0	0.41	0.23	0.73	0.00	84	7.49	11.04	30	1
18	2	28	107	7.13	2.96	10	0	0.44	0.32	1.00	0.00	57	5.77	9.44	30	1
19	0	30	111	7.4	2.64	10	1	0.49	0.23	1.00	0.00	0	2.17	0.89	4	1
20	2	28	95	6.33	3.14	10	0	0.41	0.24	0.9	0.00	57	5.86	9.42	30	1

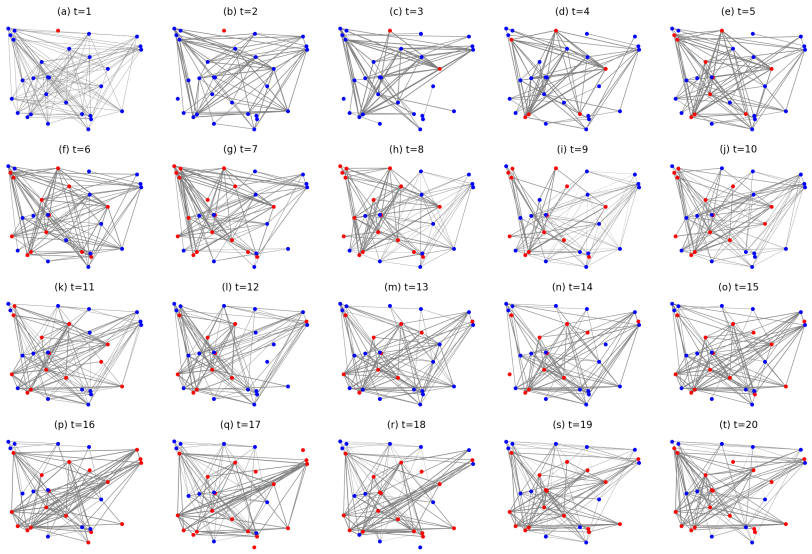


Figure D.8: The evolving social networks driven by egocentric nodes in an epidemic outbreak.

D.1.2.4 Cooperative

Table D.9: Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 10.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	121	8.07	2.1	10	0	0.26	0.08	0.4	0.00	29	3.63	7.07	30	1
2	8	22	74	4.93	3.8	10	0	0.28	0.27	1.00	0.00	204	15.09	14.02	30	1
3	3	27	85	5.67	3.51	10	0	0.31	0.28	1.00	0.00	84	7.6	10.99	30	1
4	4	26	77	5.13	3.48	10	0	0.29	0.28	1.00	0.00	110	9.26	12.1	30	1
5	4	26	80	5.33	3.59	10	0	0.24	0.28	0.79	0.00	110	9.23	12.11	30	1
6	4	26	76	5.07	3.62	10	0	0.27	0.31	1.00	0.00	110	9.3	12.07	30	1
7	4	26	77	5.13	3.69	10	0	0.29	0.28	0.78	0.00	110	9.32	12.06	30	1
8	7	23	79	5.27	4.04	10	0	0.29	0.25	0.8	0.00	224	16.31	14.11	30	1
9	11	19	70	4.67	4.38	10	0	0.27	0.29	0.79	0.00	298	21.04	13.22	30	1
10	9	21	73	4.87	4.19	10	0	0.27	0.3	0.87	0.00	225	16.43	14.06	30	1
11	4	26	77	5.13	3.53	10	0	0.31	0.26	1.00	0.00	110	9.17	12.14	30	1
12	7	23	76	5.07	4.17	10	0	0.25	0.25	0.71	0.00	242	17.47	14.04	30	1
13	3	27	84	5.6	3.51	10	0	0.43	0.3	1.00	0.00	84	7.58	11.00	30	1
14	4	26	94	6.27	3.3	10	0	0.22	0.22	1.00	0.00	110	8.99	12.24	30	1
15	4	26	78	5.2	3.57	10	0	0.34	0.27	1.00	0.00	158	12.25	13.43	30	1
16	9	21	70	4.67	4.16	10	0	0.29	0.29	1.00	0.00	225	16.5	13.99	30	1
17	5	25	83	5.53	3.47	10	0	0.32	0.27	1.00	0.00	135	10.69	12.97	30	1
18	9	21	81	5.4	4.0	10	0	0.39	0.29	1.00	0.00	225	16.36	14.13	30	1
19	5	25	78	5.2	3.75	10	0	0.33	0.26	1.00	0.00	181	13.59	13.86	30	1
20	5	25	88	5.87	3.66	10	0	0.16	0.18	0.5	0.00	135	10.64	13.0	30	1

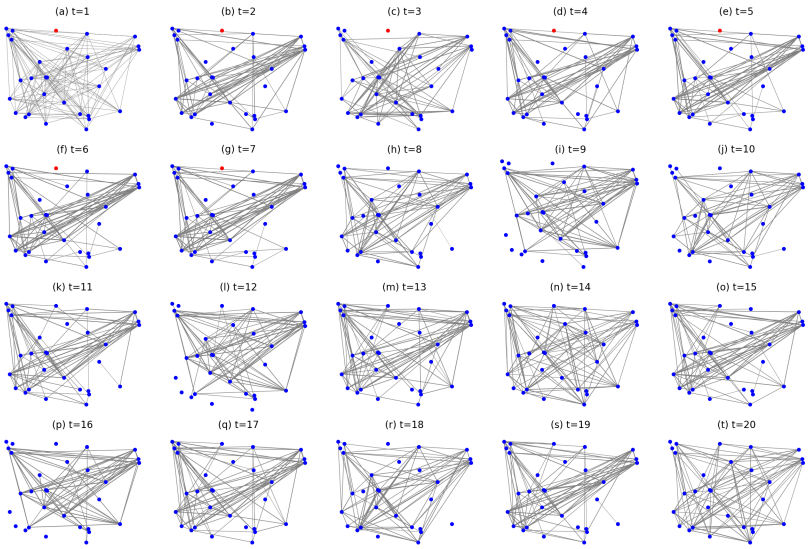


Figure D.9: The evolving social networks driven by cooperative nodes in an epidemic outbreak.

D.1.2.5 Collaborative

Table D.10: Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 10.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	225	15.0	4.18	26	0	0.56	0.11	0.67	0.00	29	3.35	7.14	30	1
2	1	29	214	14.27	5.09	28	0	0.59	0.13	0.89	0.00	29	3.37	7.13	30	1
3	1	29	211	14.07	5.16	28	0	0.59	0.13	0.79	0.00	29	3.38	7.13	30	1
4	1	29	216	14.4	5.14	28	0	0.59	0.12	0.72	0.00	29	3.37	7.13	30	1
5	1	29	213	14.2	5.06	28	0	0.59	0.12	0.69	0.00	29	3.38	7.13	30	1
6	1	29	213	14.2	5.26	28	0	0.6	0.13	0.76	0.00	29	3.38	7.13	30	1
7	0	30	220	14.67	4.64	25	1	0.57	0.13	0.75	0.00	0	1.53	0.57	3	1
8	1	29	222	14.8	4.84	24	0	0.59	0.13	0.73	0.00	29	3.36	7.14	30	1
9	0	30	214	14.27	3.34	20	8	0.51	0.06	0.65	0.33	0	1.51	0.5	2	1
10	0	30	229	15.27	2.95	22	11	0.53	0.05	0.68	0.45	0	1.47	0.5	2	1
11	0	30	217	14.47	4.17	26	7	0.58	0.09	0.93	0.45	0	1.5	0.5	2	1
12	0	30	235	15.67	3.32	22	8	0.57	0.05	0.71	0.49	0	1.46	0.5	2	1
13	0	30	227	15.13	3.63	22	7	0.55	0.06	0.69	0.41	0	1.48	0.5	2	1
14	0	30	221	14.73	3.3	22	9	0.53	0.05	0.65	0.44	0	1.49	0.5	2	1
15	0	30	219	14.6	3.37	20	6	0.53	0.05	0.64	0.42	0	1.5	0.5	2	1
16	0	30	234	15.6	3.01	23	6	0.54	0.04	0.62	0.47	0	1.46	0.5	2	1
17	0	30	227	15.13	3.32	21	6	0.55	0.04	0.64	0.49	0	1.48	0.5	2	1
18	0	30	230	15.33	3.63	24	9	0.57	0.08	0.81	0.45	0	1.47	0.5	2	1
19	0	30	215	14.33	3.21	19	6	0.5	0.07	0.58	0.27	0	1.51	0.5	2	1
20	0	30	219	14.6	3.27	21	7	0.5	0.07	0.64	0.31	0	1.5	0.5	2	1

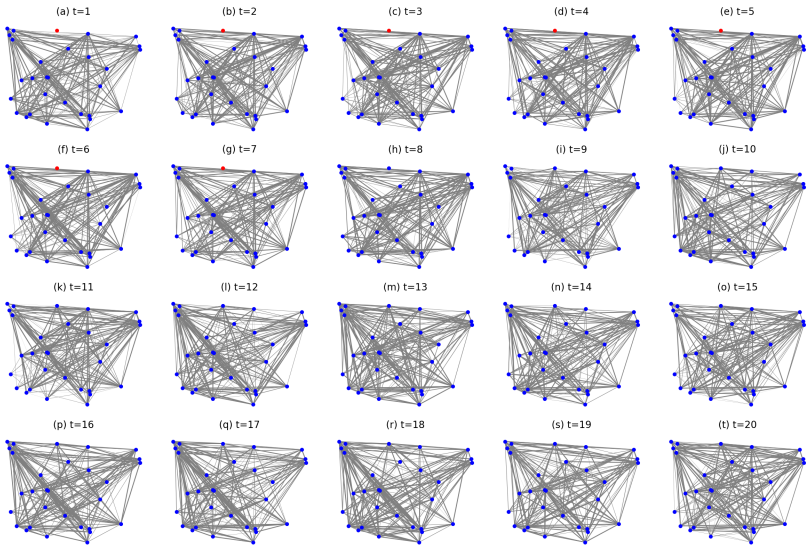


Figure D.10: The evolving social networks driven by collaborative nodes in an epidemic outbreak.

D.1.3 Social Capital Limit at 15

D.1.3.1 Inactive

Table D.11: Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	192	12.8	2.87	15	0	0.46	0.09	0.58	0.00	29	3.43	7.12	30	1
2	1	29	199	13.27	2.66	15	0	0.45	0.09	0.54	0.00	29	3.41	7.12	30	1
3	1	29	191	12.73	3.12	15	0	0.43	0.09	0.56	0.00	29	3.43	7.12	30	1
4	1	29	192	12.8	2.91	15	0	0.44	0.09	0.55	0.00	29	3.43	7.12	30	1
5	1	29	195	13.0	3.06	15	0	0.44	0.09	0.53	0.00	29	3.42	7.12	30	1
6	1	29	196	13.07	2.83	15	0	0.45	0.09	0.51	0.00	29	3.42	7.12	30	1
7	1	29	194	12.93	2.89	15	0	0.44	0.09	0.52	0.00	29	3.42	7.12	30	1
8	1	29	194	12.93	2.78	15	0	0.44	0.09	0.53	0.00	29	3.42	7.12	30	1
9	0	30	201	13.4	1.74	15	8	0.44	0.03	0.52	0.37	0	1.54	0.5	2	1
10	0	30	200	13.33	1.97	15	7	0.43	0.05	0.53	0.33	0	1.54	0.5	2	1
11	0	30	202	13.47	1.43	15	9	0.45	0.03	0.5	0.36	0	1.54	0.5	2	1
12	0	30	200	13.33	1.35	15	10	0.42	0.03	0.47	0.37	0	1.54	0.5	2	1
13	0	30	207	13.8	1.05	15	11	0.45	0.03	0.54	0.38	0	1.52	0.5	2	1
14	0	30	196	13.07	1.81	15	7	0.42	0.04	0.55	0.33	0	1.55	0.5	2	1
15	0	30	199	13.27	1.59	15	9	0.42	0.03	0.48	0.36	0	1.54	0.5	2	1
16	0	30	207	13.8	1.6	15	8	0.45	0.04	0.51	0.36	0	1.52	0.5	2	1
17	0	30	199	13.27	1.53	15	10	0.43	0.04	0.51	0.33	0	1.54	0.5	2	1
18	0	30	207	13.8	1.28	15	11	0.45	0.03	0.52	0.38	0	1.52	0.5	2	1
19	0	30	207	13.8	1.3	15	11	0.46	0.03	0.55	0.42	0	1.52	0.5	2	1
20	0	30	207	13.8	1.62	15	9	0.46	0.04	0.55	0.35	0	1.52	0.5	2	1

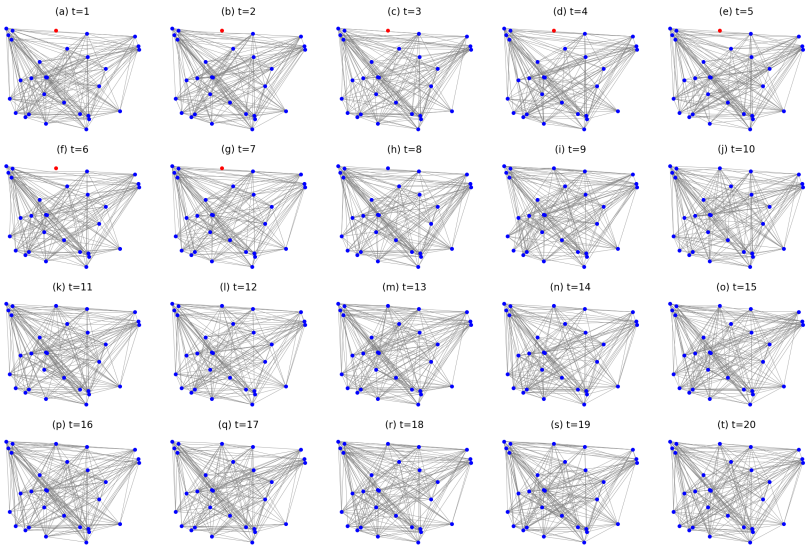


Figure D.11: The evolving social networks driven by inactive nodes in an epidemic outbreak.

D.1.3.2 Ignorant

Table D.12: Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 15.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	192	12.8	2.87	15	0	0.46	0.09	0.58	0.00	29	3.43	7.12	30	1
2	2	28	137	9.13	5.44	15	0	0.51	0.28	1.00	0.00	57	5.6	9.51	30	1
3	1	29	143	9.53	4.89	15	0	0.47	0.21	0.83	0.00	29	3.66	7.07	30	1
4	0	30	145	9.67	4.29	15	1	0.48	0.26	1.00	0.00	0	1.84	0.73	4	1
5	2	28	141	9.4	5.01	15	0	0.48	0.25	1.00	0.00	57	5.48	9.55	30	1
6	0	30	151	10.07	4.5	15	1	0.35	0.2	0.76	0.00	0	1.76	0.63	4	1
7	1	29	135	9.0	5.1	15	0	0.49	0.24	1.00	0.00	29	3.71	7.06	30	1
8	0	30	148	9.87	4.22	15	1	0.47	0.24	1.00	0.00	0	1.78	0.64	4	1
9	2	28	136	9.07	5.46	15	0	0.53	0.28	1.00	0.00	57	5.46	9.55	30	1
10	4	26	128	8.53	5.55	15	0	0.5	0.25	1.00	0.00	110	8.85	12.32	30	1
11	4	26	127	8.47	5.63	15	0	0.51	0.31	1.00	0.00	110	8.92	12.28	30	1
12	5	25	128	8.53	5.91	15	0	0.5	0.34	1.00	0.00	135	10.49	13.1	30	1
13	3	27	145	9.67	5.24	15	0	0.5	0.24	1.00	0.00	84	7.14	11.2	30	1
14	4	26	128	8.53	5.58	15	0	0.45	0.29	1.00	0.00	110	8.87	12.31	30	1
15	2	28	144	9.6	4.72	15	0	0.5	0.23	1.00	0.00	57	5.43	9.56	30	1
16	2	28	125	8.33	5.41	15	0	0.54	0.31	1.00	0.00	57	5.56	9.51	30	1
17	5	25	132	8.8	5.05	15	0	0.57	0.3	1.00	0.00	135	10.44	13.13	30	1
18	3	27	121	8.07	4.77	15	0	0.53	0.33	1.00	0.00	84	7.27	11.14	30	1
19	4	26	99	6.6	4.62	15	0	0.42	0.31	1.00	0.00	110	8.99	12.24	30	1
20	3	27	91	6.07	4.09	15	0	0.4	0.32	1.00	0.00	84	7.4	11.08	30	1

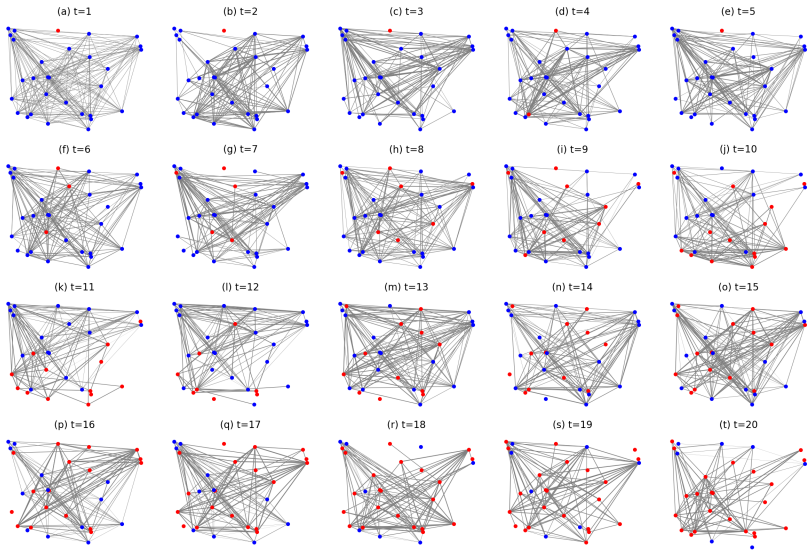


Figure D.12: The evolving social networks driven by ignorant nodes in an epidemic outbreak.

D.1.3.3 Egocentric

Table D.13: Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 15.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	192	12.8	2.87	15	0	0.46	0.09	0.58	0.00	29	3.43	7.12	30	1
2	5	25	139	9.27	6.02	15	0	0.36	0.31	0.79	0.00	135	10.43	13.14	30	1
3	0	30	146	9.73	5.3	15	1	0.58	0.29	1.00	0.00	0	1.91	0.77	4	1
4	1	29	147	9.8	5.22	15	0	0.45	0.31	0.8	0.00	29	3.71	7.06	30	1
5	2	28	165	11.00	4.5	15	0	0.4	0.2	1.00	0.00	57	5.34	9.59	30	1
6	0	30	160	10.67	4.17	15	4	0.65	0.27	1.00	0.14	0	1.87	0.77	4	1
7	2	28	135	9.0	5.0	15	0	0.63	0.35	1.00	0.00	57	5.59	9.51	30	1
8	2	28	139	9.27	4.33	15	0	0.64	0.3	1.00	0.00	57	5.59	9.51	30	1
9	3	27	128	8.53	4.73	15	0	0.69	0.31	1.00	0.00	84	7.89	10.91	30	1
10	1	29	146	9.73	4.77	15	0	0.68	0.28	1.00	0.00	29	4.13	7.01	30	1
11	1	29	148	9.87	4.54	15	0	0.57	0.21	0.8	0.00	29	3.82	7.05	30	1
12	0	30	152	10.13	3.97	15	2	0.59	0.2	1.00	0.21	0	1.8	0.68	3	1
13	0	30	153	10.2	3.67	15	3	0.51	0.19	1.00	0.27	0	1.8	0.69	4	1
14	1	29	155	10.33	3.67	15	0	0.45	0.19	1.00	0.00	29	3.59	7.08	30	1
15	1	29	160	10.67	3.88	15	0	0.46	0.17	0.83	0.00	29	3.56	7.09	30	1
16	0	30	148	9.87	4.19	15	1	0.66	0.23	1.00	0.00	0	1.91	0.79	4	1
17	2	28	136	9.07	4.15	15	0	0.53	0.25	1.00	0.00	57	5.59	9.51	30	1
18	2	28	144	9.6	4.42	15	0	0.48	0.27	1.00	0.00	57	5.51	9.53	30	1
19	1	29	150	10.00	4.43	15	0	0.38	0.23	1.00	0.00	29	3.69	7.07	30	1
20	2	28	121	8.07	4.05	15	0	0.44	0.29	1.00	0.00	57	5.59	9.5	30	1

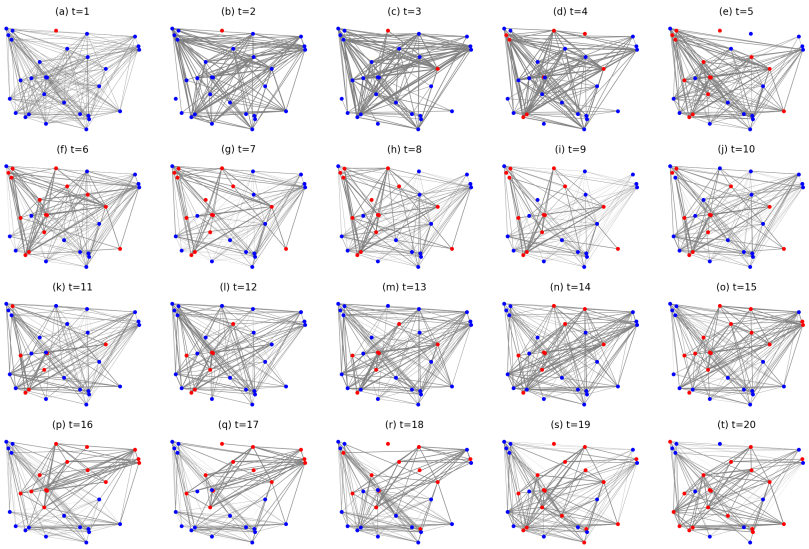


Figure D.13: The evolving social networks driven by egocentric nodes in an epidemic outbreak.

D.1.3.4 Cooperative

Table D.14: Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 15.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	192	12.8	2.87	15	0	0.46	0.09	0.58	0.00	29	3.43	7.12	30	1
2	4	26	145	9.67	5.42	15	0	0.5	0.26	1.00	0.00	110	8.83	12.33	30	1
3	1	29	151	10.07	4.75	15	0	0.44	0.24	1.00	0.00	29	3.63	7.08	30	1
4	2	28	147	9.8	5.1	15	0	0.46	0.22	0.67	0.00	57	5.42	9.56	30	1
5	2	28	150	10.00	5.22	15	0	0.46	0.3	1.00	0.00	57	5.45	9.56	30	1
6	1	29	155	10.33	5.21	15	0	0.47	0.27	0.76	0.00	29	3.75	7.06	30	1
7	1	29	151	10.07	5.05	15	0	0.46	0.29	0.83	0.00	29	3.72	7.07	30	1
8	1	29	153	10.2	4.84	15	0	0.45	0.24	1.00	0.00	29	3.62	7.08	30	1
9	1	29	140	9.33	5.88	15	0	0.42	0.32	0.78	0.00	29	3.93	7.03	30	1
10	0	30	147	9.8	5.53	15	1	0.5	0.38	1.00	0.00	0	2.07	0.93	4	1
11	0	30	159	10.6	4.24	15	2	0.48	0.22	1.00	0.00	0	1.77	0.67	4	1
12	4	26	136	9.07	6.26	15	0	0.46	0.35	0.86	0.00	110	9.05	12.22	30	1
13	0	30	154	10.27	4.69	15	1	0.5	0.23	1.00	0.00	0	1.83	0.74	4	1
14	3	27	140	9.33	6.02	15	0	0.52	0.34	1.00	0.00	84	7.26	11.15	30	1
15	0	30	159	10.6	4.69	15	1	0.48	0.23	0.93	0.00	56	5.31	9.51	30	1
16	4	26	135	9.0	6.03	15	0	0.5	0.31	0.84	0.00	110	8.97	12.26	30	1
17	1	29	156	10.4	4.64	15	0	0.53	0.22	1.00	0.00	29	3.65	7.08	30	1
18	6	24	136	9.07	6.45	15	0	0.5	0.36	1.00	0.00	159	12.02	13.66	30	1
19	1	29	136	9.07	6.03	15	0	0.47	0.4	0.92	0.00	29	4.17	7.01	30	1
20	0	30	159	10.6	4.63	15	1	0.51	0.23	1.00	0.00	0	1.84	0.8	4	1

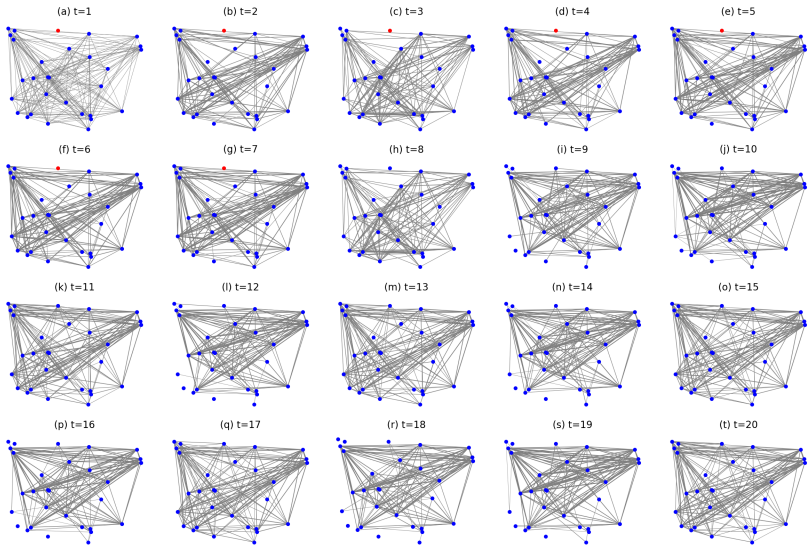


Figure D.14: The evolving social networks driven by cooperative nodes in an epidemic outbreak.

D.1.3.5 Collaborative

Table D.15: Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 15.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	302	20.13	4.45	26	0	0.72	0.14	0.8	0.00	29	3.17	7.18	30	1
2	1	29	284	18.93	5.4	28	0	0.73	0.15	0.95	0.00	29	3.21	7.17	30	1
3	1	29	290	19.33	5.09	28	0	0.72	0.14	0.83	0.00	29	3.2	7.18	30	1
4	1	29	282	18.8	5.35	28	0	0.73	0.14	0.88	0.00	29	3.22	7.17	30	1
5	1	29	290	19.33	4.96	28	0	0.71	0.14	0.79	0.00	29	3.2	7.18	30	1
6	0	30	279	18.6	5.23	28	1	0.71	0.14	0.81	0.00	0	1.37	0.51	3	1
7	1	29	283	18.87	5.45	28	0	0.74	0.15	0.88	0.00	29	3.22	7.17	30	1
8	0	30	276	18.4	5.29	29	1	0.7	0.14	0.85	0.00	0	1.37	0.48	2	1
9	0	30	287	19.13	4.78	29	8	0.73	0.07	0.86	0.64	0	1.34	0.47	2	1
10	0	30	286	19.07	4.17	25	10	0.71	0.06	0.82	0.61	0	1.34	0.47	2	1
11	0	30	303	20.2	3.39	28	11	0.72	0.04	0.79	0.66	0	1.3	0.46	2	1
12	0	30	277	18.47	4.59	29	7	0.71	0.07	0.9	0.59	0	1.36	0.48	2	1
13	0	30	300	20.00	3.56	26	12	0.71	0.04	0.82	0.65	0	1.31	0.46	2	1
14	0	30	278	18.53	4.84	29	7	0.72	0.07	0.9	0.59	0	1.36	0.48	2	1
15	0	30	280	18.67	5.37	29	8	0.76	0.09	0.93	0.62	0	1.36	0.48	2	1
16	0	30	264	17.6	4.83	29	8	0.67	0.07	0.78	0.51	0	1.39	0.49	2	1
17	0	30	303	20.2	4.69	29	11	0.76	0.07	0.93	0.65	0	1.3	0.46	2	1
18	0	30	288	19.2	4.45	28	10	0.72	0.05	0.85	0.64	0	1.34	0.47	2	1
19	0	30	288	19.2	4.76	29	6	0.74	0.07	1.00	0.64	0	1.34	0.47	2	1
20	0	30	276	18.4	4.39	26	8	0.69	0.07	0.83	0.56	0	1.37	0.48	2	1

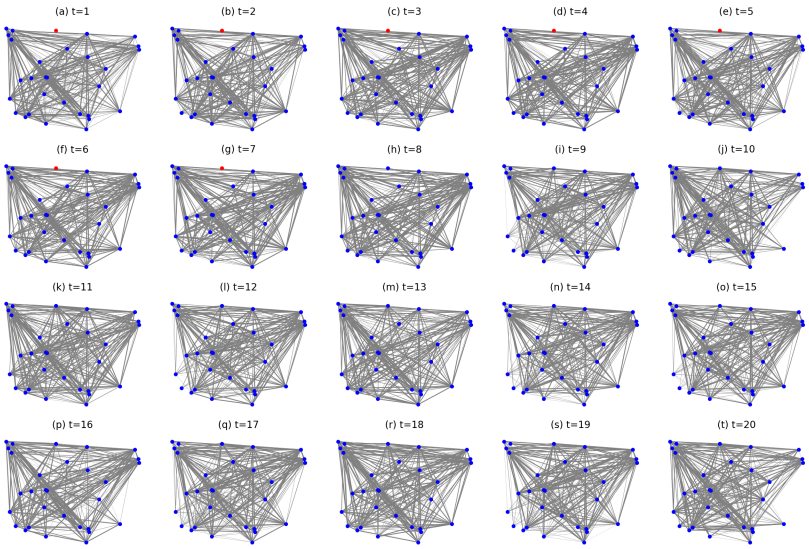


Figure D.15: The evolving social networks driven by collaborative nodes in an epidemic outbreak.

D.1.4 Social Capital Limit at 20

D.1.4.1 Inactive

Table D.16: Topological information of the network simulations driven by inactive mutation style under a social capital limit at 5.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	268	17.87	3.73	20	0	0.63	0.12	0.71	0.00	29	3.25	7.16	30	1
2	1	29	271	18.07	3.68	20	0	0.63	0.12	0.68	0.00	29	3.24	7.17	30	1
3	1	29	266	17.73	3.84	20	0	0.63	0.12	0.69	0.00	29	3.26	7.16	30	1
4	1	29	272	18.13	3.67	20	0	0.64	0.12	0.69	0.00	29	3.24	7.17	30	1
5	1	29	274	18.27	3.6	20	0	0.64	0.12	0.69	0.00	29	3.24	7.17	30	1
6	1	29	271	18.07	3.68	20	0	0.64	0.12	0.73	0.00	29	3.24	7.17	30	1
7	1	29	273	18.2	3.68	20	0	0.64	0.12	0.72	0.00	29	3.24	7.17	30	1
8	1	29	264	17.6	3.59	20	0	0.61	0.12	0.7	0.00	29	3.26	7.16	30	1
9	0	30	282	18.8	1.47	20	15	0.63	0.02	0.67	0.6	0	1.35	0.48	2	1
10	0	30	275	18.33	2.13	20	11	0.63	0.03	0.68	0.55	0	1.37	0.48	2	1
11	0	30	281	18.73	1.39	20	15	0.63	0.01	0.66	0.6	0	1.35	0.48	2	1
12	0	30	280	18.67	1.6	20	14	0.63	0.02	0.67	0.57	0	1.36	0.48	2	1
13	0	30	282	18.8	1.08	20	17	0.62	0.02	0.66	0.57	0	1.35	0.48	2	1
14	0	30	278	18.53	1.54	20	13	0.63	0.02	0.67	0.55	0	1.36	0.48	2	1
15	0	30	280	18.67	1.51	20	14	0.63	0.02	0.66	0.59	0	1.36	0.48	2	1
16	0	30	279	18.6	1.4	20	15	0.62	0.02	0.66	0.57	0	1.36	0.48	2	1
17	0	30	279	18.6	1.47	20	15	0.63	0.02	0.67	0.57	0	1.36	0.48	2	1
18	0	30	278	18.53	1.73	20	13	0.62	0.03	0.68	0.57	0	1.36	0.48	2	1
19	0	30	281	18.73	1.41	20	15	0.64	0.02	0.68	0.6	0	1.35	0.48	2	1
20	0	30	280	18.67	1.3	20	16	0.63	0.02	0.67	0.59	0	1.36	0.48	2	1

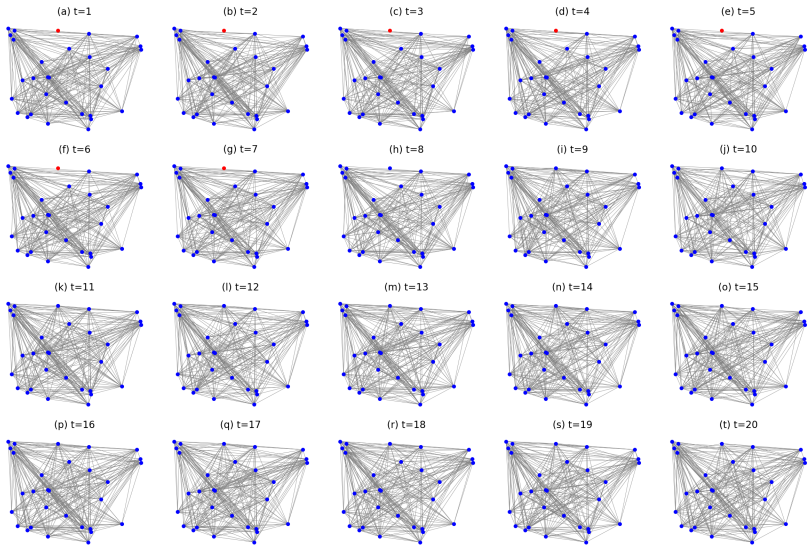


Figure D.16: The evolving social networks driven by inactive nodes in an epidemic outbreak.

D.1.4.2 Ignorant

Table D.17: Topological information of the network simulations driven by ignorant mutation style under a social capital limit at 20.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	268	17.87	3.73	20	0	0.63	0.12	0.71	0.00	29	3.25	7.16	30	1
2	1	29	216	14.4	6.48	20	0	0.64	0.24	0.97	0.00	29	3.42	7.13	30	1
3	1	29	223	14.87	6.06	20	0	0.67	0.18	1.00	0.00	29	3.38	7.13	30	1
4	0	30	227	15.13	4.91	20	4	0.64	0.13	1.00	0.39	0	1.5	0.54	3	1
5	1	29	225	15.0	5.54	20	0	0.64	0.18	1.00	0.00	29	3.38	7.13	30	1
6	0	30	229	15.27	5.35	20	2	0.59	0.15	1.00	0.2	0	1.51	0.56	3	1
7	1	29	220	14.67	6.26	20	0	0.65	0.21	1.00	0.00	29	3.38	7.13	30	1
8	0	30	226	15.07	5.14	20	5	0.64	0.14	1.00	0.2	0	1.5	0.54	3	1
9	3	27	216	14.4	7.0	20	0	0.71	0.25	1.00	0.00	84	6.93	11.3	30	1
10	4	26	198	13.2	7.62	20	0	0.72	0.3	1.00	0.00	110	8.64	12.44	30	1
11	4	26	154	10.27	6.65	20	0	0.62	0.34	1.00	0.00	110	8.8	12.35	30	1
12	4	26	147	9.8	6.92	20	0	0.56	0.4	1.00	0.00	110	8.81	12.34	30	1
13	2	28	216	14.4	7.05	20	0	0.65	0.28	0.96	0.00	57	5.22	9.64	30	1
14	2	28	170	11.33	6.9	20	0	0.63	0.33	1.00	0.00	57	5.37	9.59	30	1
15	2	28	171	11.4	5.66	20	0	0.57	0.29	1.00	0.00	57	5.31	9.6	30	1
16	4	26	87	5.8	4.97	19	0	0.52	0.38	1.00	0.00	110	8.99	12.23	30	1
17	5	25	84	5.6	4.18	17	0	0.47	0.35	1.00	0.00	135	10.66	12.99	30	1
18	6	24	87	5.8	4.04	13	0	0.42	0.36	1.00	0.00	159	12.14	13.57	30	1
19	4	26	99	6.6	5.0	20	0	0.5	0.34	1.00	0.00	110	8.94	12.26	30	1
20	4	26	101	6.73	4.6	20	0	0.51	0.33	1.00	0.00	110	8.95	12.26	30	1

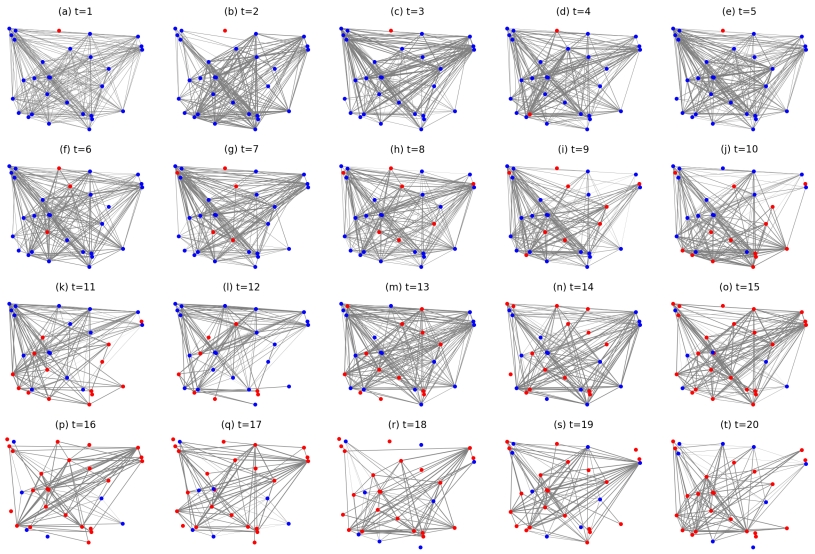


Figure D.17: The evolving social networks driven by ignorant nodes in an epidemic outbreak.

D.1.4.3 Egocentric

Table D.18: Topological information of the network simulations driven by egocentric mutation style under a social capital limit at 20.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	268	17.87	3.73	20	0	0.63	0.12	0.71	0.00	29	3.25	7.16	30	1
2	3	27	215	14.33	7.61	20	0	0.62	0.31	0.88	0.00	84	6.94	11.29	30	1
3	0	30	224	14.93	7.03	20	1	0.63	0.35	0.91	0.00	0	1.6	0.69	4	1
4	0	30	232	15.47	5.64	20	4	0.59	0.24	0.85	0.13	0	1.51	0.57	3	1
5	3	27	220	14.67	6.77	20	0	0.67	0.25	0.94	0.00	84	6.91	11.3	30	1
6	1	29	143	9.53	5.04	20	0	0.6	0.3	1.00	0.00	29	3.7	7.06	30	1
7	2	28	140	9.33	4.38	19	0	0.55	0.28	0.93	0.00	57	5.4	9.57	30	1
8	7	23	50	3.33	2.74	9	0	0.22	0.35	1.00	0.00	182	14.02	13.58	30	1
9	3	27	81	5.4	2.95	12	0	0.35	0.34	1.00	0.00	84	7.78	10.92	30	1
10	1	29	110	7.33	3.72	14	0	0.47	0.39	1.00	0.00	29	4.1	6.99	30	1
11	3	27	111	7.4	3.67	13	0	0.5	0.35	1.00	0.00	84	7.37	11.09	30	1
12	1	29	106	7.07	4.13	18	0	0.59	0.36	1.00	0.00	29	3.93	7.01	30	1
13	1	29	147	9.8	4.37	20	0	0.59	0.24	0.9	0.00	29	3.56	7.09	30	1
14	4	26	114	7.6	5.44	17	0	0.39	0.37	0.93	0.00	110	8.98	12.25	30	1
15	1	29	125	8.33	4.6	20	0	0.38	0.29	1.00	0.00	29	3.74	7.05	30	1
16	3	27	103	6.87	3.5	13	0	0.34	0.26	0.87	0.00	84	7.34	11.1	30	1
17	5	25	89	5.93	3.77	13	0	0.43	0.34	1.00	0.00	135	10.66	12.99	30	1
18	1	29	124	8.27	3.38	13	0	0.45	0.31	1.00	0.00	29	3.82	7.03	30	1
19	0	30	106	7.07	3.38	15	2	0.54	0.27	1.00	0.00	0	2.16	0.89	5	1
20	1	29	117	7.8	4.13	17	0	0.5	0.28	1.00	0.00	29	3.77	7.04	30	1

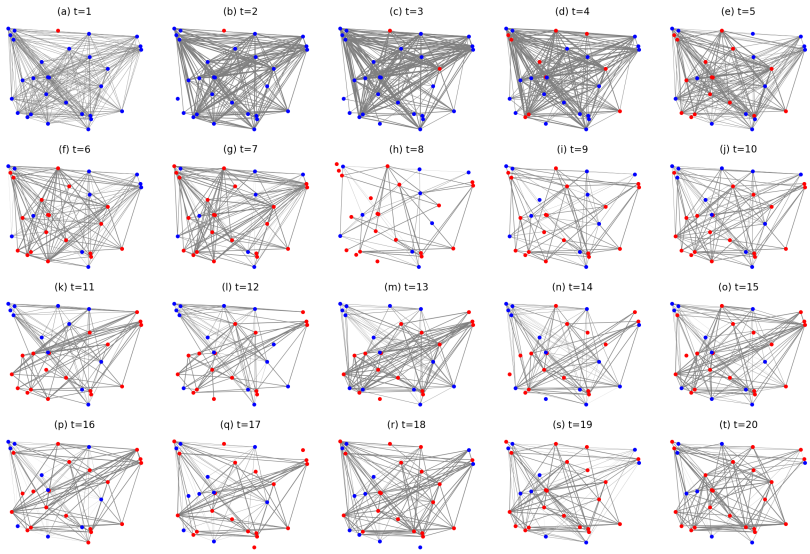


Figure D.18: The evolving social networks driven by egocentric nodes in an epidemic outbreak.

D.1.4.4 Cooperative

Table D.19: Topological information of the network simulations driven by cooperative mutation style under a social capital limit at 20.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	268	17.87	3.73	20	0	0.63	0.12	0.71	0.00	29	3.25	7.16	30	1
2	3	27	219	14.6	7.17	20	0	0.66	0.27	1.00	0.00	84	6.93	11.3	30	1
3	0	30	243	16.2	5.42	20	1	0.64	0.21	0.84	0.00	0	1.49	0.58	3	1
4	1	29	227	15.13	6.68	20	0	0.61	0.27	0.84	0.00	29	3.43	7.13	30	1
5	0	30	240	16.0	5.61	20	1	0.6	0.25	0.84	0.00	0	1.49	0.58	3	1
6	1	29	234	15.6	6.23	20	0	0.59	0.28	0.85	0.00	29	3.37	7.14	30	1
7	1	29	240	16.0	5.68	20	0	0.6	0.25	0.84	0.00	29	3.34	7.15	30	1
8	1	29	232	15.47	6.18	20	0	0.59	0.27	0.83	0.00	29	3.37	7.14	30	1
9	1	29	225	15.0	6.82	20	0	0.67	0.27	0.92	0.00	29	3.41	7.13	30	1
10	0	30	228	15.2	5.74	20	2	0.62	0.17	0.83	0.00	0	1.51	0.56	3	1
11	2	28	225	15.0	6.84	20	0	0.66	0.28	0.89	0.00	57	5.16	9.66	30	1
12	1	29	226	15.07	6.14	20	0	0.62	0.22	0.82	0.00	29	3.37	7.14	30	1
13	1	29	222	14.8	6.97	20	0	0.61	0.31	0.88	0.00	29	3.4	7.13	30	1
14	0	30	226	15.07	6.1	20	1	0.64	0.23	1.00	0.00	0	1.55	0.62	3	1
15	1	29	224	14.93	6.79	20	0	0.62	0.3	0.88	0.00	29	3.42	7.13	30	1
16	0	30	228	15.2	6.02	20	1	0.63	0.22	1.00	0.00	0	1.54	0.61	3	1
17	2	28	226	15.07	6.72	20	0	0.65	0.28	0.91	0.00	57	5.19	9.65	30	1
18	0	30	227	15.13	6.0	20	1	0.64	0.21	1.00	0.00	0	1.57	0.66	4	1
19	1	29	226	15.07	6.41	20	0	0.61	0.28	0.86	0.00	29	3.37	7.14	30	1
20	2	28	224	14.93	6.32	20	0	0.63	0.23	1.00	0.00	57	5.16	9.66	30	1

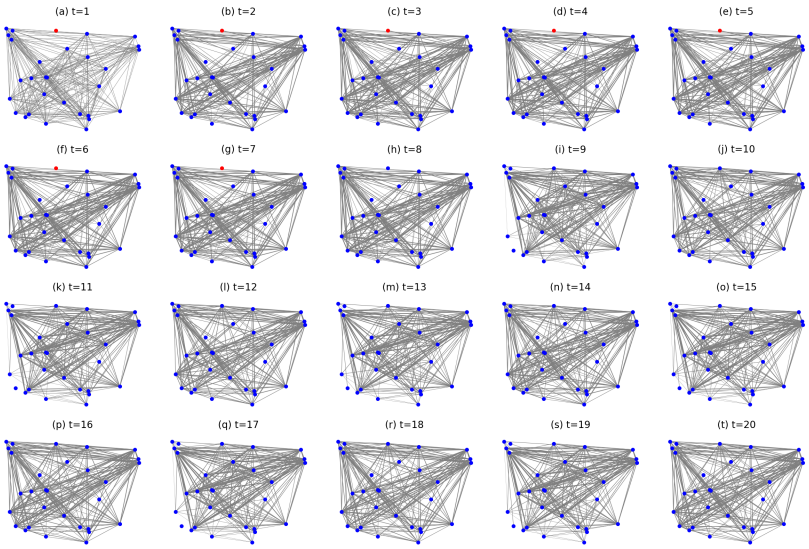


Figure D.19: The evolving social networks driven by cooperative nodes in an epidemic outbreak.

D.1.4.5 Collaborative

Table D.20: Topological information of the network simulations driven by collaborative mutation style under a social capital limit at 20.

Iteration	Nodes		Edges	Node Degree				Clustering coefficient				Shortest path length				
	Connected	Unconnected		Avg.	Std.	Max.	Min.	Avg.	Std.	Max.	Min.	Fake Paths	Avg.	Std.	Max.	Min.
0	30	0	0	0.00	0.00	0	0	0.00	0.00	0	0	435	30.00	0.00	30	30
1	1	29	356	23.73	4.85	27	0	0.85	0.16	0.9	0.00	29	3.05	7.21	30	1
2	0	30	323	21.53	5.71	29	1	0.81	0.16	0.95	0.00	0	1.26	0.44	2	1
3	0	30	329	21.93	5.88	29	1	0.83	0.17	0.98	0.00	0	1.24	0.43	2	1
4	0	30	325	21.67	5.89	29	1	0.83	0.16	0.94	0.00	0	1.25	0.43	2	1
5	0	30	329	21.93	6.07	29	1	0.84	0.17	1.00	0.00	0	1.24	0.43	2	1
6	0	30	303	20.2	6.12	29	1	0.8	0.16	1.00	0.00	0	1.3	0.46	2	1
7	0	30	324	21.6	5.87	29	1	0.82	0.16	1.00	0.00	0	1.26	0.44	2	1
8	0	30	319	21.27	5.85	29	1	0.81	0.16	0.95	0.00	0	1.27	0.44	2	1
9	0	30	321	21.4	5.15	29	8	0.82	0.07	0.97	0.71	0	1.26	0.44	2	1
10	0	30	339	22.6	4.25	29	12	0.83	0.04	0.94	0.76	0	1.22	0.41	2	1
11	0	30	324	21.6	4.74	29	8	0.8	0.06	0.9	0.7	0	1.26	0.44	2	1
12	0	30	350	23.33	3.82	29	13	0.84	0.04	0.94	0.78	0	1.2	0.4	2	1
13	0	30	342	22.8	4.36	29	10	0.84	0.05	0.94	0.76	0	1.21	0.41	2	1
14	0	30	322	21.47	5.21	29	8	0.82	0.07	0.94	0.71	0	1.26	0.44	2	1
15	0	30	348	23.2	4.06	29	11	0.84	0.05	0.95	0.78	0	1.2	0.4	2	1
16	0	30	317	21.13	5.26	29	10	0.81	0.08	0.96	0.71	0	1.27	0.44	2	1
17	0	30	356	23.73	3.85	29	15	0.86	0.05	0.94	0.79	0	1.18	0.39	2	1
18	0	30	316	21.07	5.32	29	10	0.81	0.08	0.95	0.68	0	1.27	0.45	2	1
19	0	30	351	23.4	3.83	29	13	0.84	0.04	0.94	0.78	0	1.19	0.39	2	1
20	0	30	321	21.4	5.1	29	9	0.81	0.06	0.93	0.7	0	1.26	0.44	2	1

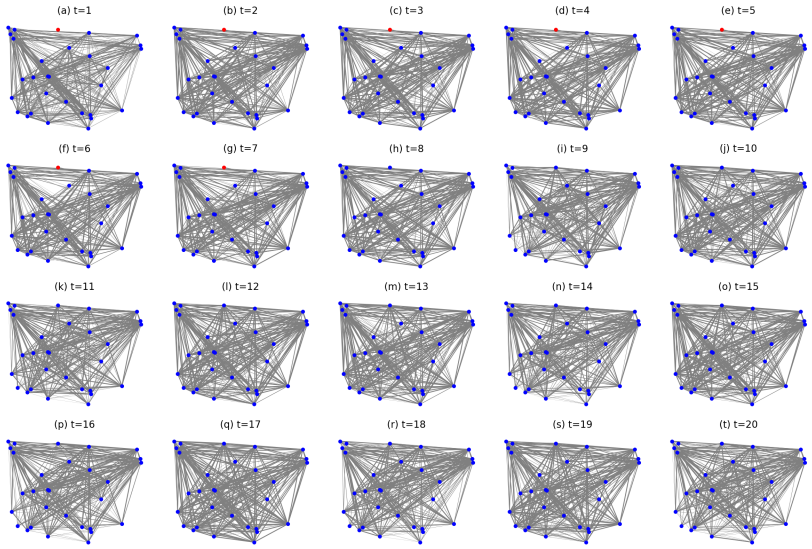


Figure D.20: The evolving social networks driven by collaborative nodes in an epidemic outbreak.

BIBLIOGRAPHY

[1] Jiaqi Wen, Bogdan Gabrys, and Katarzyna Musial.
Toward digital twin oriented modeling of complex networked systems and their dynamics: A comprehensive survey.
IEEE Access, 10:66886–66923, 2022.

[2] Jiaqi Wen, Bogdan Gabrys, and Katarzyna Musial.
Review and assessment of digital twin-oriented social network simulators.
IEEE Access, 11:97503–97521, 2023.

[3] Jiaqi Wen, Bogdan Gabrys, and Katarzyna Musial.
Towards digital twin-oriented complex networked systems: Introducing heterogeneous node features and interaction rules.
Plos one, 19(1):e0296426, 2024.

[4] Jiaqi Wen, Bogdan Gabrys, and Katarzyna Musial.
Heterogeneous feature representation for digital twin-oriented complex networked systems.
arXiv preprint arXiv:2309.13229, 2023.

[5] Alexander W Rives and Timothy Galitski.
Modular organization of cellular networks.
Proceedings of the national Academy of sciences, 100(3):1128–1133, 2003.

[6] Edward JS Hearnshaw and Mark MJ Wilson.
A complex network approach to supply chain network theory.
International Journal of Operations & Production Management, 33(4):442–469, 2013.

[7] Stella Heras, Katie Atkinson, Vicente Botti, Floriana Grasso, Vicente Julián, and Peter McBurney.
Research opportunities for argumentation in social networks.

Artificial Intelligence Review, 39:39–62, 2013.

[8] Linton Freeman.
The development of social network analysis.
A Study in the Sociology of Science, 1(687):159–167, 2004.

[9] Fei Gao, Katarzyna Musial, and Bogdan Gabrys.
A community bridge boosting social network link prediction model.
In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 683–689, 2017.

[10] Linyuan Lü and Tao Zhou.
Link prediction in complex networks: A survey.
Physica A: statistical mechanics and its applications, 390(6):1150–1170, 2011.

[11] Víctor Martínez, Fernando Berzal, and Juan-Carlos Cubero.
A survey of link prediction in complex networks.
ACM computing surveys (CSUR), 49(4):1–33, 2016.

[12] Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F Samatova.
Anomaly detection in dynamic networks: a survey.
Wiley Interdisciplinary Reviews: Computational Statistics, 7(3):223–247, 2015.

[13] PV Bindu, P Santhi Thilagam, and Deepesh Ahuja.
Discovering suspicious behavior in multilayer social networks.
Computers in Human Behavior, 73:568–582, 2017.

[14] Dariusz Król, Damien Fay, and Bogdan Gabryś.
Propagation phenomena in real world networks.
Springer, 2015.

[15] Cameron Nowzari, Victor M Preciado, and George J Pappas.
Analysis and control of epidemics: A survey of spreading processes on complex networks.
IEEE Control Systems Magazine, 36(1):26–46, 2016.

[16] Sen Pei and Hernán A Makse.
Spreading dynamics in complex networks.
Journal of Statistical Mechanics: Theory and Experiment, 2013(12):P12002, 2013.

[17] Junhuan Zhang, Peter McBurney, and Katarzyna Musial.
Convergence of trading strategies in continuous double auction markets with boundedly-rational networked traders.
Review of Quantitative Finance and Accounting, 50(1):301–352, 2018.

[18] Piotr Bródka, Katarzyna Musial, and Jaroslaw Jankowski.
Interacting spreading processes in multilayer networks: a systematic review.
IEEE Access, 8:10316–10341, 2020.

[19] Joakim Skardinga, Bogdan Gabrys, and Katarzyna Musial.
Foundations and modelling of dynamic networks using dynamic graph neural networks: A survey.
IEEE Access, 2021.

[20] Giulio Rossetti and Rémy Cazabet.
Community discovery in dynamic networks: a survey.
ACM Computing Surveys (CSUR), 51(2):1–37, 2018.

[21] Diego R Amancio.
Probing the topological properties of complex networks modeling short written texts.
PloS one, 10(2):e0118394, 2015.

[22] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas.
Characterization of complex networks: A survey of measurements.
Advances in physics, 56(1):167–242, 2007.

[23] Santo Fortunato, Alessandro Flammini, and Filippo Menczer.
Scale-free network growth by ranking.
Physical review letters, 96(21):218701, 2006.

[24] Viplove Arora and Mario Ventresca.
Action-based modeling of complex networks.
Scientific reports, 7(1):1–10, 2017.

[25] Leonardo Bargigli and Gabriele Tedeschi.
Interaction in agent-based economics: A survey on the network approach.
Physica A: Statistical Mechanics and its Applications, 399:1–15, 2014.

[26] Lorenzo Zino and Ming Cao.
Analysis, prediction, and control of epidemics: A survey from scalar to dynamic network models.
IEEE Circuits and Systems Magazine, 21(4):4–23, 2021.

[27] Sebastian Mežnar, Nada Lavrač, and Blaž Škrlj.
Prediction of the effects of epidemic spreading with graph neural networks.
In *International Conference on Complex Networks and Their Applications*, pages 420–431. Springer, 2020.

[28] Sebastian Mežnar, Nada Lavrač, and Blaž Škrlj.
Transfer learning for node regression applied to spreading prediction.
arXiv preprint arXiv:2104.00088, 2021.

[29] Mohammad Reza Sanatkar, Warren N White, Balasubramaniam Natarajan, Caterina M Scoglio, and Karen A Garrett.
Epidemic threshold of an sis model in dynamic switching networks.
IEEE Transactions on Systems, Man, and Cybernetics: Systems, 46(3):345–355, 2015.

[30] Huanyu Zhao and Ju H Park.
Dynamic output feedback consensus of continuous-time networked multiagent systems.
Complexity, 20(5):35–42, 2015.

[31] Romualdo Pastor-Satorras and Alessandro Vespignani.
Epidemic spreading in scale-free networks.
Physical review letters, 86(14):3200, 2001.

[32] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley.
The effect of network topology on the spread of epidemics.
In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1455–1466. IEEE, 2005.

[33] Eric W Fox, Martin B Short, Frederic P Schoenberg, Kathryn D Coronges, and Andrea L Bertozzi.
Modeling e-mail networks and inferring leadership using self-exciting point processes.
Journal of the American Statistical Association, 111(514):564–584, 2016.

[34] Akanda Wahid-Ul Ashraf, Marcin Budka, and Katarzyna Musial.
Simulation and augmentation of social networks for building deep learning models.
arXiv preprint arXiv:1905.09087, 2019.

[35] Francesco Sanna Passino and Nicholas A Heard.
Mutually exciting point process graphs for modeling dynamic networks.
Journal of Computational and Graphical Statistics, 32(1):116–130, 2023.

[36] Stefan Boschert, Christoph Heinrich, and Roland Rosen.
Next generation digital twin.
In *Proc. tmce*, pages 209–218. Las Palmas de Gran Canaria, Spain, 2018.

[37] Michael Grieves.
Digital twin: manufacturing excellence through virtual factory replication.
White paper, 1:1–7, 2014.

[38] Melanie Jans-Singh, Kathryn Leeming, Ruchi Choudhary, and Mark Girolami.
Digital twin of an urban-integrated hydroponic farm.
Data-Centric Engineering, 1, 2020.

[39] Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui.
Digital twin-driven product design, manufacturing and service with big data.
The International Journal of Advanced Manufacturing Technology, 94(9):3563–3576, 2018.

[40] Fei Tao, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, Zirong Guo, Stephen C-Y Lu, and AYC Nee.
Digital twin-driven product design framework.
International Journal of Production Research, 57(12):3935–3953, 2019.

[41] Angira Sharma, Edward Kosasih, Jie Zhang, Alexandra Brintrup, and Anisoara Calinescu.
Digital twins: State of the art theory and practice, challenges, and open research questions.
Journal of Industrial Information Integration, page 100383, 2022.

[42] Luciano da Fontoura Costa, Osvaldo N Oliveira Jr, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha.

Analyzing and modeling real-world phenomena with complex networks: a survey of applications.
Advances in Physics, 60(3):329–412, 2011.

[43] Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Bacşar, and Jean-Pierre Hubaux.
Game theory meets network security and privacy.
ACM Computing Surveys (CSUR), 45(3):1–39, 2013.

[44] Mikkel N Schmidt and Morten Morup.
Nonparametric bayesian modeling of complex networks: An introduction.
IEEE Signal Processing Magazine, 30(3):110–128, 2013.

[45] Albert-László Barabási and Réka Albert.
Emergence of scaling in random networks.
science, 286(5439):509–512, 1999.

[46] Akanda Wahid-Ul-Ashraf, Marcin Budka, and Katarzyna Musial.
How to predict social relationships‚Äîphysics-inspired approach to link prediction.
Physica A: Statistical Mechanics and its Applications, 523:1110–1129, 2019.

[47] Aili Asikainen, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski, and Mikko Kivelä.
Cumulative effects of triadic closure and homophily in social networks.
Science Advances, 6(19):eaax7310, 2020.

[48] Jonathan PK Doye.
Network topology of a potential energy landscape: A static scale-free network.
Physical review letters, 88(23):238701, 2002.

[49] Junhuan Zhang, Yuqian Xu, and Daniel Houser.
Vulnerability of scale-free cryptocurrency networks to double-spending attacks.
The European Journal of Finance, pages 1–15, 2021.

[50] Junhuan Zhang.
Influence of individual rationality on continuous double auction markets with networked traders.
Physica A: Statistical Mechanics and its Applications, 495:353–392, 2018.

[51] Wentao Wang, Junhuan Zhang, Shangmei Zhao, and Yanglin Zhang.
Simulation of asset pricing in information networks.
Physica A: Statistical Mechanics and its Applications, 513:620–634, 2019.

[52] Wentao Wang, Shangmei Zhao, and Junhuan Zhang.
Multi-asset pricing modeling using holding-based networks in energy markets.
Finance Research Letters, page 102483, 2021.

[53] Hawoong Jeong, Zoltan Néda, and Albert-László Barabási.
Measuring preferential attachment in evolving networks.
EPL (Europhysics Letters), 61(4):567, 2003.

[54] Dimitrios Tsiotas.
Preferential attachment: a multi-attribute growth process generating scale-free networks of different topologies.
arXiv preprint arXiv:2001.05167, 2020.

[55] Alireza Abbasi, Liaquat Hossain, and Loet Leydesdorff.
Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks.
Journal of Informetrics, 6(3):403–412, 2012.

[56] Miller McPherson, Lynn Smith-Lovin, and James M Cook.
Birds of a feather: Homophily in social networks.
Annual review of sociology, 27(1):415–444, 2001.

[57] Gueorgi Kossinets and Duncan J Watts.
Origins of homophily in an evolving social network.
American journal of sociology, 115(2):405–450, 2009.

[58] Vincent Boucher.
Structural homophily.
International Economic Review, 56(1):235–264, 2015.

[59] Gokhan Ertug, Julia Brennecke, Balázs Kovács, and Tengjian Zou.
What does homophily do? a review of the consequences of homophily.
Academy of Management Annals, 16(1):38–69, 2022.

[60] Abayomi M Ajofoyinbo, Vincent OS Olunloyo, Oye Ibidapo-Obe, et al.

On development of fuzzy controller: The case of gaussian and triangular membership functions.
Journal of Signal and Information Processing, 2(04):257, 2011.

[61] Duncan J Watts and Steven H Strogatz.
Collective dynamics of ‚Äòsmall-world‚Äônetworks.
nature, 393(6684):440–442, 1998.

[62] İsmail Güneş, Şule Gündüz-Öğüdücü, and Zehra Çataltepe.
Link prediction using time series of neighborhood-based node similarity scores.
Data Mining and Knowledge Discovery, 30(1):147–180, 2016.

[63] Giovanni Petri and Alain Barrat.
Simplicial activity driven model.
Physical review letters, 121(22):228301, 2018.

[64] Per Block, Marion Hoffman, Isabel J Raabe, Jennifer Beam Dowd, Charles Rahal, Ridhi Kashyap, and Melinda C Mills.
Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world.
Nature Human Behaviour, 4(6):588–596, 2020.

[65] Zsófia Boda, Timon Elmer, András Vörös, and Christoph Stadtfeld.
Short-term and long-term effects of a social network intervention on friendships among university students.
Scientific reports, 10(1):1–12, 2020.

[66] Liuyan Shi, Liang Zhang, and Yun Lu.
Evaluating social network-based weight loss interventions in chinese population: An agent-based simulation.
Plos one, 15(8):e0236716, 2020.

[67] Feng Liu, Xin Li, and Gaofeng Zhu.
Using the contact network model and metropolis-hastings sampling to reconstruct the covid-19 spread on the ‚Äúdiamond princess‚Äù.
Science bulletin, 65(15):1297–1305, 2020.

[68] Joon-Seok Kim, Hamdi Kavak, Chris Ovi Rouly, Hyunjee Jin, Andrew Crooks, Dieter Pfoser, Carola Wenk, and Andreas Züfle.

Location-based social simulation for prescriptive analytics of disease spread.
SIGSPATIAL Special, 12(1):53–61, 2020.

[69] Marko Gosak, Moritz UG Kraemer, Heinrich H Nax, Matjaž Perc, and Bary SR Pradelski.
Endogenous social distancing and its underappreciated impact on the epidemic curve.
Scientific reports, 11(1):1–10, 2021.

[70] Alessandra Fogli and Laura Veldkamp.
Germs, social networks, and growth.
The Review of Economic Studies, 88(3):1074–1100, 2021.

[71] Paul W Holland and Samuel Leinhardt.
Transitivity in structural models of small groups.
Comparative group studies, 2(2):107–124, 1971.

[72] Wen, Jiaqi, Bogdan Gabrys, and Katarzyna Musial.
Evolutionary Digital Twin-Oriented Complex Networked Systems Driven by Node Features and the Mutation of Feature Preferences.
2023.

[73] Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse.
Knowledge discovery in cryptocurrency transactions: A survey.
IEEE Access, 2021.

[74] Clayson Celes, Azzedine Boukerche, and Antonio AF Loureiro.
Mobility trace analysis for intelligent vehicular networks: Methods, models, and applications.
ACM Computing Surveys (CSUR), 54(3):1–38, 2021.

[75] Erfan Babaee Tirkolaee, Alireza Goli, and Gerhard-Wilhelm Weber.
Multi-objective aggregate production planning model considering overtime and outsourcing options under fuzzy seasonal demand.
In *International Scientific-Technical Conference Manufacturing*, pages 81–96. Springer, 2019.

[76] Komal Batool and Muaz A Niazi.

Modeling the internet of things: a hybrid modeling approach using complex networks and agent-based models.
Complex Adaptive Systems Modeling, 5(1):1–19, 2017.

[77] Frauke Degenhardt, Stephan Seifert, and Silke Szymczak.
Evaluation of variable selection methods for random forests and omics data sets.
Briefings in bioinformatics, 20(2):492–503, 2019.

[78] Jake Lever, Martin Krzywinski, and Naomi Altman.
Points of significance: model selection and overfitting.
Nature methods, 13(9):703–705, 2016.

[79] Luis A Aguirre, Leonardo L Portes, and Christophe Letellier.
Structural, dynamical and symbolic observability: From dynamical systems to networks.
PLoS One, 13(10):e0206180, 2018.

[80] Thomas Kailath.
Linear systems, volume 156.
Prentice-Hall Englewood Cliffs, NJ, 1980.

[81] Dániel Leitold, Ágnes Vathy-Fogarassy, and János Abonyi.
Controllability and observability in complex networks–the effect of connection types.
Scientific reports, 7(1):1–9, 2017.

[82] Irene Sendiña-Nadal and Christophe Letellier.
Observability of dynamical networks from graphic and symbolic approaches.
In *International Workshop on Complex Networks*, pages 3–15. Springer, 2019.

[83] Francesco Lo Iudice, Francesco Sorrentino, and Franco Garofalo.
On node controllability and observability in complex dynamical networks.
IEEE Control Systems Letters, 3(4):847–852, 2019.

[84] Andrew J Whalen, Sean N Brennan, Timothy D Sauer, and Steven J Schiff.
Observability and controllability of nonlinear networks: The role of symmetry.
Physical Review X, 5(1):011005, 2015.

[85] Christophe Letellier, Irene Sendina-Nadal, and Luis A Aguirre.
Nonlinear graph-based theory for dynamical network observability.

Physical Review E, 98(2):020303, 2018.

[86] Hamidreza Alvari, Alireza Hajibagheri, and Gita Sukthankar.
Community detection in dynamic social networks: A game-theoretic approach.
In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 101–107. IEEE, 2014.

[87] Meng Qin, Di Jin, Kai Lei, Bogdan Gabrys, and Katarzyna Musial-Gabrys.
Adaptive community detection incorporating topology and content in social networks.
Knowledge-based systems, 161:342–356, 2018.

[88] Di Jin, Bogdan Gabrys, and Jianwu Dang.
Combined node and link partitions method for finding overlapping communities in complex networks.
Scientific reports, 5(1):1–8, 2015.

[89] Martin Rosvall and Carl T Bergstrom.
Mapping change in large networks.
PloS one, 5(1):e8694, 2010.

[90] Nam P Nguyen, Thang N Dinh, Ying Xuan, and My T Thai.
Adaptive algorithms for detecting community structure in dynamic social networks.
In *2011 Proceedings IEEE INFOCOM*, pages 2282–2290. IEEE, 2011.

[91] Urs Braun, Axel Schäfer, Henrik Walter, Susanne Erk, Nina Romanczuk-Seiferth, Leila Haddad, Janina I Schweiger, Oliver Grimm, Andreas Heinz, Heike Tost, et al.
Dynamic reconfiguration of frontal brain networks during executive cognition in humans.
Proceedings of the National Academy of Sciences, 112(37):11678–11683, 2015.

[92] Kevin T Macon, Peter J Mucha, and Mason A Porter.
Community structure in the united nations general assembly.
Physica A: Statistical Mechanics and its Applications, 391(1-2):343–361, 2012.

[93] David Liben-Nowell and Jon Kleinberg.
The link-prediction problem for social networks.

Journal of the American society for information science and technology, 58(7):1019–1031, 2007.

[94] Aaron Clauset, Cristopher Moore, and Mark EJ Newman.
Hierarchical structure and the prediction of missing links in networks.
Nature, 453(7191):98–101, 2008.

[95] Francois Fouss, Luh Yen, Alain Pirotte, and Marco Saerens.
An experimental investigation of graph kernels on a collaborative recommendation task.
In *Sixth International Conference on Data Mining (ICDM'06)*, pages 863–868. IEEE, 2006.

[96] Mojtaba Nayyeri, Gokce Muge Cil, Sahar Vahdati, Francesco Osborne, Mahfuzur Rahman, Simone Angioni, Angelo Salatino, Diego Reforgiato Recupero, Nadezhda Vassilyeva, Enrico Motta, et al.
Trans4e: Link prediction on scholarly knowledge graphs.
Neurocomputing, 461:530–542, 2021.

[97] Linyuan Lü, Ci-Hang Jin, and Tao Zhou.
Similarity index based on local paths for link prediction of complex networks.
Physical Review E, 80(4):046122, 2009.

[98] Giovanni Battista Rossi.
Measurability.
Measurement, 40(6):545–562, 2007.

[99] Kai Lei, Meng Qin, Bo Bai, Gong Zhang, and Min Yang.
Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks.
In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 388–396. IEEE, 2019.

[100] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, and Bryan Hooi.
Nodeaug: Semi-supervised node classification with data augmentation.
In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 207–217, 2020.

[101] Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou.

Synchronization in complex networks.
Physics reports, 469(3):93–153, 2008.

[102] Florian Dorfler and Francesco Bullo.
Synchronization and transient stability in power networks and nonuniform kuramoto oscillators.
SIAM Journal on Control and Optimization, 50(3):1616–1642, 2012.

[103] Florian Dörfler, Michael Chertkov, and Francesco Bullo.
Synchronization in complex oscillator networks and smart grids.
Proceedings of the National Academy of Sciences, 110(6):2005–2010, 2013.

[104] Ed Bullmore and Olaf Sporns.
The economy of brain network organization.
Nature Reviews Neuroscience, 13(5):336–349, 2012.

[105] Yang Tang, Huijun Gao, and Jürgen Kurths.
Distributed robust synchronization of dynamical networks with stochastic coupling.
IEEE Transactions on Circuits and Systems I: Regular Papers, 61(5):1508–1519, 2013.

[106] Constance Hammond, Hagai Bergman, and Peter Brown.
Pathological synchronization in parkinson's disease: networks, models and treatments.
Trends in neurosciences, 30(7):357–364, 2007.

[107] Rachel CW Wong, Tony WY Yuen, Kwok Wah Fung, and Janny MY Leung.
Optimizing timetable synchronization for rail mass transit.
Transportation Science, 42(1):57–69, 2008.

[108] Mahesh Wickramasinghe and István Z Kiss.
Spatially organized dynamical states in chemical oscillator networks: Synchronization, dynamical differentiation, and chimera patterns.
PloS one, 8(11):e80586, 2013.

[109] Insoo Sohn.
Small-world and scale-free network models for iot systems.
Mobile Information Systems, 2017, 2017.

[110] Remy Cazabet and Frederic Amblard.
Simulate to detect: a multi-agent system for community detection.
In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, volume 2, pages 402–408. IEEE, 2011.

[111] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang.
Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks.
In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2672–2681, 2018.

[112] Lucy Kendrick, Katarzyna Musial, and Bogdan Gabrys.
Change point detection in social networks,Äîcritical review with experiments.
Computer Science Review, 29:1–13, 2018.

[113] Marina Thottan and Chuanyi Ji.
Anomaly detection in ip networks.
IEEE Transactions on signal processing, 51(8):2191–2204, 2003.

[114] Sutharshan Rajasegarar, Christopher Leckie, and Marimuthu Palaniswami.
Anomaly detection in wireless sensor networks.
IEEE Wireless Communications, 15(4):34–40, 2008.

[115] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita.
Network anomaly detection: methods, systems and tools.
Ieee communications surveys & tutorials, 16(1):303–336, 2013.

[116] Cancer Genome Atlas Research Network et al.
Comprehensive genomic characterization defines human glioblastoma genes and core pathways.
Nature, 455(7216):1061, 2008.

[117] Bin Jia, Cailing Dong, Zhijiang Chen, Kuo-Chu Chang, Nichole Sullivan, and Genshe Chen.
Pattern discovery and anomaly detection via knowledge graph.
In *2018 21st International Conference on Information Fusion (FUSION)*, pages 2392–2399. IEEE, 2018.

[118] Ramon Ferrer i Cancho, Ricard V Solé, and Reinhard Köhler.
Patterns in syntactic dependency networks.
Physical Review E, 69(5):051915, 2004.

[119] André Ricardo Backes, Dalcimar Casanova, and Odemir Martinez Bruno.
A complex network-based approach for boundary shape analysis.
Pattern Recognition, 42(1):54–67, 2009.

[120] Sara Kaviani and Insoo Sohn.
Influence of random topology in artificial neural networks: A survey.
ICT Express, 6(2):145–150, 2020.

[121] Matt J Keeling and Ken TD Eames.
Networks and epidemic models.
Journal of the Royal Society Interface, 2(4):295–307, 2005.

[122] Frédéric Amblard and Guillaume Deffuant.
The role of network topology on extremism propagation with the relative agreement opinion dynamics.
Physica A: Statistical Mechanics and its Applications, 343:725–738, 2004.

[123] Youngsang Cho, Junseok Hwang, and Daeho Lee.
Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach.
Technological Forecasting and Social Change, 79(1):97–106, 2012.

[124] Xinyue Ye, Lanxue Dang, Jay Lee, Ming-Hsiang Tsou, and Zhuo Chen.
Open source social network simulator focusing on spatial meme diffusion.
In *Human dynamics research in smart and connected communities*, pages 203–222. Springer, 2018.

[125] Luigi Atzori, Antonio Iera, Giacomo Morabito, and Michele Nitti.
The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization.
Computer networks, 56(16):3594–3608, 2012.

[126] Harold Soh, Sonja Lim, Tianyou Zhang, Xiuju Fu, Gary Kee Khoon Lee, Terence Gih Guang Hung, Pan Di, Silvester Prakasam, and Limsoon Wong.

Weighted complex network analysis of travel routes on the singapore public transportation system.
Physica A: Statistical Mechanics and its Applications, 389(24):5852–5863, 2010.

[127] Romero Morais, Vuong Le, Svetha Venkatesh, and Truyen Tran.
Learning asynchronous and sparse human-object interaction in videos.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16041–16050, 2021.

[128] Michele Coscia, Fosca Giannotti, and Dino Pedreschi.
A classification for community discovery methods in complex networks.
Statistical Analysis and Data Mining: The ASA Data Science Journal, 4(5):512–546, 2011.

[129] Rudolf Emil Kalman.
Mathematical description of linear dynamical systems.
Journal of the Society for Industrial and Applied Mathematics, Series A: Control, 1(2):152–192, 1963.

[130] Marco Cremonini and Francesca Casamassima.
Controllability of social networks and the strategic use of random information.
Computational social networks, 4(1):1–22, 2017.

[131] Stefan Wuchty.
Controllability in protein interaction networks.
Proceedings of the National Academy of Sciences, 111(19):7156–7160, 2014.

[132] Teresa M Karrer, Jason Z Kim, Jennifer Stiso, Ari E Kahn, Fabio Pasqualetti, Ute Habel, and Danielle S Bassett.
A practical guide to methodological considerations in the controllability of structural brain networks.
Journal of neural engineering, 17(2):026031, 2020.

[133] Marco Rinaldi.
Controllability of transportation networks.
Transportation Research Part B: Methodological, 118:381–406, 2018.

[134] Krzysztof Juszczyszyn, Przemyslaw Kazienko, Katarzyna Musial, and Bogdan Gabrys.

Temporal changes in connection patterns of an email-based social network.
In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 3, pages 9–12. IEEE, 2008.

[135] Mingshan Jia, Bogdan Gabrys, and Katarzyna Musial.
Directed closure coefficient and its patterns.
Plos one, 16(6):e0253822, 2021.

[136] Mingshan Jia, Bogdan Gabrys, and Katarzyna Musial.
Measuring quadrangle formation in complex networks.
IEEE Transactions on Network Science and Engineering, 2021.

[137] Reza Lotfi, Yahia Zare Mehrjerdi, Mir Saman Pishvaee, Ahmad Sadeghieh, and Gerhard-Wilhelm Weber.
A robust optimization model for sustainable and resilient closed-loop supply chain network design considering conditional value at risk.
Numerical Algebra, Control & Optimization, 11(2):221, 2021.

[138] Erfan Babaee Tirkolaee and Nadi Serhan Aydin.
Integrated design of sustainable supply chain and transportation network using a fuzzy bi-level decision support system for perishable products.
Expert Systems with Applications, page 116628, 2022.

[139] Reza Lotfi, Bahareh Kargar, Alireza Gharehbaghi, and Gerhard-Wilhelm Weber.
Viable medical waste chain network design by considering risk and robustness.
Environmental Science and Pollution Research, pages 1–16, 2021.

[140] Alireza Goli, Erfan Babaee Tirkolaee, and Gerhard-Wilhelm Weber.
A perishable product sustainable supply chain network design problem with lead time and customer satisfaction using a hybrid whale-genetic algorithm.
In *Logistics operations and management for recycling and reuse*, pages 99–124. Springer, 2020.

[141] Erfan Babaee Tirkolaee, Alireza Goli, Peiman Ghasemi, and Fariba Goodarzian.
Designing a sustainable closed-loop supply chain network of face masks during the covid-19 pandemic: Pareto-based algorithms.
Journal of Cleaner Production, 333:130056, 2022.

[142] Sebastian Haag and Reiner Anderl.

Digital twin - proof of concept.
Manufacturing Letters, 15:64 – 66, 2018.

[143] Dmitry Ivanov, Alexandre Dolgui, Ajay Das, and Boris Sokolov.
Digital supply chain twins: Managing the ripple effect, resilience, and disruption risks by data-driven optimization, simulation, and visibility.
In *Handbook of ripple effects in the supply chain*, pages 309–332. Springer, 2019.

[144] Azad M Madni, Carla C Madni, and Scott D Lucero.
Leveraging digital twin technology in model-based systems engineering.
Systems, 7(1):7, 2019.

[145] Fei Tao, Meng Zhang, Yushan Liu, and AYC Nee.
Digital twin driven prognostics and health management for complex equipment.
Cirp Annals, 67(1):169–172, 2018.

[146] Thomas H-J Uhlemann, Christian Lehmann, and Rolf Steinhilper.
The digital twin: Realizing the cyber-physical production system for industry 4.0.
Procedia Cirp, 61:335–340, 2017.

[147] Jing Du, Qi Zhu, Yangming Shi, Qi Wang, Yingzi Lin, and Daniel Zhao.
Cognition digital twins for personalized information systems of smart cities: Proof of concept.
Journal of Management in Engineering, 36(2):04019052, 2020.

[148] Ying Liu, Lin Zhang, Yuan Yang, Longfei Zhou, Lei Ren, Fei Wang, Rong Liu, Zhibo Pang, and M Jamal Deen.
A novel cloud-based framework for the elderly healthcare services using digital twin.
IEEE Access, 7:49088–49101, 2019.

[149] Andrea Castellani, Sebastian Schmitt, and Stefano Squartini.
Real-world anomaly detection by using digital twin systems and weakly supervised learning.
IEEE Transactions on Industrial Informatics, 17(7):4733–4742, 2020.

[150] Kazi Masudul Alam and Abdulmotaleb El Saddik.
C2ps: A digital twin architecture reference model for the cloud-based cyber-physical systems.
IEEE access, 5:2050–2062, 2017.

[151] Dhananjay Singh, Gaurav Tripathi, and Antonio J Jara.
A survey of internet-of-things: Future vision, architecture, challenges and services.
In *2014 IEEE world forum on Internet of Things (WF-IoT)*, pages 287–292. IEEE, 2014.

[152] Yunhua He, Hong Li, Xiuzhen Cheng, Yan Liu, Chao Yang, and Limin Sun.
A blockchain based truthful incentive mechanism for distributed p2p applications.
IEEE Access, 6:27324–27335, 2018.

[153] Weifeng Hao, Jiajie Zeng, Xiaohai Dai, Jiang Xiao, Qiang-Sheng Hua, Hanhua Chen, Kuan-Ching Li, and Hai Jin.
Towards a trust-enhanced blockchain p2p topology for enabling fast and reliable broadcast.
IEEE Transactions on Network and Service Management, 17(2):904–917, 2020.

[154] Yuri Breitbart, Minos Garofalakis, Cliff Martin, Rajeev Rastogi, S Seshadri, and Avi Silberschatz.
Topology discovery in heterogeneous ip networks.
In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064)*, volume 1, pages 265–274. IEEE, 2000.

[155] Joshua R Tyler, Dennis M Wilkinson, and Bernardo A Huberman.
Email as spectroscopy: Automated discovery of community structure within organizations.
In *Communities and technologies*, pages 81–96. Springer, 2003.

[156] Neil Shah, Alex Beutel, Bryan Hooi, Leman Akoglu, Stephan Gunnemann, Disha Makhija, Mohit Kumar, and Christos Faloutsos.
Edgecentric: Anomaly detection in edge-attributed networks.
In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 327–334. IEEE, 2016.

[157] Chengbin Hou, Shan He, and Ke Tang.
Rosane: Robust and scalable attributed network embedding for sparse networks.
Neurocomputing, 409:231–243, 2020.

[158] Tianjin Huang, Yulong Pei, Vlado Menkovski, and Mykola Pechenizkiy.
Hop-count based self-supervised anomaly detection on attributed networks.

pages 225–241, 2022.

[159] Katarzyna Musiał and Przemysław Kazienko.
Social networks on the internet.
World Wide Web, 16(1):31–72, 2013.

[160] Di Jin, Ziyang Liu, Dongxiao He, Bogdan Gabrys, and Katarzyna Musial.
Robust detection of communities with multi-semantics in large attributed networks.
In *International Conference on Knowledge Science, Engineering and Management*, pages 362–376. Springer, 2018.

[161] Jundong Li, Kewei Cheng, Liang Wu, and Huan Liu.
Streaming link prediction on dynamic attributed networks.
In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 369–377, 2018.

[162] Jian Li, Leonardo Duenas-Osorio, Changkun Chen, and Congling Shi.
Connectivity reliability and topological controllability of infrastructure networks: A comparative assessment.
Reliability Engineering & System Safety, 156:24–33, 2016.

[163] Hui Xia, Li Li, Xiangguo Cheng, Chao Liu, and Tie Qiu.
A dynamic virus propagation model based on social attributes in city iot.
IEEE Internet of Things Journal, 7(9):8036–8048, 2020.

[164] Xueyan Liu, Wenzhuo Song, Katarzyna Musial, Xuehua Zhao, Wanli Zuo, and Bo Yang.
Semi-supervised stochastic blockmodel for structure analysis of signed networks.
Knowledge-Based Systems, 195:105714, 2020.

[165] Shuai Liu, Zhijian Ji, and Huizi Ma.
Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs.
Complexity, 2020, 2020.

[166] Thuy Van T Duong and Dinh Que Tran.
An effective approach for mobility prediction in wireless network based on temporal weighted mobility rule.

International Journal of Computer Science and Telecommunications, 3(2):29–36, 2012.

[167] Aditya G Nair, Steven L Brunton, and Kunihiko Taira.
Networked-oscillator-based modeling and control of unsteady wake flows.
Physical Review E, 97(6):063107, 2018.

[168] Qinma Kang, Yunfan Kang, and Hanzhang Kong.
Edge-attributed community search for large graphs.
In *Proceedings of the 2nd International Conference on Big Data Research*, pages 114–118, 2018.

[169] Hayat Dino Bedru, Shuo Yu, Xinru Xiao, Da Zhang, Liangtian Wan, He Guo, and Feng Xia.
Big networks: A survey.
Computer Science Review, 37:100247, 2020.

[170] Tao Fei, Cheng Jiangfeng, Qi Qinglin, Meng Zhang, He Zhang, and Sui Fangyuan.
Digital twin-driven product design, manufacturing and service with big data.
The International Journal of Advanced Manufacturing Technology, 94(9-12):3563–3576, 2018.

[171] Petr Chunaev.
Community detection in node-attributed social networks: a survey.
Computer Science Review, 37:100286, 2020.

[172] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang.
Spectral redemption in clustering sparse networks.
Proceedings of the National Academy of Sciences, 110(52):20935–20940, 2013.

[173] Peter V Marsden and Nan Lin.
Social structure and network analysis.
Sage Publications California, 1982.

[174] Jill E Perry-Smith and Christina E Shalley.
The social side of creativity: A static and dynamic social network perspective.
Academy of management review, 28(1):89–106, 2003.

[175] Pavlos Kefalas, Panagiotis Symeonidis, and Yannis Manolopoulos.
Recommendations based on a heterogeneous spatio-temporal social network.
World Wide Web, 21(2):345–371, 2018.

[176] Xiaoli Li, Min Wu, Chee-Keong Kwoh, and See-Kiong Ng.
Computational approaches for detecting protein complexes from protein interaction networks: a survey.
BMC genomics, 11(1):1–19, 2010.

[177] Bolin Chen, Weiwei Fan, Juan Liu, and Fang-Xiang Wu.
Identifying protein complexes and functional modules‚Äîfrom static ppi networks to dynamic ppi networks.
Briefings in bioinformatics, 15(2):177–194, 2014.

[178] Christoph E Mandl.
Evaluation and optimization of urban public transportation networks.
European Journal of Operational Research, 5(6):396–404, 1980.

[179] Yi-Hwa Wu and Harvey J Miller.
Computational tools for measuring space-time accessibility within dynamic flow transportation networks.
Journal of Transportation and Statistics, 4(2/3):1–14, 2001.

[180] Mark W Bruner, Karl Erickson, Brian Wilson, and Jean Côté.
An appraisal of athlete development models through citation network analysis.
Psychology of sport and exercise, 11(2):133–139, 2010.

[181] Fei Gao and Katarzyna Musial-Gabrys.
Hybrid structure-based link prediction model.
In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1221–1228. IEEE, 2016.

[182] Petter Holme and Jari Saramäki.
Temporal networks.
Physics reports, 519(3):97–125, 2012.

[183] Marc Barthélemy.
Spatial networks.
Physics Reports, 499(1-3):1–101, 2011.

[184] Paolo Crucitti, Vito Latora, and Sergio Porta.
Centrality measures in spatial networks of urban streets.
Physical Review E, 73(3):036125, 2006.

[185] Jerome Buhl, Jacques Gautrais, N Reeves, Ricard V Solé, Sergi Valverde, Pascale Kuntz, and Guy Theraulaz.
Topological patterns in street networks of self-organized urban settlements.
The European Physical Journal B-Condensed Matter and Complex Systems, 49(4):513–522, 2006.

[186] Alessio Cardillo, Salvatore Scellato, Vito Latora, and Sergio Porta.
Structural properties of planar graphs of urban street patterns.
Physical Review E, 73(6):066107, 2006.

[187] Frank Goetzke, Regine Gerike, Antonio Páez, and Elenna Dugundji.
Social interactions in transportation: Analyzing groups and spatial networks.
Transportation, 42(5):723–731, 2015.

[188] Giorgio Topa.
Social and spatial networks in labour markets.
Oxford Review of Economic Policy, 35(4):722–745, 2019.

[189] Monica Stephens and Ate Poorthuis.
Follow thy neighbor: Connecting the social and the spatial networks on twitter.
Computers, Environment and Urban Systems, 53:87–95, 2015.

[190] Yihui Ren, Mária Ercsey-Ravasz, Pu Wang, Marta C González, and Zoltán Toroczkai.
Predicting commuter flows in spatial networks using a radiation model based on temporal ranges.
Nature communications, 5(1):1–9, 2014.

[191] Shuo Shang, Jiajun Liu, Kai Zheng, Hua Lu, Torben Bach Pedersen, and Ji-Rong Wen.
Planning unobstructed paths in traffic-aware spatial networks.
GeoInformatica, 19(4):723–746, 2015.

[192] Ray Solomonoff and Anatol Rapoport.
Connectivity of random nets.
The bulletin of mathematical biophysics, 13(2):107–117, 1951.

[193] Edgar N Gilbert.
Random graphs.
The Annals of Mathematical Statistics, 30(4):1141–1144, 1959.

[194] Deb Verhoeven, Katarzyna Musial, Stuart Palmer, Sarah Taylor, Shaukat Abidi, Vejune Zemaityte, and Lachlan Simpson.
Controlling for openness in the male-dominated collaborative networks of the global film industry.
PloS one, 15(6):e0234460, 2020.

[195] Kaihao Zhang, Yongzhen Huang, Yong Du, and Liang Wang.
Facial expression recognition based on deep evolutional spatial-temporal networks.
IEEE Transactions on Image Processing, 26(9):4193–4203, 2017.

[196] Shuangjie Xu, Yu Cheng, Kang Gu, Yang Yang, Shiyu Chang, and Pan Zhou.
Jointly attentive spatial-temporal pooling networks for video-based person re-identification.
In *Proceedings of the IEEE international conference on computer vision*, pages 4733–4742, 2017.

[197] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han.
Mining hidden community in heterogeneous social networks.
In *Proceedings of the 3rd international workshop on Link discovery*, pages 58–65, 2005.

[198] Wenjun Wang, Pengfei Jiao, Dongxiao He, Di Jin, Lin Pan, and Bogdan Gabrys.
Autonomous overlapping community detection in temporal networks: A dynamic bayesian nonnegative matrix factorization approach.
Knowledge-Based Systems, 110:121–134, 2016.

[199] Junhuan Zhang, Jiaqi Wen, and Jing Chen.
Modeling market fluctuations under investor sentiment with a hawkes-contact process.
The European Journal of Finance, pages 1–16, 2021.

[200] Robert W Krause, Mark Huisman, Christian Steglich, and Tom Snijders.
Missing data in cross-sectional networks–an extensive comparison of missing data treatment methods.
Social Networks, 62:99–112, 2020.

[201] Anna D Broido and Aaron Clauset.
Scale-free networks are rare.
Nature communications, 10(1):1–10, 2019.

[202] David R Hunter, Steven M Goodreau, and Mark S Handcock.
Goodness of fit of social network models.
Journal of the american statistical association, 103(481):248–258, 2008.

[203] Katarzyna Musial, Marcin Budka, and Krzysztof Juszczyszyn.
Creation and growth of online social network.
World Wide Web, 16(4):421–447, 2013.

[204] Antonia Godoy-Lorite and Nick S Jones.
Inference and influence of network structure using snapshot social behavior without network data.
Science Advances, 7(23):eabb8762, 2021.

[205] Junhuan Zhang, Jun Wang, and Jiguang Shao.
Finite-range contact process on the market return intervals distributions.
Advances in Complex Systems, 13(05):643–657, 2010.

[206] Junhuan Zhang and Jun Wang.
Modeling and simulation of the market fluctuations by the finite range contact systems.
Simulation Modelling Practice and Theory, 18(6):910–925, 2010.

[207] Byoung Kyu Choi and Donghun Kang.
Modeling and simulation of discrete event systems.
John Wiley & Sons, 2013.

[208] Augusto Q Macedo, Leandro B Marinho, and Rodrygo LT Santos.
Context-aware event recommendation in event-based social networks.
In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 123–130, 2015.

[209] Tuan-Anh Nguyen Pham, Xutao Li, Gao Cong, and Zhenjie Zhang.
A general graph-based model for recommendation in event-based social networks.
In *2015 IEEE 31st international conference on data engineering*, pages 567–578. IEEE, 2015.

[210] Xingjie Liu, Qi He, Yuanyuan Tian, Wang-Chien Lee, John McPherson, and Jiawei Han.
Event-based social networks: linking the online and offline social worlds.
In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1032–1040, 2012.

[211] Joshua M Epstein.
Generative social science: Studies in agent-based computational modeling, volume 13.
Princeton University Press, 2006.

[212] Amanda L Andrei, Kevin Comer, and Matthew Koehler.
An agent-based model of network effects on tax compliance and evasion.
Journal of Economic Psychology, 40:119–133, 2014.

[213] Srinivasan Venkatramanan, Bryan Lewis, Jiangzhuo Chen, Dave Higdon, Anil Vullikanti, and Madhav Marathe.
Using data-driven agent-based models for forecasting emerging infectious diseases.
Epidemics, 22:43–49, 2018.

[214] Qingqi Long.
Data-driven decision making for supply chain networks with agent-based computational experiment.
Knowledge-Based Systems, 141:55–66, 2018.

[215] Angelia Nedic, Asuman Ozdaglar, and Pablo A Parrilo.
Constrained consensus and optimization in multi-agent networks.
IEEE Transactions on Automatic Control, 55(4):922–938, 2010.

[216] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani.
Activity driven modeling of time varying networks.
Scientific reports, 2(1):1–7, 2012.

[217] Xiangyu Meng and Tongwen Chen.
Event based agreement protocols for multi-agent networks.
Automatica, 49(7):2125–2132, 2013.

[218] Masaki Ogura, Victor M Preciado, and Naoki Masuda.

Optimal containment of epidemics over temporal activity-driven networks.
SIAM Journal on Applied Mathematics, 79(3):986–1006, 2019.

[219] Gerardo Iñiguez, Federico Battiston, and Márton Karsai.
Bridging the gap between graphs and networks.
Communications Physics, 3(1):1–5, 2020.

[220] Tero Aittokallio and Benno Schwikowski.
Graph-based methods for analysing networks in cell biology.
Briefings in bioinformatics, 7(3):243–255, 2006.

[221] Gemini George and Sabu M Thampi.
A graph-based security framework for securing industrial iot networks from vulnerability exploitations.
IEEE Access, 6:43586–43601, 2018.

[222] Leonhard Euler.
Solutio problematis ad geometriam situs pertinentis.
Commentarii academiae scientiarum Petropolitanae, pages 128–140, 1741.

[223] Gerald Alexanderson.
About the cover: Euler and königsberg,Äôs bridges: A historical view.
Bulletin of the american mathematical society, 43(4):567–573, 2006.

[224] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al.
Graph theory with applications, volume 290.
Macmillan London, 1976.

[225] Hristo Djidjev, Gary Sandine, Curtis Storlie, and Scott Vander Wiel.
Graph based statistical analysis of network traffic.
In *Proceedings of the Ninth Workshop on Mining and Learning with Graphs*, 2011.

[226] Takashi Washio and Hiroshi Motoda.
State of the art of graph-based data mining.
Acm Sigkdd Explorations Newsletter, 5(1):59–68, 2003.

[227] Nasrin Hassanlou, Maryam Shoaran, and Alex Thomo.
Probabilistic graph summarization.
In *International Conference on Web-Age Information Management*, pages 545–556. Springer, 2013.

[228] Luis Enrique Sucar.
Probabilistic graphical models.
Advances in Computer Vision and Pattern Recognition. London: Springer London. doi, 10(978):1, 2015.

[229] Yuichiro Anzai.
Pattern recognition and machine learning.
Elsevier, 2012.

[230] John Lafferty, Andrew McCallum, and Fernando CN Pereira.
Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
2001.

[231] Ben Taskar, Pieter Abbeel, Ming-Fai Wong, and Daphne Koller.
Relational markov networks.
Introduction to statistical relational learning, pages 175–200, 2007.

[232] Di Jin, Binbin Zhang, Yue Song, Dongxiao He, Zhiyong Feng, Shizhan Chen, Weihao Li, and Katarzyna Musial.
Modmrf: A modularity-based markov random field method for community detection.
Neurocomputing, 405:218–228, 2020.

[233] Ilya Sutskever and Geoffrey E Hinton.
Deep, narrow sigmoid belief networks are universal approximators.
Neural computation, 20(11):2629–2636, 2008.

[234] Serena H Chen and Carmel A Pollino.
Good practice in bayesian network modelling.
Environmental Modelling & Software, 37:134–145, 2012.

[235] Matthew J Beal, Zoubin Ghahramani, and Carl Edward Rasmussen.
The infinite hidden markov model.
Advances in neural information processing systems, 1:577–584, 2002.

[236] Nico Potyka.
Abstract argumentation with markov networks.
In *ECAI 2020*, pages 865–872. IOS Press, 2020.

[237] David Heckerman.
A tutorial on learning with bayesian networks.
Innovations in Bayesian networks, pages 33–82, 2008.

[238] Chenzhao Li, Sankaran Mahadevan, You Ling, Sergio Choze, and Liping Wang.
Dynamic bayesian network for aircraft wing health monitoring digital twin.
Aiaa Journal, 55(3):930–941, 2017.

[239] Jinsong Yu, Yue Song, Diyin Tang, and Jing Dai.
A digital twin approach based on nonparametric bayesian network for complex system health monitoring.
Journal of Manufacturing Systems, 58:293–304, 2021.

[240] Peter Orbanz and Daniel M Roy.
Bayesian models of graphs, arrays and other exchangeable random structures.
IEEE transactions on pattern analysis and machine intelligence, 37(2):437–461, 2014.

[241] Hichem Boudali and Joanne Bechta Dugan.
A continuous-time bayesian network reliability modeling, and analysis framework.
IEEE transactions on reliability, 55(1):86–97, 2006.

[242] David F Anderson and Thomas G Kurtz.
Continuous time markov chain models for chemical reaction networks.
In *Design and analysis of biomolecular circuits*, pages 3–42. Springer, 2011.

[243] Brendan J Frey and Nebojsa Jojic.
A comparison of algorithms for inference and learning in probabilistic graphical models.
IEEE Transactions on pattern analysis and machine intelligence, 27(9):1392–1416, 2005.

[244] Martyn Plummer et al.
Jags: A program for analysis of bayesian graphical models using gibbs sampling.
In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–10. Vienna, Austria., 2003.

[245] Erik B Sudderth, Alexander T Ihler, Michael Isard, William T Freeman, and Alan S Willsky.

Nonparametric belief propagation.
Communications of the ACM, 53(10):95–103, 2010.

[246] Joao V Graça, Kuzman Ganchev, and Ben Taskar.
Expectation maximization and posterior constraints.
2007.

[247] Erik Kropat, Ayşe Özmen, Gerhard-Wilhelm Weber, Silja Meyer-Nieberg, and Ozlem Defterli.
Fuzzy prediction strategies for gene-environment networks–fuzzy regression analysis for two-modal regulatory systems.
RAIRO-Operations Research, 50(2):413–435, 2016.

[248] Erik Kropat, Gerhard-Wilhelm Weber, and Selma Belen.
Dynamical gene-environment networks under ellipsoidal uncertainty: set-theoretic regression analysis based on ellipsoidal or.
In *Dynamics, games and science I*, pages 545–571. Springer, 2011.

[249] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu.
A survey on network embedding.
IEEE Transactions on Knowledge and Data Engineering, 31(5):833–852, 2018.

[250] Marcin Budka, Krzysztof Juszczyszyn, Katarzyna Musial, and Anna Musial.
Molecular model of dynamic social network based on e-mail communication.
Social Network Analysis and Mining, 3(3):543–563, 2013.

[251] William L Hamilton, Rex Ying, and Jure Leskovec.
Representation learning on graphs: Methods and applications.
arXiv preprint arXiv:1709.05584, 2017.

[252] Ketan Rajawat and Sandeep Kumar.
Stochastic multidimensional scaling.
IEEE Transactions on Signal and Information Processing over Networks, 3(2):360–375, 2017.

[253] Bing Li, Yigang He, Fengming Guo, and Lei Zuo.
A novel localization algorithm based on isomap and partial least squares for wireless sensor networks.
IEEE Transactions on Instrumentation and Measurement, 62(2):304–314, 2012.

[254] Johan Bollen, Herbert Van de Sompel, Aric Hagberg, and Ryan Chute.
A principal component analysis of 39 scientific impact measures.
PloS one, 4(6):e6022, 2009.

[255] Padma Polash Paul, Marina L Gavrilova, and Reda Alhajj.
Decision fusion for multimodal biometrics using social network analysis.
IEEE transactions on systems, man, and cybernetics: systems, 44(11):1522–1533, 2014.

[256] Zhirong Yang, Jaakko Peltonen, and Samuel Kaski.
Optimization equivalence of divergences improves neighbor embedding.
In *International Conference on Machine Learning*, pages 460–468. PMLR, 2014.

[257] Florent Olivon, Nicolas Elie, Gwendal Grelier, Fanny Roussi, Marc Litaudon, and David Touboul.
Metgem software for the generation of molecular networks based on the t-sne algorithm.
Analytical chemistry, 90(23):13900–13908, 2018.

[258] Yaojing Wang, Yuan Yao, Hanghang Tong, Feng Xu, and Jian Lu.
A brief review of network embedding.
Big Data Mining and Analytics, 2(1):35–47, 2020.

[259] David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks.
A closer look at skip-gram modelling.
In *LREC*, volume 6, pages 1222–1225. Citeseer, 2006.

[260] Ajit P Singh and Geoffrey J Gordon.
A unified view of matrix factorization models.
In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 358–373. Springer, 2008.

[261] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang.
Network representation learning: A survey.
IEEE transactions on Big Data, 6(1):3–28, 2018.

[262] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang.
A comprehensive survey of graph embedding: Problems, techniques, and applications.
IEEE Transactions on Knowledge and Data Engineering, 30(9):1616–1637, 2018.

[263] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena.
Deepwalk: Online learning of social representations.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.

[264] Aditya Grover and Jure Leskovec.
node2vec: Scalable feature learning for networks.
In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[265] Pinar Yanardag and SVN Vishwanathan.
Deep graph kernels.
In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.

[266] Yuchong Gong, Yanqing Niu, Wen Zhang, and Xiaohong Li.
A network embedding-based multiple information integration method for the mirna-disease association prediction.
BMC bioinformatics, 20(1):1–13, 2019.

[267] Shaosheng Cao, Wei Lu, and Qiongkai Xu.
Deep neural networks for learning graph representations.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[268] Hongxu Chen, Hongzhi Yin, Xiangguo Sun, Tong Chen, Bogdan Gabrys, and Katarzyna Musial.
Multi-level graph convolutional networks for cross-platform anchor link prediction.
In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1503–1511, 2020.

[269] Jianing Sun, Wei Guo, Dengcheng Zhang, Yingxue Zhang, Florence Regol, Yaochen Hu, Huifeng Guo, Ruiming Tang, Han Yuan, Xiuqiang He, et al.
A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks.
In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2030–2039, 2020.

[270] Meng Qu, Yoshua Bengio, and Jian Tang.
Gmnn: Graph markov neural networks.

In *International conference on machine learning*, pages 5241–5250. PMLR, 2019.

[271] Davide Bacciu, Federico Errica, and Alessio Micheli.
Contextual graph markov model: A deep and generative approach to graph processing.
In *International Conference on Machine Learning*, pages 294–303. PMLR, 2018.

[272] Zijie Huang, Yizhou Sun, and Wei Wang.
Coupled graph ode for learning interacting system dynamics.
In *Proc. of 2021 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'21)*, 2021.

[273] Finale Doshi-Velez and Been Kim.
Towards a rigorous science of interpretable machine learning.
arXiv preprint arXiv:1702.08608, 2017.

[274] Katarzyna Musial, Bogdan Gabrys, and Marcin Buczko.
What kind of network are you? using local and global characteristics in network categorisation tasks.
In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1366–1373, 2013.

[275] Rashad Eletreby, Yong Zhuang, Kathleen M Carley, Osman Yağan, and H Vincent Poor.
The effects of evolutionary adaptations on spreading processes in complex networks.
Proceedings of the National Academy of Sciences, 117(11):5664–5670, 2020.

[276] Mustapha Ait Rami, Vahid S Bokharaie, Oliver Mason, and Fabian R Wirth.
Stability criteria for sis epidemiological models under switching policies.
arXiv preprint arXiv:1306.0135, 2013.

[277] Yupeng Gu, Yizhou Sun, and Jianxi Gao.
The co-evolution model for social network evolving and opinion migration.
In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 175–184, 2017.

[278] Yingkui Wang, Di Jin, Katarzyna Musial, and Jianwu Dang.
Community detection in social networks considering topic correlations.

In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 321–328, 2019.

[279] Mohammad Al Hasan and Mohammed J Zaki.
A survey of link prediction in social networks.
In *Social network data analytics*, pages 243–275. Springer, 2011.

[280] Zhan Bu, Yuyao Wang, Hui-Jia Li, Jiuchuan Jiang, Zhiang Wu, and Jie Cao.
Link prediction in temporal networks: Integrating survival analysis and game theory.
Information Sciences, 498:41–61, 2019.

[281] Albert-Laszlo Barabâsi, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek.
Evolution of the social network of scientific collaborations.
Physica A: Statistical mechanics and its applications, 311(3-4):590–614, 2002.

[282] Justin Basilico and Thomas Hofmann.
Unifying collaborative and content-based filtering.
In *Proceedings of the twenty-first international conference on Machine learning*, page 9, 2004.

[283] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki.
Link prediction using supervised learning.
In *SDM06: workshop on link analysis, counter-terrorism and security*, volume 30, pages 798–805, 2006.

[284] Glen Jeh and Jennifer Widom.
Simrank: a measure of structural-context similarity.
In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 538–543, 2002.

[285] Richang Hong, Yuan He, Le Wu, Yong Ge, and Xindong Wu.
Deep attributed network embedding by preserving structure and attribute information.
IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019.

[286] Ben Taskar, Ming-Fai Wong, Pieter Abbeel, and Daphne Koller.
Link prediction in relational data.
Advances in neural information processing systems, 16:659–666, 2003.

[287] Caiyan Dai, Ling Chen, Bin Li, and Yun Li.
Link prediction in multi-relational networks based on relational similarity.
Information Sciences, 394:198–216, 2017.

[288] Hongchang Gao and Heng Huang.
Deep attributed network embedding.
In *Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[289] Xueyan Liu, Bo Yang, Wenzhuo Song, Katarzyna Musial, Wanli Zuo, Hongxu Chen, and Hongzhi Yin.
A block-based generative model for attributed network embedding.
World Wide Web, 24(5):1439–1464, 2021.

[290] Benjamin Taskar, Eran Segal, and Daphne Koller.
Probabilistic classification and clustering in relational data.
In *International joint conference on artificial intelligence*, volume 17, pages 870–878. Lawrence Erlbaum Associates LTD, 2001.

[291] Jon Kleinberg and Eva Tardos.
Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields.
Journal of the ACM (JACM), 49(5):616–639, 2002.

[292] Frank McSherry.
Spectral partitioning of random graphs.
In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.

[293] Smriti Bhagat, Graham Cormode, and S Muthukrishnan.
Node classification in social networks.
In *Social network data analytics*, pages 115–148. Springer, 2011.

[294] Yang Zhou, Hong Cheng, and Jeffrey Xu Yu.
Graph clustering based on structural/attribute similarities.
Proceedings of the VLDB Endowment, 2(1):718–729, 2009.

[295] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty.
Semi-supervised learning using gaussian fields and harmonic functions.

In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.

[296] Sami Abu-El-Haija, Amol Kapoor, Bryan Perozzi, and Joonseok Lee.
N-gcn: Multi-scale graph convolution for semi-supervised node classification.
In *uncertainty in artificial intelligence*, pages 841–851. PMLR, 2020.

[297] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy.
Local probabilistic models for link prediction.
In *Seventh IEEE international conference on data mining (ICDM 2007)*, pages 322–331. IEEE, 2007.

[298] Xiaoyi Li, Nan Du, Hui Li, Kang Li, Jing Gao, and Aidong Zhang.
A deep learning approach to link prediction in dynamic networks.
In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 289–297. SIAM, 2014.

[299] Jinyin Chen, Xueke Wang, and Xuanheng Xu.
Gc-lstm: graph convolution embedded lstm for dynamic network link prediction.
Applied Intelligence, pages 1–16, 2021.

[300] Xiaoke Ma, Penggang Sun, and Yu Wang.
Graph regularized nonnegative matrix factorization for temporal link prediction in dynamic networks.
Physica A: Statistical mechanics and its applications, 496:121–136, 2018.

[301] Purnamrita Sarkar, Deepayan Chakrabarti, and Michael Jordan.
Nonparametric link prediction in large scale dynamic networks.
Electronic Journal of Statistics, 8(2):2022–2065, 2014.

[302] Huiyuan Chen and Jing Li.
Exploiting structural and temporal evolution in dynamic link prediction.
In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 427–436, 2018.

[303] Fragkiskos D Malliaros and Michalis Vazirgiannis.
Clustering and community detection in directed networks: A survey.
Physics reports, 533(4):95–142, 2013.

[304] Michele Coscia.
Discovering communities of community discovery.
In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1–8, 2019.

[305] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu.
A comprehensive survey on graph anomaly detection with deep learning.
IEEE Transactions on Knowledge and Data Engineering, 2021.

[306] Leman Akoglu, Hanghang Tong, and Danai Koutra.
Graph based anomaly detection and description: a survey.
Data mining and knowledge discovery, 29(3):626–688, 2015.

[307] Sebastian Funk, Marcel Salathé, and Vincent AA Jansen.
Modelling the influence of human behaviour on the spread of infectious diseases: a review.
Journal of the Royal Society Interface, 7(50):1247–1256, 2010.

[308] Hanlin Sun, David Saad, and Andrey Y Lokhov.
Competition, collaboration, and optimization in multiple interacting spreading processes.
Physical Review X, 11(1):011048, 2021.

[309] I Ameen, Dumitru Baleanu, and Hegagi Mohamed Ali.
An efficient algorithm for solving the fractional optimal control of sirv epidemic model with a combination of vaccination and treatment.
Chaos, Solitons & Fractals, 137:109892, 2020.

[310] Faryad Darabi Sahneh and Caterina Scoglio.
Competitive epidemic spreading over arbitrary multilayer networks.
Physical Review E, 89(6):062817, 2014.

[311] Ian Cooper, Argha Mondal, and Chris G Antonopoulos.
A sir model assumption for the spread of covid-19 in different communities.
Chaos, Solitons & Fractals, 139:110057, 2020.

[312] Side Syafruddin and MSM Noorani.

A sir model for spread of dengue fever disease (simulation for south sulawesi, indonesia and selangor, malaysia).
World Journal of Modelling and Simulation, 9(2):96–105, 2013.

[313] Pascal P Klamser, Marc Wiedermann, Jonathan F Donges, and Reik V Donner.
Zealotry effects on opinion dynamics in the adaptive voter model.
Physical Review E, 96(5):052315, 2017.

[314] Tong Xu, Hengshu Zhu, Hao Zhong, Guannan Liu, Hui Xiong, and Enhong Chen.
Exploiting the dynamic mutual influence for predicting social event participation.
IEEE Transactions on Knowledge and Data Engineering, 31(6):1122–1135, 2018.

[315] Celia Laur, Jack Bell, Renata Valaitis, Sumantra Ray, and Heather Keller.
The sustain and spread framework: strategies for sustaining and spreading nutrition care improvements in acute care based on thematic analysis from the more-2-eat study.
BMC health services research, 18(1):1–11, 2018.

[316] Amy E Kendig, Elizabeth T Borer, Emily N Boak, Tashina C Picard, and Eric W Seabloom.
Host nutrition mediates interactions between plant viruses, altering transmission and predicted disease spread.
Ecology, 101(11):e03155, 2020.

[317] Panagiotis D Karampourniotis, Sameet Sreenivasan, Boleslaw K Szymanski, and Gyorgy Korniss.
The impact of heterogeneous thresholds on social contagion with multiple initiators.
PloS one, 10(11):e0143020, 2015.

[318] Istvan Z Kiss, Gergely Röst, and Zsolt Vizi.
Generalization of pairwise models to non-markovian epidemics on networks.
Physical review letters, 115(7):078701, 2015.

[319] Qiang Liu and Piet Van Mieghem.
Burst of virus infection and a possibly largest epidemic threshold of non-markovian susceptible-infected-susceptible processes on networks.
Physical Review E, 97(2):022309, 2018.

[320] Arturo Melo, Carolyn L Beck, José Ismael Peña, and Philip E Paré.
Knowledge transfer from universities to regions as a network spreading process.
In *2018 IEEE International Systems Engineering Symposium (ISSE)*, pages 1–8. IEEE, 2018.

[321] Piet Van Mieghem.
The n-intertwined sis epidemic network model.
Computing, 93(2-4):147–169, 2011.

[322] Hyoung Jun Ahn and Babak Hassibi.
Global dynamics of epidemic spread over complex networks.
In *52nd IEEE Conference on Decision and Control*, pages 4579–4585. IEEE, 2013.

[323] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora.
Simplicial models of social contagion.
Nature communications, 10(1):1–9, 2019.

[324] Thilo Gross and Bernd Blasius.
Adaptive coevolutionary networks: a review.
Journal of the Royal Society Interface, 5(20):259–271, 2008.

[325] Jack Leitch, Kathleen A Alexander, and Srijan Sengupta.
Toward epidemic thresholds on temporal networks: a review and open questions.
Applied Network Science, 4(1):1–21, 2019.

[326] Istvan Z Kiss, Luc Berthouze, Timothy J Taylor, and Péter L Simon.
Modelling approaches for simple dynamic networks and applications to disease transmission models.
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 468(2141):1332–1355, 2012.

[327] Michael Taylor, Timothy J Taylor, and Istvan Z Kiss.
Epidemic threshold and control in a dynamic network.
Physical Review E, 85(1):016103, 2012.

[328] Federico Baldo, Lorenzo Dall'Olio, Mattia Ceccarelli, Riccardo Scheda, Michele Lombardi, Andrea Borghesi, Stefano Diciotti, and Michela Milano.
Deep learning for virus-spreading forecasting: a brief survey.
arXiv preprint arXiv:2103.02346, 2021.

[329] Brian Karrer and Mark EJ Newman.
Competing epidemics on complex networks.
Physical Review E, 84(3):036106, 2011.

[330] Mark EJ Newman.
Threshold effects for two pathogens spreading on a network.
Physical review letters, 95(10):108701, 2005.

[331] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi.
Limiting the spread of misinformation in social networks.
In *Proceedings of the 20th international conference on World wide web*, pages 665–674, 2011.

[332] B Aditya Prakash, Alex Beutel, Roni Rosenfeld, and Christos Faloutsos.
Winner takes all: competing viruses or ideas on fair-play networks.
In *Proceedings of the 21st international conference on World Wide Web*, pages 1037–1046, 2012.

[333] Byungjoon Min and Maxi San Miguel.
Competing contagion processes: Complex contagion triggered by simple contagion.
Scientific reports, 8(1):1–8, 2018.

[334] Guiyuan Fu, Feier Chen, Jianguo Liu, and Jingti Han.
Analysis of competitive information diffusion in a group-based population over social networks.
Physica A: Statistical Mechanics and its Applications, 525:409–419, 2019.

[335] Irem Koprulu, Yoora Kim, and Ness B Shroff.
Battle of opinions over evolving social networks.
IEEE/ACM Transactions on Networking, 27(2):532–545, 2019.

[336] Cheng Pan, Lu-Xing Yang, Xiaofan Yang, Yingbo Wu, and Yuan Yan Tang.
An effective rumor-containing strategy.
Physica A: Statistical Mechanics and its Applications, 500:80–91, 2018.

[337] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen.
The spread of awareness and its impact on epidemic outbreaks.
Proceedings of the National Academy of Sciences, 106(16):6872–6877, 2009.

[338] Xiu-Xiu Zhan, Chuang Liu, Ge Zhou, Zi-Ke Zhang, Gui-Quan Sun, Jonathan JH Zhu, and Zhen Jin.
Coupling dynamics of epidemic spreading and information diffusion on complex networks.
Applied Mathematics and Computation, 332:437–448, 2018.

[339] Giacomo Baggio, Danielle S Bassett, and Fabio Pasqualetti.
Data-driven control of complex networks.
Nature communications, 12(1):1–13, 2021.

[340] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási.
Controllability of complex networks.
nature, 473(7346):167–173, 2011.

[341] Zhengzhong Yuan, Chen Zhao, Zengru Di, Wen-Xu Wang, and Ying-Cheng Lai.
Exact controllability of complex networks.
Nature communications, 4(1):1–9, 2013.

[342] Yang Lou, Lin Wang, Kim-Fung Tsang, and Guanrong Chen.
Towards optimal robustness of network controllability: An empirical necessary condition.
IEEE Transactions on Circuits and Systems I: Regular Papers, 67(9):3163–3174, 2020.

[343] Yang Lou, Yaodong He, Lin Wang, and Guanrong Chen.
Predicting network controllability robustness: A convolutional neural network approach.
IEEE Transactions on Cybernetics, 2020.

[344] Bahare Kiumarsi, Kyriakos G Vamvoudakis, Hamidreza Modares, and Frank L Lewis.
Optimal and autonomous control using reinforcement learning: A survey.
IEEE transactions on neural networks and learning systems, 29(6):2042–2062, 2017.

[345] Florian Dörfler and Francesco Bullo.
Synchronization in complex networks of phase oscillators: A survey.
Automatica, 50(6):1539–1564, 2014.

[346] Yang Tang, Feng Qian, Huijun Gao, and Jürgen Kurths.
Synchronization in complex networks and its application–a survey of recent advances and challenges.
Annual Reviews in Control, 38(2):184–198, 2014.

[347] Davide Fiore, Giovanni Russo, and Mario Di Bernardo.
Exploiting nodes symmetries to control synchronization and consensus patterns in multiagent systems.
IEEE control systems letters, 1(2):364–369, 2017.

[348] Jiejie Chen, Boshan Chen, and Zhigang Zeng.
Synchronization and consensus in networks of linear fractional-order multi-agent systems via sampled-data control.
IEEE transactions on neural networks and learning systems, 31(8):2955–2964, 2019.

[349] Wenlian Lu and Tianping Chen.
Quad-condition, synchronization, consensus of multiagents, and anti-synchronization of complex networks.
IEEE transactions on cybernetics, 2019.

[350] Eli Meirom, Haggai Maron, Shie Mannor, and Gal Chechik.
Controlling graph dynamics with reinforcement learning and graph neural networks.
In *International Conference on Machine Learning*, pages 7565–7577. PMLR, 2021.

[351] Dimitris Kiritsis.
Closed-loop plm for intelligent products in the era of the internet of things.
Computer-Aided Design, 43(5):479–501, 2011.

[352] Michael G Kapteyn, Jacob VR Pretorius, and Karen E Willcox.
A probabilistic graphical model foundation for enabling predictive digital twins at scale.
Nature Computational Science, 1(5):337–347, 2021.

[353] Evgeny Kharlamov, Francisco Martin-Recuerda, Brandon Perry, David Cameron, Roar Fjellheim, and Arild Waaler.
Towards semantically enhanced digital twins.

In *2018 IEEE International Conference on Big Data (Big Data)*, pages 4189–4193. IEEE, 2018.

[354] Agniva Banerjee, Raka Dalal, Sudip Mittal, and Karuna Pande Joshi.
Generating digital twin models using knowledge graphs for industrial production lines.
UMBC Information Systems Department, 2017.

[355] Thomas Hubauer, Steffen Lamparter, Peter Haase, and Daniel Markus Herzig.
Use cases of the industrial knowledge graph at siemens.
In *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.

[356] Jinzhi Lu, Xiaochen Zheng, Ali Gharaei, Kostas Kalaboukas, and Dimitris Kiritsis.
Cognitive twins for supporting decision-makings of internet of things systems.
In *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing*, pages 105–115. Springer, 2020.

[357] Chao Fan, Cheng Zhang, Alex Yahja, and Ali Mostafavi.
Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management.
International Journal of Information Management, 56:102049, 2021.

[358] Claudio Mandolla, Antonio Messeni Petruzzelli, Gianluca Percoco, and Andrea Urbinati.
Building a digital twin for additive manufacturing through the exploitation of blockchain: A case analysis of the aircraft industry.
Computers in Industry, 109:134–152, 2019.

[359] Hong-Ning Dai, Zibin Zheng, and Yan Zhang.
Blockchain for internet of things: A survey.
IEEE Internet of Things Journal, 6(5):8076–8094, 2019.

[360] Abir EL Azzaoui, Tae Woo Kim, Vincenzo Loia, and Jong Hyuk Park.
Blockchain-based secure digital twin framework for smart healthy city.
In *Advanced Multimedia and Ubiquitous Engineering*, pages 107–113. Springer, 2021.

[361] Junjie Pang, Yan Huang, Zhenzhen Xie, Jianbo Li, and Zhipeng Cai.

Collaborative city digital twin for the covid-19 pandemic: A federated learning solution.
Tsinghua Science and Technology, 26(5):759–771, 2021.

[362] Chen Chen, Jingrui He, Nadya Bliss, and Hanghang Tong.
Towards optimal connectivity on multi-layered networks.
IEEE transactions on knowledge and data engineering, 29(10):2332–2346, 2017.

[363] Constantin Cronrath, Abolfazl R Aderiani, and Bengt Lennartson.
Enhancing digital twins through reinforcement learning.
In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pages 293–298. IEEE, 2019.

[364] Hao Zhang, Qiang Liu, Xin Chen, Ding Zhang, and Jiewu Leng.
A digital twin-based approach for designing and multi-objective optimization of hollow glass production line.
Ieee Access, 5:26901–26911, 2017.

[365] Sajjad Rahmanzadeh, Mir Saman Pishvaee, and Kannan Govindan.
Emergence of open supply chain management: the role of open innovation in the future smart industry using digital twin network.
Annals of Operations Research, pages 1–29, 2022.

[366] Louise Wright and Stuart Davidson.
How to tell the difference between a model and a digital twin.
Advanced Modeling and Simulation in Engineering Sciences, 7(1):1–13, 2020.

[367] Lin Zhang, Longfei Zhou, and Berthold KP Horn.
Building a right digital twin with model engineering.
Journal of Manufacturing Systems, 59:151–164, 2021.

[368] Hatem Ben Sta.
Quality and the efficiency of data in ‚Äúsmart-cities‚Äù.
Future Generation Computer Systems, 74:409–416, 2017.

[369] J Kent Crawford.
Project management maturity model.
Auerbach Publications, 2006.

[370] Peichao Zhang, Chao Huo, and Mladen Kezunovic.
A novel measure of component importance considering cost for all-digital protection systems.
In *2007 IEEE Power Engineering Society General Meeting*, pages 1–6. IEEE, 2007.

[371] Jinjiang Wang, Lunkuan Ye, Robert X Gao, Chen Li, and Laibin Zhang.
Digital twin for rotating machinery fault diagnosis in smart manufacturing.
International Journal of Production Research, 57(12):3920–3934, 2019.

[372] Yuxiao Dong, Jing Zhang, Jie Tang, Nitesh V Chawla, and Bai Wang.
Coupledlp: Link prediction in coupled networks.
In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 199–208, 2015.

[373] Ruthwik R Junuthula, Kevin S Xu, and Vijay K Devabhaktuni.
Evaluating link prediction accuracy in dynamic networks with added and removed edges.
In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 377–384. IEEE, 2016.

[374] Rushabh Patel, Yanhui Guo, Adi Alhudhaif, Fayadh Alenezi, Sara A Althubiti, Kemal Polat, et al.
Graph-based link prediction between human phenotypes and genes.
Mathematical Problems in Engineering, 2022, 2021.

[375] George Cybenko.
Parallel computing for machine learning in social network analysis.
In *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 1464–1471. IEEE, 2017.

[376] Ruizhe Yang, F Richard Yu, Pengbo Si, Zhaoxin Yang, and Yanhua Zhang.
Integrated blockchain and edge computing systems: A survey, some research issues and challenges.
IEEE Communications Surveys & Tutorials, 21(2):1508–1532, 2019.

[377] Jin Sun, Lu Yin, Minhui Zou, Yi Zhang, Tianqi Zhang, and Junlong Zhou.

Makespan-minimization workflow scheduling for complex networks with social groups in edge computing.
Journal of Systems Architecture, 108:101799, 2020.

[378] M Ahuja, Randeep Kaur, and Dinesh Kumar.
Trend towards the use of complex networks in cloud computing environment.
Int J Hybrid Inf Technol, 8(3):297–306, 2015.

[379] Yan Zhang, Yu Guo, Zhang Zhang, Mengyuan Chen, Shuo Wang, and Jiang Zhang.
Automated discovery of interactions and dynamics for large networked dynamical systems.
arXiv preprint arXiv:2101.00179, 2021.

[380] Zheng Yang, Longfei Shangguan, Weixi Gu, Zimu Zhou, Chenshu Wu, and Yunhao Liu.
Sherlock: Micro-environment sensing for smartphones.
IEEE Transactions on Parallel and Distributed Systems, 25(12):3295–3305, 2014.

[381] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao.
Federated learning in mobile edge networks: A comprehensive survey.
IEEE Communications Surveys & Tutorials, 22(3):2031–2063, 2020.

[382] Mehdi Salehi Heydar Abad, Emre Ozfatura, Deniz Gunduz, and Ozgur Ercetin.
Hierarchical federated learning across heterogeneous cellular networks.
In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8866–8870. IEEE, 2020.

[383] Yunlong Lu, Xiaohong Huang, Yueyue Dai, Sabita Maharjan, and Yan Zhang.
Blockchain and federated learning for privacy-preserved data sharing in industrial iot.
IEEE Transactions on Industrial Informatics, 16(6):4177–4186, 2019.

[384] Yunlong Lu, Xiaohong Huang, Ke Zhang, Sabita Maharjan, and Yan Zhang.
Low-latency federated learning and blockchain for edge association in digital twin empowered 6g networks.
IEEE Transactions on Industrial Informatics, 2020.

[385] Hamdi Kavak, Joon-Seok Kim, Andrew Crooks, Dieter Pfoser, Carola Wenk, and Andreas Züfle.

Location-based social simulation.
In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pages 218–221, 2019.

[386] Katarzyna Musial, Krzysztof Juszczyszyn, and Marcin Budka.
Triad transition probabilities characterize complex networks.
Awareness Magazine, 2012.

[387] Peter Robin Hiesinger.
The Self-Assembling Brain: How Neural Networks Grow Smarter.
Princeton University Press, 2021.

[388] Viviana Amati, Alessandro Lomi, and Antonietta Mira.
Social network modeling.
Annual Review of Statistics and Its Application, 5:343–369, 2018.

[389] Rasha Elhesha, Aisharjya Sarkar, Pietro Cinaglia, Christina Boucher, and Tamer Kahveci.
Co-evolving patterns in temporal networks of varying evolution.
In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 494–503, 2019.

[390] Gunnar Brinkmann.
Fast generation of cubic graphs.
Journal of Graph Theory, 23(2):139–149, 1996.

[391] Markus Meringer.
Fast generation of regular graphs and construction of cages.
Journal of Graph Theory, 30(2):137–146, 1999.

[392] Ralf Gugisch, Adalbert Kerber, Axel Kohnert, Reinhard Laue, Markus Meringer, Christoph Rücker, and Alfred Wassermann.
Molgen 5.0, a molecular structure generator.
In *Advances in mathematical chemistry and applications*, pages 113–138. Elsevier, 2015.

[393] Ingo Althöfer, Jan Kristian Haugland, Karl Scherer, Frank Schneider, and Nicolas Van Cleemput.
Alternating plane graphs.
ARS MATHEMATICA CONTEMPORANEA, 8(2):337–363, 2015.

[394] Átila A Jones, Fábio Protti, and Renata R Del-Vecchio.
Cograph generation with linear delay.
Theoretical Computer Science, 713:1–10, 2018.

[395] Jan Goedgebeur and Oliver Schaudt.
Exhaustive generation of k-critical-free graphs.
Journal of Graph Theory, 87(2):188–207, 2018.

[396] Jan Goedgebeur and Brendan D McKay.
Recursive generation of ipr fullerenes.
Journal of Mathematical Chemistry, 53(8):1702–1724, 2015.

[397] Igor Fabrici, Tomáš Madaras, Mária Timková, Nico Van Cleemput, and Carol T Zamfirescu.
Non-hamiltonian graphs in which every edge-contracted subgraph is hamiltonian.
Applied Mathematics and Computation, 392:125714, 2021.

[398] Jan Goedgebeur, Barbara Meersman, and Carol Zamfirescu.
Graphs with few hamiltonian cycles.
Mathematics of Computation, 89(322):965–991, 2020.

[399] Katherine A Seaton and Lisa M Hackett.
Stations, trains and small-world networks.
Physica A: Statistical Mechanics and its Applications, 339(3-4):635–644, 2004.

[400] Gunnar Brinkmann and Jan Goedgebeur.
Generation of cubic graphs and snarks with large girth.
Journal of Graph Theory, 86(2):255–272, 2017.

[401] Carol T Zamfirescu and Jan Goedgebeur.
On almost hypohamiltonian graphs.
Discrete Mathematics & Theoretical Computer Science, 21, 2019.

[402] Brendan D McKay and Adolfo Piperno.
Practical graph isomorphism, ii.
Journal of symbolic computation, 60:94–112, 2014.

[403] Gunnar Brinkmann, Brendan D McKay, et al.
Fast generation of planar graphs.
MATCH Commun. Math. Comput. Chem, 58(2):323–357, 2007.

[404] Andrea Lancichinetti and Santo Fortunato.
Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities.
Physical Review E, 80(1):016118, 2009.

[405] Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L Tseng.
Facetnet: a framework for analyzing communities and their evolutions in dynamic networks.
In *Proceedings of the 17th international conference on World Wide Web*, pages 685–694, 2008.

[406] Petar Jovanovski, Igor Tomovski, and Ljupco Kocarev.
Modeling the spread of multiple contagions on multilayer networks.
Physica A: Statistical Mechanics and its Applications, 563:125410, 2021.

[407] Vincenza Carchiolo, Alessandro Longheu, Michele Malgeri, Giuseppe Mangioni, and Marialaura Previti.
Mutual influence of users credibility and news spreading in online social networks.
Future Internet, 13(5):107, 2021.

[408] Joon-Seok Kim, Hamdi Kavak, Umar Manzoor, and Andreas Züfle.
Advancing simulation experimentation capabilities with runtime interventions.
In *2019 Spring Simulation Conference (SpringSim)*, pages 1–11. IEEE, 2019.

[409] Marialisa Scatà, Alessandro Di Stefano, Pietro Liò, and Aurelio La Corte.
The impact of heterogeneity and awareness in modeling epidemic spreading on multiplex networks.
Scientific reports, 6(1):1–13, 2016.

[410] Peican Zhu, Xinyu Wang, Shudong Li, Yangming Guo, and Zhen Wang.
Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties.
Applied mathematics and computation, 359:512–524, 2019.

[411] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon.
Superfamilies of evolved and designed networks.
Science, 303(5663):1538–1542, 2004.

[412] Stephen Della Pietra, Vincent Della Pietra, and John Lafferty.
Inducing features of random fields.
IEEE transactions on pattern analysis and machine intelligence, 19(4):380–393, 1997.

[413] Kalyanmoy Deb.
A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-2.
IEEE Trans. Evol. Comput., 6(2):182–197, 2002.

[414] Annu Lambora, Kunal Gupta, and Kriti Chopra.
Genetic algorithm-a literature review.
In *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 380–384. IEEE, 2019.

[415] Andrzej P Wierzbicki.
The use of reference objectives in multiobjective optimization.
In *Multiple Criteria Decision Making Theory and Application: Proceedings of the Third Conference Hagen/Königswinter, West Germany, August 20–24, 1979*, pages 468–486. Springer, 1980.

[416] Wayne W Zachary.
An information flow model for conflict and fission in small groups.
Journal of anthropological research, 33(4):452–473, 1977.

[417] Lekui Zhou, Yang Yang, Xiang Ren, Fei Wu, and Yueting Zhuang.
Dynamic network embedding by modeling triadic closure process.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[418] Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua.
Attributed social network embedding.
IEEE Transactions on Knowledge and Data Engineering, 30(12):2257–2270, 2018.

[419] Abida Sadaf, Luke Mathieson, Piotr Bródka, and Katarzyna Musial.
Maximising influence spread in complex networks by utilising community-based driver nodes as seeds.
In *Annual International Conference on Information Management and Big Data*, pages 126–141. Springer, 2022.

[420] Fabian Lorig, Emil Johansson, and Paul Davidsson.
Agent-based social simulation of the covid-19 pandemic: A systematic review.
Journal of Artificial Societies and Social Simulation, 24(3), 2021.

[421] Cesar Henrique Comin and Luciano da Fontoura Costa.
Identifying the starting point of a spreading process in complex networks.
Physical Review E, 84(5):056105, 2011.

[422] Artur Karczmarczyk, Kamil Bortko, Piotr Bartków, Patryk Pazura, and Jaroslaw Jankowski.
Influencing information spreading processes in complex networks with probability spraying.
In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1038–1046. IEEE, 2018.

[423] Anne Chao, Nicholas J Gotelli, TC Hsieh, Elizabeth L Sander, KH Ma, Robert K Colwell, and Aaron M Ellison.
Rarefaction and extrapolation with hill numbers: a framework for sampling and estimation in species diversity studies.
Ecological monographs, 84(1):45–67, 2014.

[424] M Faisal, EM Zamzami, et al.
Comparative analysis of inter-centroid k-means performance using euclidean distance, canberra distance and manhattan distance.
In *Journal of Physics: Conference Series*, volume 1566, page 012112. IOP Publishing, 2020.

[425] Danai Koutra, Joshua T Vogelstein, and Christos Faloutsos.
Deltacon: A principled massive-graph similarity function.
In *Proceedings of the 2013 SIAM international conference on data mining*, pages 162–170. SIAM, 2013.

[426] Giannis Nikolentzos, Polykarpos Meladianos, and Michalis Vazirgiannis.
Matching node embeddings for graph similarity.
In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[427] Teng Wang, Guoliang Lu, Jie Liu, and Peng Yan.
Graph-based change detection for condition monitoring of rotating machines: Techniques for graph similarity.

IEEE Transactions on Reliability, 68(3):1034–1049, 2018.

[428] Yu Wang, Michael Conrad Meyer, and Junbo Wang.
Real-time delay minimization for data processing in wirelessly networked disaster areas.
IEEE Access, 7:2928–2937, 2018.

[429] Qawi K Telesford, Karen E Joyce, Satoru Hayasaka, Jonathan H Burdette, and Paul J Laurienti.
The ubiquity of small-world networks.
Brain connectivity, 1(5):367–375, 2011.

[430] Thomas Courtat, Catherine Gloaguen, and Stephane Douady.
Mathematics and morphogenesis of cities: A geometrical approach.
Physical Review E, 83(3):036106, 2011.

[431] Peter Ivie and Douglas Thain.
Reproducibility in scientific computing.
ACM Computing Surveys (CSUR), 51(3):1–36, 2018.

[432] Xiaoyan Qiu, Laijun Zhao, Jiajia Wang, Xiaoli Wang, and Qin Wang.
Effects of time-dependent diffusion behaviors on the rumor spreading in social networks.
Physics Letters A, 380(24):2054–2063, 2016.

[433] Timothy W Russell, Joel Hellewell, Christopher I Jarvis, Kevin Van Zandvoort, Sam Abbott, Ruwan Ratnayake, Stefan Flasche, Rosalind M Eggo, W John Edmunds, Adam J Kucharski, et al.
Estimating the infection and case fatality ratio for coronavirus disease (covid-19) using age-adjusted data from the outbreak on the diamond princess cruise ship, february 2020.
Eurosurveillance, 25(12):2000256, 2020.

[434] Anne Chao, Chun-Huo Chiu, and Lou Jost.
Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers.
Annual review of ecology, evolution, and systematics, 45:297–324, 2014.

[435] Antton Alberdi and M Thomas P Gilbert.

A guide to the application of hill numbers to dna-based diversity analyses.
Molecular Ecology Resources, 19(4):804–817, 2019.

[436] Maziar Nekovee, Yamir Moreno, Ginestra Bianconi, and Matteo Marsili.
Theory of rumour spreading in complex social networks.
Physica A: Statistical Mechanics and its Applications, 374(1):457–470, 2007.

[437] Chun-Hsien Li, Chiung-Chiou Tsai, and Suh-Yuh Yang.
Analysis of epidemic spreading of an sirs model in complex heterogeneous networks.
Communications in Nonlinear Science and Numerical Simulation, 19(4):1042–1054, 2014.

[438] Clara Stegehuis, Remco Van Der Hofstad, and Johan SH Van Leeuwaarden.
Epidemic spreading on complex networks with community structures.
Scientific reports, 6(1):1–7, 2016.

[439] Ye Sun, Chuang Liu, Chu-Xu Zhang, and Zi-Ke Zhang.
Epidemic spreading on weighted complex networks.
Physics Letters A, 378(7-8):635–640, 2014.

[440] Alrik Thiem.
Membership function sensitivity of descriptive statistics in fuzzy-set relations.
International Journal of Social Research Methodology, 17(6):625–642, 2014.

[441] Omar Adil M Ali, Aous Y Ali, and Balasem Salem Sumait.
Comparison between the effects of different types of membership functions on fuzzy logic controller performance.
International Journal, 76:76–83, 2015.

[442] Keon-Myung Lee, KYUNG-MI LEE, and KRZYSZTOF J CIOS.
Comparison of interval-valued fuzzy sets, intuitionistic fuzzy sets, and bipolar-valued fuzzy sets.
In *Computing and Information Technologies: Exploring Emerging Technologies*, pages 433–439. World Scientific, 2001.

[443] Mehmet Kaya and Reda Alhajj.
Utilizing genetic algorithms to optimize membership functions for fuzzy weighted association rules mining.
Applied Intelligence, 24(1):7–15, 2006.

[444] Marylu L Lagunes, Oscar Castillo, and Jose Soria.
Optimization of membership function parameters for fuzzy controllers of an autonomous mobile robot using the firefly algorithm.
In *Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications*, pages 199–206. Springer, 2018.

[445] Timothy C Havens, James C Bezdek, Christopher Leckie, Kotagiri Ramamohanarao, and Marimuthu Palaniswami.
A soft modularity function for detecting fuzzy communities in social networks.
IEEE Transactions on Fuzzy Systems, 21(6):1170–1175, 2013.

[446] Jian Wu and Francisco Chiclana.
A social network analysis trust–consensus based approach to group decision-making problems with interval-valued fuzzy reciprocal preference relations.
Knowledge-Based Systems, 59:97–107, 2014.

[447] Guangquan Xu, Bingyan Liu, Litao Jiao, Xiaotong Li, Meiqi Feng, Kaitai Liang, Lei Ma, and Xi Zheng.
Trust2privacy: a novel fuzzy trust-to-privacy mechanism for mobile social networks.
IEEE Wireless Communications, 27(3):72–78, 2020.

[448] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al.
Social contacts and mixing patterns relevant to the spread of infectious diseases.
PLoS medicine, 5(3):e74, 2008.

[449] Lander Willem, Thang Van Hoang, Sebastian Funk, Pietro Coletti, Philippe Beutels, and Niel Hens.
Socrates: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for covid-19.
BMC Research Notes, 13(1):1–8, 2020.

[450] Amin Haghnegahdar and Saman Razavi.
Insights into sensitivity analysis of earth and environmental systems models: On the impact of parameter perturbation scale.
Environmental Modelling & Software, 95:115–131, 2017.

[451] Saman Razavi and Hoshin V Gupta.
A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. theory.
Water Resources Research, 52(1):423–439, 2016.

[452] Bruno Sudret.
Global sensitivity analysis using polynomial chaos expansions.
Reliability engineering & system safety, 93(7):964–979, 2008.

[453] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola.
Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index.
Computer physics communications, 181(2):259–270, 2010.

[454] Fernando Nogueira.
Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–.

[455] Hsin Hsin Chang and Shuang-Shii Chuang.
Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator.
Information & management, 48(1):9–18, 2011.

[456] Nan Lin.
Building a network theory of social capital.
Social capital, pages 3–28, 2017.

[457] Henrik Walter, Birgit Abler, Angela Ciaramidaro, and Susanne Erk.
Motivating forces of human actions: Neuroimaging reward and social interaction.
Brain research bulletin, 67(5):368–381, 2005.

[458] Sören Krach, Frieder M Paulus, Maren Bodden, and Tilo Kircher.
The rewarding nature of social interactions.
Frontiers in behavioral neuroscience, 4:1141, 2010.