

C02029: Doctor of Philosophy

CRICOS Code: 009469A

Subject Code: 32903

October 2021

*Quantifying COVID-19:
Modeling and Evaluation*

Qing Liu

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Qing Liu* 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 Information Technology* 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.

Production Note:

SIGNATURE: Signature removed prior to publication.

Qing Liu

DATE: 25th October, 2021

PLACE: Sydney, Australia

ABSTRACT

The coronavirus disease 2019 (COVID-19) has evolved to a global pandemic and poses significant demands and challenges in modeling its complex epidemic transmission, infection, and contagion. Moreover, it has shown to be vastly different from known epidemics. To address the COVID-19 pandemic, significant efforts have been made to model COVID-19 transmission, diagnoses, interventions, and pathological and influence analysis, etc. However, due to the unique and unknown problem and data complexity, the related studies of COVID-19 still face numerous challenges, including undocumented infections, asymptomatic contagion, uncertainty and quality issues in the reported data, flexible external non-pharmaceutical interventions, unknown resurgence patterns or periodicity, and multiple mutations.

This thesis aims to understand COVID-19 concerning the COVID-19 research landscape, transmission complexity, non-pharmaceutical interventions, and COVID-19 resurgence. Focusing on the COVID-19 challenges, this thesis first compares the key characteristics of COVID-19 disease with several known epidemics, and it summarizes the COVID-19 modeling complexities caused by these attributes. Starting from this basic knowledge, this thesis further explores COVID-19 modeling, which results in the following four contributions. (1) This thesis tracks the current COVID-19 modeling progress with natural language techniques and statistically summarizes the major facts of COVID-19 disease and COVID-19 modelling. This work structures a transdisciplinary research landscape and provides a holistic picture of COVID-19 modeling. (2) It infers the possible quantity of undocumented infections in the early stage of the COVID-19 outbreak with the proposed density-based Bayesian probabilistic compartmental model. This work examines the COVID-19 transmission complexities, in other words, undocumented infections, contagion reinforcement, and the imperfect conditions existing in COVID-19 reported data, that is, noise, sparsity, and uncertainty. (3) With the proposed event-driven generalized Susceptible-Exposed-Infectious-Recovered compartmental model, this thesis studies the impact of external interventions and activities in the dynamic COVID-19 evolving process and quantifies the efficacy of control policies and relaxation measures. (4) This thesis compares the differences between multiple COVID-19 waves, including the epidemiological attributes and the countermeasures, and it simulates the possible scenarios with different interventions and virus mutations. This exploration illustrates the possible reasons for COVID-19 resurgence and provides reliable guidance for society resuscitation.

Extensive experiments, including mean-field Bayesian inference, backward-looking

empirical evaluations, forward-looking simulations, and short-term forecast, demonstrate the effectiveness of the proposed methods for modeling the COVID-19 complexities aforementioned. The findings and quantitative results in this thesis indicate clues, evidence, and guidance for governments and policymakers to appropriately manage and mitigate the COVID-19 pandemic.

DEDICATION

To my beloved parents, Fenhai Liu and Yuhuan Cheng ...

ACKNOWLEDGMENTS

I am deeply grateful to my supervisor Prof. Longbing Cao for his support, guidance, and patience. Longbing is insightful, thoughtful, critical, and hardworking. He has taught me how to research independently and led me into the data science realm. Longbing has spent much time discussing new ideas with me, cultivating my research sense, and helping me with academic writing. Each time when I become stuck and feel depressed, he will encourage me, offer me professional suggestions, and re-ignite my enthusiasm for the research. I am extremely thankful that Longbing took me on as a Ph.D. student and continued to have faith in me over the four years. I have benefited greatly from working with him.

I offer my sincere thanks to my colleagues. I am fortunate to have been a part of the UTS Advanced Analytics Institute, where I have met many kind team mates: Dr. Guansong Pang, Dr. Trong Dinh Thac Do, Dr. Liang Hu, Dr. Shoujin Wang, Dr. Chengzhang Zhu, Dr. Wei Wang, Qi Zhang, Zhilin Zhao, Oliver Chi, Wenfeng Hou, Usman Naseem. I appreciate our intensive and enlightening discussions, academic collaboration, and the happy time we had during the four years.

I would also like to thank the visiting scholars and students in our laboratory. We shared unforgettable times at UTS. In particular, I am grateful to A/Prof. Pengfei Guo, A/Prof. Kelei Sun, A/Prof. Wenjuan Ren, Dr. Songlei Jian, Dr. Gaihong Yu, Dr. Jinjin Guo, Dr. Guifeng Wang, Dr. Ziyue Shi, Dr. Shufeng Hao, Qingfen Wang, Siyuan Ren, Yuteng Zhang, Ke Liu. Their companionship warms my daily life in this strange land.

Very special thanks are owed to Dr. Arman Abrahamyan and the excellent internship experience in Commonwealth Bank. Arman is experienced, professional, and kind. During the internship, Arman not only taught me much about how to approach an industry project, but he also shared his life and Ph.D. experiences with me. He has always been supportive and understanding. I am grateful for all his efforts to encourage me and boost my confidence.

I gratefully acknowledge China Scholarship Council for funding my Ph.D. study in UTS. I am fortunate to have been granted this valuable opportunity. Stepping outside of my comfort zone, I see a wider world. The insights and knowledge gained by studying abroad are the greatest assets in my life.

Finally, I wholeheartedly thank my family for their unconditional, unequivocal, and loving support and encouragement. They deserve endless gratitude. To my family, I offer everything, including this.

LIST OF PUBLICATIONS

PUBLISHED :

1. Liu Q, Do T D T, Cao L. Answer Keyword Generation for Community Question Answering by Multi-aspect Gamma-Poisson Matrix Completion[J]. IEEE Intelligent Systems, 2020.
2. Hou W, Liu Q, Cao L. Cognitive aspects-based short text representation with named entity, concept and knowledge[J]. Applied Sciences, 2020, 10(14): 4893.

UNDER REVIEW OR SUBMISSION:

3. Cao L and Liu Q. COVID-19 Modeling: A Review[J]. arXiv preprint arXiv:2104.12556, 2021. submitted to Machine Learning (submission ID: MACH-D-21-00689).
4. Liu Q and Cao L. Modeling COVID-19 uncertainties evolving with time and density-dependent social reinforcement and asymptomatic infections[J]. arXiv preprint arXiv: 2108.10029, 2021. submitted to Scientific Reports (submission ID: 2e67d7c3-4b47-4f49-ae8c-c9f03924caa3).
5. Cao L and Liu Q. How control and relaxation interventions and virus mutations influence the resurgence of COVID-19[J]. medRxiv, 2021. <https://www.medrxiv.org/content/10.1101/2021.08.31.21262897v1>, to be submitted.

TABLE OF CONTENTS

List of Publications	xi
List of Figures	xvii
List of Tables	xxi
1 Introduction	1
1.1 Background	1
1.2 Challenges and Gaps	3
1.2.1 COVID-19 Modeling Challenges	3
1.2.2 Modeling Gap Analysis	4
1.3 Research Problems and Objectives	7
1.4 Thesis Contributions	10
1.5 Thesis Organization	12
2 Related Work	15
2.1 Introduction	15
2.2 COVID-19 Epidemic Modeling	16
2.2.1 Compartmental Models	16
2.2.2 Data-driven Models	18
2.2.3 Probabilistic Models	20
2.3 COVID-19 Non-pharmaceutical Intervention Modeling	22
2.4 COVID-19 Resurgence Modeling	23
2.5 Summary	25
3 COVID-19 Characteristics and Complexities	27
3.1 Introduction	27
3.2 COVID-19 Disease Characteristics	28
3.2.1 High Contagiousness and Rapid Spread	28

TABLE OF CONTENTS

3.2.2	The Wide Range of Incubation Period	29
3.2.3	The Large Quantity of Asymptomatic and Undocumented Infections	29
3.2.4	High Mutation With Mysterious Strains and High Contagion . . .	30
3.3	COVID-19 Complexities	30
3.3.1	Data Challenges	30
3.3.2	Modeling Complexities	32
3.4	Summary	34
4	COVID-19 Modeling: the Research Landscape	37
4.1	Introduction	37
4.2	Method	38
4.3	Broad COVID-19 Modeling Landscape	38
4.3.1	Objectives of COVID-19 Modeling	38
4.3.2	Categorization of COVID-19 Modeling	42
4.4	Broad COVID-19 Modeling Research	44
4.4.1	COVID-19 Mathematical Modeling	44
4.4.2	COVID-19 Data-driven Learning	48
4.4.3	COVID-19 Domain-driven Modeling	52
4.4.4	COVID-19 Influence and Impact Modeling	58
4.4.5	COVID-19 Simulation Modeling	63
4.4.6	COVID-19 Hybrid Modeling	64
4.5	Summary	66
5	Probabilistic Compartmental Modeling of Social Reinforcement and Asymptomatic Infections	71
5.1	Introduction	71
5.2	The SUDR Design	74
5.2.1	Modeling COVID-19 Transmission Mechanisms	75
5.2.2	Modeling the Unknown Infections	76
5.2.3	Modeling the Contagion Reinforcement	77
5.2.4	Modeling Data Uncertainty, Sparsity and Noise	77
5.2.5	The SUDR Model Summary	78
5.2.6	Model Implementation	79
5.3	Experiment and Evaluation	80
5.3.1	Inferring the Undocumented Infections	80
5.3.2	Inferring the Epidemiological Attributes	85

5.3.3	Robustness Analysis	87
5.4	Discussion	90
5.5	Summary	91
6	Event-driven Compartmental Modeling of Non-pharmaceutical Inter- ventions	93
6.1	Introduction	93
6.2	The Event-driven SPEIQRD Model	94
6.2.1	SPEIQRD Model Mechanism	94
6.2.2	Modeling Dynamic Control and Lifting Processes	98
6.2.3	Parameter Estimation	98
6.3	Experiment and Evaluation	99
6.3.1	Interpretation of Public Data	99
6.3.2	The Epidemiological Attributes Inferred by SPEIQRD	100
6.3.3	The Event Impact Inferred by SPEIQRD	102
6.3.4	Short-term Forecast with the SPEIQRD Model	106
6.4	Summary	107
7	Estimating the Influence of Interventions and Viral Mutations on Resur- gence	109
7.1	Introduction	109
7.2	Data and Method	113
7.3	The Evaluation of COVID-19 Resurgence	115
7.3.1	How do Epidemiological Attributes Change Over the First-to- second Wave Transfer in Different Countries?	115
7.3.2	How do Different Impact Levels, Numbers and Timings of Imple- menting Interventions Affect the Trends of the Two Waves?	118
7.3.3	What may Happen When Tightening or Relaxing Containment With More Infectious Coronavirus Variants?	129
7.3.4	How Will the Epidemic Evolve Over Next 30 Days Under Different Containment Scenarios?	131
7.3.5	How Will the Epidemic Trend Evolve Over Next 30 Days With More Infectious Coronavirus Variants?	135
7.4	Discussion	136
7.5	Summary	138

TABLE OF CONTENTS

8	Conclusions and Future Work	141
8.1	Conclusion	141
8.2	COVID-19 Modeling Opportunities	142
	Bibliography	145

LIST OF FIGURES

FIGURE	Page
4.1 A transdisciplinary research landscape of COVID-19 modeling. From right: business problems and modeling objectives to the left: modeling techniques and methods.	43
5.1 The SUDR rationale: modeling the social interaction density of asymptomatic COVID-19 infections.	73
5.2 Compartment representation of the SUDR model.	75
5.3 The graphical model of the inference for SUDR.	79
5.4 The undocumented infections inferred by the SUDR model.	83
5.5 The inference of the prevalence-varying infection rate $\beta(I^U)$ of 11 countries. .	86
5.6 The performance comparison on the COVID-19 data with different levels of sparsity.	89
6.1 The dynamic compartmental model SPEIQRD for modeling both epidemiological attributes and intervention impact in the COVID-19 first wave evolving process. Event impact attributes: the protection rate α , and the deconfinement rate τ ; and epidemiological attributes: the infection rate β , the incubation rate γ , the quarantine rate δ , the recovery rate λ , and the mortality rate κ . Additionally, α , τ , λ and κ are time-varying parameters.	96
6.2 The time-varying recovery and mortality rates of the first waves in Germany, France and Italy.	103
6.3 The cumulative impact of sequential events on the daily new cases during the first wave in Germany, France and Italy.	105

LIST OF FIGURES

6.4 The prediction of numbers of active cases, recovered cases and deaths in three European countries by the SPEIQRD model. The case numbers during the first wave periods (in star) are the actual and used to train the model and infer its parameters. The corresponding lines show the inferred numbers of active, recovered and deceased cases. 106

7.1 The daily new cases in Germany, France and Italy from 01 March to 01 December, 2020. Two evolving waves of COVID-19 epidemic appear in each country, with all countries presenting similar but different resurgence attributes. 111

7.2 The time-varying recovery and mortality rates of two waves in Germany. . . . 117

7.3 The time-varying recovery and mortality rates of two waves in France. 118

7.4 The time-varying recovery and mortality rates of two waves in Italy. 119

7.5 The cumulative impact of sequential events on the daily new cases during the two waves in Germany. The blue axis represents the daily new cases, and the red axis represents the cumulative impact of sequentially-enforced events. The blue bars represent the daily new cases, the red solid line represents the cumulative impact of control events (i.e., the change of protection rate α), and the red dash line represents relaxing events (i.e., the change of deconfinement rate τ). The sequentially-implemented controlling and relaxing events lead to the increase of α and τ , respectively. 122

7.6 The impact of individual events on the movement of daily new cases during the two waves in Germany. The blue axis represents the daily new cases, and the red axis represents the event impact. The blue curve represents the change of daily new case number, the red regular triangles represent relaxing events, and the red inverted triangles represent control events. The deeper the color of red triangles, the greater the impact of the events on the daily case movement. 122

7.7 The cumulative impact of sequential events on the daily new cases during the two waves in France. 124

7.8 The impact of individual events on the movement of daily new cases during the two waves in France. 124

7.9 The cumulative impact of sequential events on the daily new cases during the two waves in Italy. 126

7.10 The impact of individual events on the movement of daily new cases during the two waves in Italy. 127

7.11	The simulation of the second wave with stronger mitigation (i.e., enforcing the same control measures in the first wave) but more infectious coronavirus variants.	131
7.12	The simulation of the second wave with more infectious coronavirus variants but mild mitigation strategies (no change of the relaxed containment).	131
7.13	The prediction of the next 30-day infection trend following the 2 nd COVID-19 wave in Germany by relaxing or enforcing containment strategies. (a)-(f) simulate six scenarios corresponding to different control and relaxing strategies with respective protection rates α and deconfinement rates τ . The diamond markers are reported data and the solid lines are the fitted data. The confirmed cases (blue line and purple diamond), the daily active cases (red line and green diamond) and the daily new cases (orange line and blue diamond) are forecasted.	133
7.14	The prediction of the next 30-day infection trend following the 2 nd COVID-19 wave in France by relaxing or enforcing containment strategies.	134
7.15	The prediction of the next 30-day infection trend following the 2 nd COVID-19 wave in Italy by relaxing or enforcing containment strategies.	135
7.16	The estimation of the epidemic with more infectious coronavirus variants. . .	136

LIST OF TABLES

TABLE	Page
1.1 Thesis overview (main body).	13
4.1 Business problems and objectives of COVID-19 modeling.	39
4.2 Examples of COVID-19 mathematical and statistical learning.	49
4.3 Examples of COVID-19 shallow and deep learning ^a	53
4.4 Examples of COVID-19 epidemiological modeling.	55
4.5 Comparison of COVID-19 infection diagnosis methods.	57
4.6 Examples of deep COVID-19 medical imaging analysis.	67
4.7 Examples of modeling for COVID-19 treatment and drug/vaccine development.	68
4.8 Examples of COVID-19 influence and impact modeling.	69
5.1 The COVID-19 data details for undocumented infection detection in the 11 European countries.	82
5.2 The quantitative comparison of the documented and undocumented infections in each country. The mean value and the 95% CI of I^D peak value, I^U peak value, and the ratio of $\max I^U / \max I^D$ are shown.	84
5.3 The detection and removal rates inferred by SUDR in each country. The mean value and the 95% CI of detection rate θ and removal rate γ are shown. A positive correlation can be found between θ and γ in most countries.	87
6.1 The major events in the first COVID-19 wave in Germany.	100
6.2 The major events in the first COVID-19 wave in France.	101
6.3 The major events in the first COVID-19 wave in Italy.	101
6.4 The epidemiological attributes of the 2020 first wave in Germany, France and Italy.	102
6.5 Attributes and effects of the events in the COVID-19 first wave in Germany.	104
6.6 Attributes and effects of the events in the COVID-19 first wave in France.	104

LIST OF TABLES

6.7	Attributes and effects of the events in the COVID-19 first wave in Italy. . . .	104
6.8	The 10-day prediction errors for Germany, France and Italy.	107
7.1	The major events undertaken during the two COVID-19 waves in Germany. .	114
7.2	The major events undertaken during the two COVID-19 waves in France. . .	114
7.3	The major events undertaken during the two COVID-19 waves in Italy. . . .	115
7.4	Epidemiological attributes of the 2020 first-to-second waves in Germany. . .	116
7.5	Epidemiological attributes of the 2020 first-to-second waves in France. . . .	116
7.6	Epidemiological attributes of the 2020 first-to-second waves in Italy.	116
7.7	Attributes of event effects of 2020 first-to-second waves in Germany.	121
7.8	Attributes of event effects of 2020 first-to-second waves in France.	125
7.9	Attributes of event effects of 2020 first-to-second waves in Italy.	126