

# **Deep Learning on Complex Graphs**

by Xu Chen

Thesis submitted in fulfilment of the requirements for  
the degree of

**Doctor of Philosophy**

under the supervision of Prof. Ivor W. Tsang

University of Technology Sydney  
Faculty of Engineering and Information Technology

Aug. 2021



# Certificate of Original Authorship

Graduate research students are required to make a declaration of original authorship when they submit the thesis for examination and in the final bound copies. Please note, the Research Training Program (RTP) statement is for all students. The Certificate of Original Authorship must be placed within the thesis, immediately after the thesis title page.

## Required wording for the certificate of original authorship

### CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Xu Chen declare that this thesis, is submitted in fulfilment of the requirements for the award of doctor of philosophy, in the 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.

*\*If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).*

*\*If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).*

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

Production Note:

Signature: Signature removed prior to publication.

Date: 2022-5-03

## Collaborative doctoral research degree statement

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Shanghai Jiao Tong University.

## Indigenous Cultural and Intellectual Property (ICIP) statement

This thesis includes Indigenous Cultural and Intellectual Property (ICIP) belonging to custodians or traditional owners. Where I have used ICIP, I have followed the relevant protocols and consulted with appropriate Indigenous people/communities about its inclusion in my thesis. ICIP rights are Indigenous heritage and will always remain with these groups. To use, adapt or reference the ICIP contained in this work, you will need to consult with the relevant Indigenous groups and follow cultural protocols.



## Acknowledgements

I would like to take this good opportunity to show my gratitude to my advisors, several professors, my colleagues, my friends and my family for their significant help during my doctoral study in University of Technology Sydney and Shanghai Jiao Tong University.

First, I would like to express my foremost and deepest gratitude to my supervisor Prof. Ivor W. Tsang, who plays an important role in my doctoral life. During my research life in University of Technology Sydney, I am deeply impressed by his broad knowledge in machine learning and his pure altitude toward research. In each discussion with him, I can learn something new about how to understand different works, how to bridge them and how to generalize them to other tasks. These cases improve my research taste to a new level and tell me what is research philosophy. He teaches me how to think, how to summarize and then how to create, which helps me to touch top conferences and journals. Besides, he also gives me a lot of courage to face difficulties, especially when I was nearly on the brink of collapse in my doctoral life. With his help, I really find the happiness of doing pure research.

Second, I also would like to appreciate my co-supervisor Prof. Ya Zhang, who gives me sufficient freedom to explore the research area I like and helps me build my research basis. She patiently corrects my bad research habits, showing me how to formalize a good research style in the doctoral life. In particular, when I was new to the doctoral research, I often made some stupid mistakes and got stuck in various problems. Prof. Ya Zhang always encourages me and pulls me back from dilemmas. Her vision on the proposal of how to judge a good idea and endless help on teaching me the writing of research papers really help me a lot in my later research life. Besides, through the discussion with her, I learned how to think and understand a research work from the view of reviewers. I cannot finish this thesis if without her high standards, unlimited patience, generous support and guidance.

I feel very lucky to attend the joint degree program between University of Technology Sydney and Shanghai Jiao Tong University. Wherein I would like to express my sincere appreciation to Prof. Ivor W. Tsang, Prof. Ya Zhang and Prof. Siheng Chen, who provide their selfless help to me and offer me a great opportunity to work in such great teams.

Within my doctoral life, I have met and made friends with many famous researchers in different areas and learned many useful experiences from them. Specifically, with the help of Jiangchao Yao, I have the chance to meet Prof. Ivor W. Tsang. Thanks to my friend Yuangang Pan, I can quickly adapt to the life in Australia. With the help of them, I have broaden my eyes on the machine learning community, and also realize what I lack for the top research level in detail.

Then, it comes to my dear colleagues and friends in my doctoral study. I would like to thank Dr.Yexun Zhang, Dr.Zhiyi Tan and Fei Ye for their selfless help in my research, Huangjie Zheng, Yujun Gu, Jie Chang, Peisen Zhao, Haoyan Guan, Jiajie Wang, Hanqing Zhao, Yu Li, and Bo Huang for being the “The ten brothers of gourd” in our team, the lovely juniors Maosen Li, Kenan Cui, Hao Wu, Zhikang Li, Jingchao Su, Chengyang Li, Yingying Xue, Tengfei Hou, Yifei Hu, Luo Chen Lyu and many others. I am happy to study with you all and wish you all have a bright future. I am also grateful to all the other friends in Sydney: Yuangang Pan, Yueming Lyu, Yinghua Yao, Muming Zhao, Jing Li, Xiaowei Zhou, Yan Zhang, Yaxin Shi, Xingrui Yu, Xiaofeng Xu and Xiaofeng Cao for their support and company.

Finally, I would like to express my deep gratitude to my parents, my sister and my girl friend. You all always get my back and stay around. It is you that give me great courage to solve so many difficulties and then I could better pursue my research goal. The love from you will be remembered by me forever. In the unforgettable five years, I experienced so many things including both happiness and sadness. I can see that I grew up a lot, with my own cognition towards life. The life is usually of regrets, but cherishing the moment now is the key to a happy life.

# DEEP LEARNING ON COMPLEX GRAPHS

## ABSTRACT

Deep learning on graphs has recently become a hot research topic in machine learning and shown great promise in a wide variety of applications such as recommendation and social network analysis. This dissertation considers a set of more practical and challenging cases of learning on complex graphs, *i.e.*, edges with attribute information, nodes with missing attributes and multiple graphs with overlapped nodes.

Our first work aims to model the topological information in signed directed networks where the edges have additional attribute information, *i.e.* signs and directions. In signed directed networks, different signs and directions have different effects in information propagation, which raises challenges to model the structural information. We propose to decouple the modeling of signs and directions with different network parameters, and meanwhile maximize the log-likelihood of observed edges by a variational *evidence lower bound* to learn the node representations. The experimental results show the effectiveness of the proposed model on both link sign prediction and node recommendation task.

Our second work considers learning on attribute-missing graphs where attributes of partial nodes are entirely missing. Previous graph learning algorithms have limitations in dealing with this kind of graphs. The random walk based methods suffer from the sampling bias issue of structures. The popular graph neural networks feed the structures and attributes to a shared network, and thus become incompatible for attribute-missing nodes. To better learn on attribute-missing graphs, we consider the structures and

attributes as two correlated views of the node information and make a shared-latent space assumption of these two views. Based on the assumption, we propose to model the two views by two different encoders and meanwhile maintain their joint distribution by a novel distribution matching scheme. Extensive experiments on seven real-world datasets show the superiority of the proposed model on both the link prediction task and the newly introduced node attribute completion task.

Moreover, single graphs may have overlapped nodes and become one complex graph. A popular case is the user-item bipartite graphs in cross domain recommendation where users are shared while items are from different domains. Previous methods usually emphasize the overlapped features of user preferences while compromise the domain-specific features or learn the domain-specific features by heuristic human knowledge. Our third work proposes to learn both features in a more practical way by an equivalent transformation assumption. The assumption hypothesizes the user preference in each domain can be mutually converted to each other by equivalent transformation. Then, a novel equivalent transformation based distribution matching scheme is developed to model the joint distribution of user behaviors across domains and conduct the recommendation task. The results on three real-world benchmarks confirm the superiority of the proposed model.

**KEY WORDS:** deep learning, graph learning, complex graphs



# Contents

<b>Acknowledgements</b>	<b>I</b>
<b>List of Figures</b>	<b>XIV</b>
<b>List of Tables</b>	<b>XVII</b>
<b>List of Algorithms</b>	<b>XIX</b>
<b>Nomenclature</b>	<b>XXI</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Deep Learning on Graphs . . . . .	2
1.3 Deep Learning on Complex Graphs . . . . .	5
1.4 Contributions and Organizations . . . . .	8
<b>Chapter 2 Background</b>	<b>11</b>
2.1 Deep Learning on Graphs . . . . .	11
2.1.1 Network Embedding . . . . .	11
2.1.2 Graph Neural Networks . . . . .	14
2.2 Learning on Signed Directed Networks . . . . .	17
2.3 Learning to Collaborate Different Bipartite Graphs . . . . .	19
2.4 Advances in Deep Generative Models . . . . .	20
<b>Chapter 3 Learning on Signed Directed Networks</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 DVE: Decoupled Variational Embedding . . . . .	25
3.2.1 Problem Definition . . . . .	25
3.2.2 Variational Auto-Encoding Formulation . . . . .	26
3.2.3 Decoupled Variational Encoder . . . . .	28
3.2.4 Structure Decoder . . . . .	31
3.2.5 Model Learning . . . . .	32

3.2.6	Comparison Between DVE and Existing Methods . . . . .	33
3.2.7	Time Complexity Analysis . . . . .	33
3.3	Experiments and Analysis . . . . .	35
3.3.1	Dataset Description . . . . .	35
3.3.2	Baselines . . . . .	35
3.3.3	Experimental Setups . . . . .	37
3.3.4	Performance Comparison . . . . .	37
3.3.5	Qualitative Visualization . . . . .	43
3.3.6	Ablation Study . . . . .	47
3.4	Summary . . . . .	50
<b>Chapter 4 Learning on Attribute-Missing Graphs</b>		<b>53</b>
4.1	Introduction . . . . .	53
4.2	SAT: Structure-Attribute Transformer . . . . .	56
4.2.1	Problem Definition . . . . .	56
4.2.2	Overview . . . . .	56
4.2.3	Paired Structure-Attribute Matching . . . . .	58
4.2.4	Adversarial Distribution Matching . . . . .	59
4.2.5	Objective Function and Implementation . . . . .	61
4.2.6	Time Complexity Analysis . . . . .	61
4.3	Experiments and Analysis . . . . .	63
4.3.1	Dataset Description . . . . .	63
4.3.2	Baselines . . . . .	64
4.3.3	Experimental Setups . . . . .	66
4.3.4	Node Attribute Completion . . . . .	67
4.3.5	Link Prediction . . . . .	74
4.3.6	Analysis of Learned Node Representations . . . . .	75
4.3.7	Learning Process Visualization . . . . .	76
4.3.8	Hyper-parameter $\lambda_c$ . . . . .	79
4.3.9	Empirical Running Time Analysis . . . . .	79
4.4	Summary . . . . .	80
<b>Chapter 5 Learning to Collaborate Different Bipartite Graphs</b>		<b>81</b>
5.1	Introduction . . . . .	82

5.2	ETL: Equivalent Transformation Learner . . . . .	83
5.2.1	Problem Definition . . . . .	83
5.2.2	Joint Distribution Modeling . . . . .	84
5.2.3	Joint Reconstruction Loss . . . . .	86
5.2.4	Prior Regularization Loss . . . . .	88
5.2.5	Objective Function and Implementation . . . . .	89
5.2.6	Discussion of Equivalent Transformation . . . . .	90
5.2.7	Time Complexity Analysis . . . . .	91
5.3	Experiments and Analysis . . . . .	92
5.3.1	Dataset Description . . . . .	92
5.3.2	Baselines . . . . .	93
5.3.3	Experimental Setups . . . . .	95
5.3.4	Performance Comparison . . . . .	97
5.3.5	Analysis of Learned User Preferences . . . . .	102
5.3.6	Ablation Study . . . . .	106
5.4	Summary . . . . .	109
	<b>Chapter 6 Conclusion</b>	<b>111</b>
6.1	Summary of Contributions . . . . .	111
6.2	Future Works . . . . .	112
	<b>Appendix A Deduction in Chapter 3</b>	<b>115</b>
A.1	Variational Evidence Lower Bound . . . . .	115
	<b>Bibliography</b>	<b>117</b>
	<b>Publications</b>	<b>133</b>



## List of Figures

1–1	An example to show the graph structured data in different scenarios. The bottom row lists three graph examples ( <i>i.e.</i> citation network, coauthor network, social network). The top row indicates an abstract graph example that summarizes the instances in different scenarios. . . . .	2
1–2	The tree to depict the studied three different scenarios of deep learning on complex graphs. In the edge level, the edges may have attribute information such as signs and directions. In the node attribute level, attributes of partial nodes may be entirely missing. In the graph level, different graphs may have overlapped nodes and become one more complex graph. This figure also indicates the studied three kinds of complex graphs in Chapter 3, Chapter 4 and Chapter 4, respectively. . . . .	3
1–3	The tree to depict different methodologies of deep learning on graphs. . . . .	4
1–4	A social network example shows the characteristics ( <i>i.e.</i> signed edges and directed edges) in signed directed networks. The structures in signed directed networks are coupled coupling signals inner signed edges, inner directed edges, between signed and directed edges. . . . .	6
1–5	A citation network example shows the characteristics of attribute-missing graphs, where attributes of only partial nodes are available and those of other nodes are entirely missing. Current popular methods including the random walk based and the GNN based cannot well handle this case. . . . .	7
1–6	An example to show that the user preference in CDR has both overlapped features and domain-specific features. With the correlations of user behaviors in different domains, knowledge transfer techniques can be employed to improve the learning of user preference and help the user behavior prediction in each domain. . . . .	7
2–1	An example to illustrate the typical learning process of random walk based network embedding. . . . .	12

2–2	An example to illustrate the typical learning process of matrix factorization based network embedding. In this figure, the adjacency matrix can be substituted with other graph matrices such as graph Laplacian matrix. Various factorization approaches can be employed such as SVD. $\otimes$ indicates a calculation function which is specific to the factorization method. . . . .	13
2–3	An example to illustrate the general process of DNN based graph learning algorithms. . . . .	13
2–4	An example to illustrate the general process of GNN based graph learning algorithms. In this figure, we take node $v_1$ as an anchor node and show two convolutional layers that involves different orders of the information propagation for $v_1$ . In the convolutional layers, the orders are visualized by different shades. . . . .	14
3–1	An example to illustrate the <i>high-order</i> topology and <i>first-order</i> topology in signed directed networks. Red arrows mean positive directed edges and blue arrows indicate negative directed edges. (a) is an example of signed directed networks $\mathcal{G}_e$ . (b) indicates the <i>high-order</i> topology of node $v_1$ in $\mathcal{G}_e$ , namely the local structures of $v_1$ . Different depth of shades represent different orders of structures for $v_1$ . (c) shows the <i>first-order</i> topology of $v_1$ , namely the closeness relationships between $v_1$ and its directly linked neighbours. The concentric circles with $v_1$ as the center indicate the closeness between $v_1$ and its positively linked nodes $v_2, v_4$ , the non-linked node $v_7$ and the negatively linked node $v_3$ . . . . .	24
3–2	The model architecture of DVE. We first decouple the signed directed graph into an undirected positive graph whose adjacency matrix is $A^p$ and an undirected negative graph whose adjacency matrix is $A^n$ . Then our decoupled variational encoder encodes $A^p$ and $A^n$ as the source node representation $Z_s$ and target node representation $Z_t$ , respectively. Finally, $Z_s$ and $Z_t$ are used to perform the balance pair-wise ranking loss which is also the structure decoder in DVE. Node $i$ is the source node and from $Z_s$ , node. Nodes $j, k, r$ are the target nodes and from $Z_t$ . $f(\cdot, \cdot)$ indicates the score of two nodes connecting with positive links. . . . .	26

3–3	The graphical model of our the decoupling idea on link direction. In this figure, $Z_s$ and $Z_t$ are the latent variables for the source node representation and target node representation, respectively. $\mathcal{E}$ indicates the observed signed directed edges. Solid arrows denote the generative process and dashed arrows denote the inference process. . . . .	27
3–4	Comparison of methods with different training data on Epinions, Slashdot and Wiki for two tasks. AUC in (a)(b)(c) is the metric for link sign prediction task and Recall@50 in (d)(e)(f) is the metric for node recommendation task. . . . .	41
3–5	Comparison of methods with different latent dimension on Epinions, Slashdot and Wiki for two tasks. AUC in (a)(b)(c) is the metric for link sign prediction task and Recall@50 in (d)(e)(f) is the metric for node recommendation task. . . . .	42
3–6	The empirical running time in each epoch of different methods. In this figure, DVE-N indicates the non-parallel DVE, DVE-H means the half-parallel one and the DVE-Q denotes the quarter-parallel one. . . . .	42
3–7	t-SNE visualization of topology preservation in Epinions. The node in red color means the sampled central node $i$ as source node. The nodes in blue color represents the positively linked neighbours $\mathcal{N}_p(i)$ and nodes in green color are the negatively linked neighbours $\mathcal{N}_n(i)$ . While the yellow ones are randomly sampled non-linked nodes $\mathcal{N}_{un}(i)$ for center node $i$ . Both $\mathcal{N}_p(i)$ and $\mathcal{N}_n(i)$ are target nodes. . . . .	44
3–8	Estimated probability density function of different types of node pairs on Slashdot for 6 methods. The red curve means the estimated PDF (Probability Density Function) of cosine similarity among positively linked node pairs. Similarly, the yellow curve and green curve denote the estimated PDF among the negatively linked node pairs and non-linked node pairs, respectively. . . . .	46
3–9	Performance of DVE with different generative functions. AUC in (a)(b)(c) refers to the metric for link sign prediction task. Recall@50 in (d)(e)(f) is the metric for node recommendation task. . . . .	47

3–10	Performance of DVE with different parameter settings. AUC in (a)(b)(c) is the metric for link sign prediction task and Recall@50 in (d)(e)(f) is the metric for node recommendation task. . . . .	49
3–11	Performance of DVE with different dropout rates. Dropout rate=0.0 means dropout keep probability is 1.0 during training. AUC in (a)(b)(c) is the metric for link sign prediction task and Recall@50 in (d)(e)(f) is the metric for node recommendation task. . . . .	50
4–1	Given a graph with attributes, based on the completeness of the node attributes, we may classify the graph into three types. . . . .	54
4–2	The comparison between recent GNN and our SAT, together with the illustration of our <i>shared-latent space</i> assumption. (a) means recent GNN usually requires structures and attributes as a whole input. $E$ is the encoder and $D$ is the decoder. (b) shows our <i>shared-latent space</i> assumption where $E_X$ and $E_A$ are two encoders and $D_X$ and $D_A$ are two decoders. (c) shows the general architecture of the proposed SAT. . . . .	54
4–3	Architecture of SAT. SAT first transforms attributes and structures into the latent space, then aligns the paired latent representations via adversarial distribution matching, and finally decodes to the original attributes and structures, namely the paired structure-attribute matching. . . . .	57
4–4	Node classification and profiling performance with less attribute-observed nodes. (a)-(c) illustrates the results of node classification with “X” setting. (d)-(f) shows the results of node classification with “A+X” setting. The dashed line is a criterion to criticize whether the restored attributes can enhance the GCN classifier. (g)-(i) shows the results of profiling task. Train ratio means the ratio of samples from the original train data. . . . .	73
4–5	Link prediction with less observed links on three datasets. We use AUC here to evaluate the performance. Train ratio means the ratio of random samples from original train data. The term “with X” means attributes are taken into consideration and “no X” means only structures are considered.	76
4–6	The t-SNE visualization of test node embeddings on Cora. Each color represents one class. Note that all methods learn node embeddings without class supervision. . . . .	76



4–7	Visualization of the training process for SAT(GCN) on Cora. (a) The self-reconstruction loss. (b) The cross-reconstruction loss. (c) The GAN loss in adversarial distribution matching. (d) Validation Recall@10 along the training steps. (e) The train and validation MMD distance between the aggregated distribution $q(z)$ and Gaussian prior $p(z)$ . (f) The train and validation MMD distance between distributions of $z_x$ and $z_a$ . . . . .	77
4–8	The effects of $\lambda_c$ on both the node classification and profiling task. (a-c) means the result for node classification with “A+X” setting on Cora, Citeseer and Pubmed. The dotted line with "only A" indicates that only the structural information is used, where GCN is the classifier. (d-f) indicates the result for profiling on Cora, Citeseer and Steam. The dotted line with "GCN" means we use the GCN as the attribute completion model. (g-i) shows the link prediction result with different $\lambda_c$ . . . . .	78
4–9	The empirical running time in each epoch of different methods. In this figure, SAT(GCN)-N indicates the non-parallel SAT(GCN), and SAT(GCN)-P means the parallel one. . . . .	79
5–1	Our equivalent transformation based model for recommendation in the U-NI scenario. Solid lines mean observed user-item interactions and dashed lines are the interactions we aim to predict with probabilities. Given a user, $z_x$ and $z_y$ are the preference representations encoded the user-item interactions in domain $\mathcal{X}$ and domain $\mathcal{Y}$ , respectively. The proposed ETL has an equivalent transformation between $z_x$ and $z_y$ , and models the joint distribution of the user behaviors across domains. . . .	84
5–2	The architecture of ETL. ETL encodes user behaviors in two domains with different encoders and then decodes the latent codes to user behaviors in each domain. The joint reconstruction loss and a prior regularization loss facilitates knowledge transfer between two domains and benefits the user behavior prediction. . . . .	85
5–3	The graphical model of our the equivalent transformation based model. In this figure, $z_x$ and $z_y$ are the latent variables of user representations encoded the behaviors $x$ of domain $\mathcal{X}$ and $y$ of domain $\mathcal{Y}$ , respectively. Solid arrows denote the generative process and dashed arrows denote the inference process. . . . .	87

5–4	The effects of different latent dimensions on Movie & Book. . . . .	99
5–5	The effects of different latent dimensions on Movie & Music. . . . .	99
5–6	The effects of different latent dimensions on Music & Book. . . . .	100
5–7	The effects of different sparsity levels on Movie & Book. Train ratio means the ratio of the original train data. . . . .	100
5–8	The effects of different sparsity levels on Movie & Music. Train ratio means the ratio of the original train data. . . . .	101
5–9	The effects of different sparsity levels on Music & Book. Train ratio means the ratio of the original train data. . . . .	101
5–10	The empirical running time in each epoch of different methods. . . . .	102
5–11	The results of model performance and MMD distance on three benchmarks. This experiment is designed to show that ETL has the capacity to learn the domain-specific features of user preferences. . . . .	104
5–12	The effect of different transformations on Movie & Book. Note that we show the results on Movie & Book here as an example to illustrate our idea. Results on other two benchmarks follow similar pattern. . . . .	106
5–13	The effects of hyper-parameters $\lambda, \eta$ on two datasets. We show the results of two datasets here to illustrate the observations and the results of another dataset follow similar pattern. . . . .	108

## List of Tables

3–1	The statistics of Epinions, Slashdot and Wiki utilized in our experiments.	36
3–2	Link sign prediction performance. Names with * refer to our methods. Compared to SiNE, the absolute improvement percentage of DE and DVE are given. . . . .	37
3–3	Node recommendation performance on Epinions. Names with * refer to our methods. The metrics for this task are Recall@k and Precision@k. We pick k=10,20,50 here. Compared to SiNE, the absolute improvement of DVE is given in the table. . . . .	39
3–4	Node recommendation performance on Slashdot. Names with * refer to our methods. The metrics for this task are Recall@k and Precision@k. We pick k=10,20,50 here. Compared to SiNE, the absolute improvement of DVE is given in the table. . . . .	39
3–5	Node recommendation performance on Wiki. Names with * refer to our methods. The metrics for this task are Recall@k and Precision@k. We pick k=10,20,50 here. Compared to SiNE, the absolute improvement of DVE is given in the table. . . . .	40
3–6	Different generative functions for $f_s$ . In this table, $[\cdot, \cdot]$ means the concatenation operation and $W_C \in \mathbb{R}^{2d \times 2d}$ is the weight of MLP for concatenation. $\odot$ indicates the element-wise product operation and $W_E \in \mathbb{R}^{d \times d}$ is the weight of MLP for element-wise product. . . . .	48
4–1	The statistics of seven datasets. In this table, “attribute form” means the attribute style. “#avg hot num” means the average hot number for multi-hot attributes of nodes. #class indicates the number of categories.	62

4–2	Node classification of the <i>node-level</i> evaluation for <i>node attribute completion</i> . The first column with “X”, “A” and “A+X” indicates three settings to do node classification with only attributes, only structures and the fused one. Note that SAT has an extendable GNN backbone, so we combine it with different models as SAT(GCN), SAT(GraphSage) and SAT(GAT). SAT(GCN)-no self, SAT(GCN)-no cross and SAT(GCN)-no adver respectively denotes SAT without self-reconstruction terms, cross-reconstruction terms and adversarial learning terms. The term “True attributes” indicates we use the ground truth attributes to do node classification. . . . .	69
4–3	Profiling of the <i>attribute-level</i> evaluation for <i>node attribute completion</i> . Note that we use top 3, 5, 10 to evaluate the performance on Steam since the average non-zero hot number of node attributes on Steam is quite small.	72
4–4	Area under curve (AUC) and average precision (AP) of Link prediction task. “A” indicates the <i>structure-only</i> aspect and only structures are used. While “A+X” means the <i>fused structure-attribute</i> aspect and both structures and attributes are used. . . . .	74
5–1	The statistics of datasets. . . . .	93
5–2	The overall comparison on Movie & Book. The underlined results are the best performance of baselines. . . . .	96
5–3	The overall comparison on Movie & Music. The underlined results are the best performance of baselines. . . . .	96
5–4	The overall performance on Music & Book. The underlined results are the best performance of baselines. Compared to ETL, the t-test results of other baselines are shown in this table. . . . .	97
5–5	The binary classification results with AUC. This experiment is designed to verify that the proposed ETL can better learn the overlapped features of user preferences in CDR. . . . .	103
5–6	The definitions of different transformations. Trans1 and Trans2 are both not equivalent transformation. Trans3 is one simple equivalent transformation. Trans4 is our extended non-linear equivalent transformation from Trans3. . . . .	106

- 5–7 The effects of different priors on three benchmarks. Uniform indicates samples from  $U(0, 1)$ , Laplace indicates samples from  $L(0, 1)$  and Gaussian means samples from  $\mathcal{N}(0, 1)$ . For multi-variate Gaussian (MVGaussian), we set it as  $MVG = \mathcal{N}(0, 1) + \mathcal{N}(3, 1)$  to form multiple peaks. . . 107



## List of Algorithms

3–1	Decoupled Variational Embedding (DVE) . . . . .	34
4–1	Structure-attribute Transformer (SAT) . . . . .	60
5–1	Equivalent Transformation Learner (ETL) . . . . .	90





## Nomenclature

$\mathcal{G}$	a graph
$\mathcal{V}$	the set of nodes
$v_i$	the $i$ th node
$\mathcal{E}$	the observed edges
$X$	the attribute matrix of nodes
$A$	the adjacent matrix
$N$	the number of nodes
$F$	the attribute dimension
$L$	the symmetric normalized Laplacian matrix
$I_N$	an identity matrix of size $N$
$d$	the latent dimension
$E$	an encoder
$D$	an decoder
$\mathcal{D}$	a discriminator in adversarial learning
$D_{KL}$	Kullback-Leibler divergence
$\mathcal{L}$	loss function
$\mathcal{N}(\mu, \sigma)$	a Gaussian distribution
GNN	graph neural networks
<i>ELBO</i>	<i>evidence lower bound</i>
CDR	cross domain recommendation

