

Artificial Intelligence: Foundations, Theory, and Algorithms

Series Editors

Barry O'Sullivan, Department of Computer Science, University College Cork, Cork,
Ireland

Michael Wooldridge, Department of Computer Science, University of Oxford,
Oxford, UK

Artificial Intelligence: Foundations, Theory and Algorithms fosters the dissemination of knowledge, technologies and methodologies that advance developments in artificial intelligence (AI) and its broad applications. It brings together the latest developments in all areas of this multidisciplinary topic, ranging from theories and algorithms to various important applications. The intended readership includes research students and researchers in computer science, computer engineering, electrical engineering, data science, and related areas seeking a convenient way to track the latest findings on the foundations, methodologies, and key applications of artificial intelligence.

This series provides a publication and communication platform for all AI topics, including but not limited to:

- Knowledge representation
- Automated reasoning and inference
- Reasoning under uncertainty
- Planning, scheduling, and problem solving
- Cognition and AI
- Search
- Diagnosis
- Constraint processing
- Multi-agent systems
- Game theory in AI
- Machine learning
- Deep learning
- Reinforcement learning
- Data mining
- Natural language processing
- Computer vision
- Human interfaces
- Intelligent robotics
- Explanation generation
- Ethics in AI
- Fairness, accountability, and transparency in AI

This series includes monographs, introductory and advanced textbooks, state-of-the-art collections, and handbooks. Furthermore, it supports Open Access publication mode.

More information about this series at <https://link.springer.com/bookseries/13900>

Chuan Shi • Xiao Wang • Philip S. Yu

Heterogeneous Graph Representation Learning and Applications

 Springer

Chuan Shi
School of Computer Science
Beijing University of Posts
and Telecommunications
Beijing, China

Xiao Wang
School of Computer Science
Beijing University of Posts
and Telecommunications
Beijing, China

Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
Chicago, IL, USA

ISSN 2365-3051 ISSN 2365-306X (electronic)
Artificial Intelligence: Foundations, Theory, and Algorithms
ISBN 978-981-16-6165-5 ISBN 978-981-16-6166-2 (eBook)
<https://doi.org/10.1007/978-981-16-6166-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Graphs and networks are ubiquitous in today's interconnected world. Among complex network models, a specific one, called *heterogeneous network* (or *heterogeneous graph*), which models real-world systems as interactions among a massive set of multi-modal and multi-typed objects, is of particular importance because the explicit modeling of the inherent structure of complex networks facilitates powerful, in-depth network analysis. In recent years, *representation learning* (which is also known as *embedding learning*), which represents high-dimensional data with lower dimensional distributions by various deep learning or embedding methods, has been rapidly developed as a powerful tool for high-dimensional data analysis. Similarly, *graph representation learning* (also called *network embedding*), which learns representations of nodes/edges in a lower-dimensional space, has demonstrated its effectiveness for various graph mining and graph analysis tasks.

This book is the first book dedicated to *heterogeneous graph representation learning*, which learns node/edge representations in a lower dimensional space while preserving the heterogeneous structures and semantics for downstream tasks (e.g., node/graph classification and link prediction). Heterogeneous graph representation learning has become a powerful, realistic, and general network modeling tool in recent years and has attracted increasing attention in both academia and industry.

This book serves as a comprehensive and extensive introduction to heterogeneous graph representation learning and its applications, including a survey of current developments and the state of the art in this booming field. It not only extensively introduces the mainstream techniques and models, including structure-preserved, attribute-assisted, and dynamic graph, but also presents wide applications in recommendation, text mining, and industry. In addition, the book provides a platform and practice of heterogeneous graph representation learning. As the first book on the theme, it summarizes the latest developments and presents cutting-edge research on heterogeneous graph representation learning. It may have double benefits: (1) providing researchers with an understanding of the fundamental issues and a good entry point for working in this rapidly expanding field, and (2) presenting the latest research on applying heterogeneous graphs to model real systems and learning structural features of interaction systems.

The authors of this book have done substantial research on heterogeneous graph representation learning and the related themes. Philip S. Yu is one of the leading experts on data mining and heterogeneous information networks. Chuan Shi is a long-term collaborator with Philip on research into heterogeneous information networks. Chuan has systematically studied the recommendation and representation learning based on heterogeneous graphs, applied heterogeneous information network modeling to e-commerce and text mining, and dived recently into heterogeneous graph representation learning. Wang Xiao is a rising-star scholar in network embedding community. The book systematically summarizes their contributions in the direction of heterogeneous graph representation learning. This book can be used not only as a guidebook for academia and industry but also as a textbook for undergraduate and graduate students. I hope you enjoy reading it.

Michael Aiken Chair Professor
University of Illinois at Urbana-Champaign

Jiawei Han

Preface

Heterogeneous graph, containing different types of nodes and links, is ubiquitous in the real world, ranging from bibliographic networks and social networks to recommendation systems. Currently, heterogeneous graph representation learning, which learns node/edge representations in a lower dimensional space while preserving the heterogeneous structures and semantics for downstream tasks (e.g., node/graph classification and link prediction), has attracted considerable attentions, and we have witnessed the impressive performance of heterogeneous graph representation learning methods on various real-world applications (e.g., recommender systems). The increasing number of works on heterogeneous graph representation learning indicates a global trend in both academic and industrial communities. Thus, there is a pressing demand for comprehensively summarizing and discussing heterogeneous graph representation learning methods.

Compared with homogeneous graph representation learning, heterogeneous graph representation learning shows different challenges because of the heterogeneity. For example, heterogeneous graph has more complex structures caused by multiple relations, where the node attributes are also heterogeneous. The heterogeneous graph representation learning is highly related with real-world applications from the heterogeneous graph construction to learning, which may need more advanced domain knowledge. All these factors heavily affect the performance of heterogeneous graph representation learning, which should be carefully considered. Therefore, researches on heterogeneous graph representation learning are of great scientific and application value.

This book serves the interests of specific reader groups. Generally, the book is intended for anyone who wishes to understand the fundamental problems, techniques, and applications of heterogeneous graph representation learning. In particular, we hope that students, researchers, and engineers will find this book inspiring.

This book is divided into four parts, and the readers are able to quickly understand this field through the first part, deeply study the techniques and applications with the second and third parts, and perform typical algorithms with a platform in the fourth part.

- In the first part, we first present an overview of recent heterogeneous graph representation learning methods from different aspects, including both of the methodology and technique levels. Open sources are also summarized so as to facilitate future research and applications in this area. This part will help readers rapidly understand the overall development of this field. In particular, in Chap. 1, the basic concepts and definitions, as well as the background of homogeneous and heterogeneous graph representation learning, will be introduced. The method taxonomy and open sources will be summarized in Chap. 2.
- In the second part, we then provide an in-depth and detailed introduction of representative heterogeneous graph representation learning techniques. This part will help readers understand the fundamental problems in this field and illustrate how to design the state-of-the-art heterogeneous graph representation learning methods for these problems. In particular, the structure-preserved heterogeneous graph representation learning methods are discussed in Chap. 3, including the meta-path structure and network schema structure. In Chap. 4, the heterogeneous graph representation learning methods with attributes are presented, mainly focusing on the heterogeneous graph neural networks. After that, we introduce the dynamic heterogeneous representation learning methods in Chap. 5, which consider the incremental learning, sequence information, and temporal interaction. Then, in Chap. 6, we discuss some emerging topics of heterogeneous graph representation learning, covering the adversarial learning, sampling, and hyperbolic representation learning.
- In the third part, we summarize the real-world applications based on heterogeneous graph representation learning. This part enables readers to know the successful applications of heterogeneous graph representation learning and the way of applying the advanced techniques to the real-world scenarios. Specifically, in Chap. 7, we show how the heterogeneous graph representation learning improves different recommender systems, e.g., the top-N recommendation, cold-start recommendation, and author-set recommendation. The application on text mining is introduced in Chap. 8, focusing on the short text classification and the news recommendation scenarios. In Chap. 9, we present the heterogeneous graph representation learning in industry applications, e.g., the cash-out user detection, intent recommendation, share recommendation, and friend-enhanced recommendation.
- In the fourth part, we introduce a platform of heterogeneous graph representation learning and conclude this book. Considering the importance of deep learning platforms, we introduce the foundation platforms on graph machine learning, especially the platform of heterogeneous graph representation learning in Chap. 10. Also, we take three representative heterogeneous graph neural networks as examples, showing how we can perform them using the platform. Finally, the future research directions and open problems are discussed in Chap. 11.

Writing a book always involves more people than just the authors. We would like to express our sincere thanks to all those who worked with us on this book. They are Deyu Bo, Jiawei Liu, Ruijia Wang, Yugang Ji, Houye Ji, Yiding Zhang, Mengmei

Zhang, Tianchi Yang, Shaohua Fan, Chunchen Wang, Hui Han, Qi Cui, Qi Zhang, Nian Liu, Yuanxin Zhuang, Zhenyi Wang, Guanyi Chu, Hongrui Liu, Chen Li, Tianyu Zhao, Xinlong Zhai, Donglin Xia, and Fengqi Liang. We also give our thanks to many students of Prof. Philip S. Yu for their careful proofreading. They are Yuwei Cao, Yingtong Dou, Ziwei Fan, He Huang, Xiaohan Li, Zhiwei Liu, and Congying Xia. In addition, the work is supported by the National Natural Science Foundation of China (No. U20B2045, U1936220, 61772082, 61702296, 62002029, 62172052). It is also supported in part by NSF under grants III-1763325, III-1909323, III-2106758, and SaTC-1930941. We also thank the supports of these grants. Finally, we thank our families for their wholehearted support throughout this book.

About the Book

Representation learning in heterogeneous graphs (HGs) is intended to provide a meaningful vector representation for each node so as to facilitate downstream applications, such as link prediction, personalized recommendation, node classification, etc. This task, however, is challenging not only because of the need to incorporate heterogeneous structural (graph) information consisting of multiple types of node and edge but also because of the need to consider heterogeneous attributes or types of content (e.g., text or image) associated with each node. Although considerable advances have been made in homogeneous (and heterogeneous) graph embedding, attributed graph embedding, and graph neural networks, few are capable of simultaneously and effectively taking into account the heterogeneous structural (graph) information as well as the heterogeneous content information of each node.

In this book, we provide a comprehensive survey of the current developments in HG representation learning. More importantly, we present the state-of-the-art in this field, including theoretical models and real applications that have been showcased at the top conferences and journals, such as TKDE, KDD, WWW, IJCAI, and AAAI. This book has two major objectives: (1) to provide researchers with an understanding of the fundamental issues and a good point of departure for working in this rapidly expanding field and (2) to present the latest research on applying heterogeneous graphs to model real systems and learning structural features of interaction systems. To the best of our knowledge, it is the first book to summarize the latest developments and present cutting-edge research on heterogeneous graph representation learning. To gain the most from it, readers should have a basic grasp of computer science, data mining, and machine learning.

Contents

1	Introduction	1
1.1	Basic Concepts and Definitions	1
1.2	Graph Representation Learning	4
1.3	Heterogeneous Graph Representation Learning and Challenges	5
1.4	Organization of the Book	6
	References	6
2	The State-of-the-Art of Heterogeneous Graph Representation	9
2.1	Method Taxonomy	9
2.1.1	Structure-Preserved Representation	9
2.1.2	Attribute-Assisted Representation	12
2.1.3	Dynamic Representation	13
2.1.4	Application-Oriented Representation	14
2.2	Technique Summary	16
2.2.1	Shallow Model	16
2.2.2	Deep Model	16
2.3	Open Sources	17
2.3.1	Benchmark Datasets	18
2.3.2	Open-Source Code	18
2.3.3	Available Tools	20
	References	21
Part I Techniques		
3	Structure-Preserved Heterogeneous Graph Representation	29
3.1	Introduction	29
3.2	Meta-Path Based Random Walk	30
3.2.1	Overview	30
3.2.2	The HERec Model	31
3.2.3	Experiments	35

3.3	Meta-Path Based Decomposition	41
3.3.1	Overview	41
3.3.2	The NeuACF Model	42
3.3.3	Experiments	47
3.4	Relation Structure Awareness	48
3.4.1	Overview	48
3.4.2	Preliminary	51
3.4.3	The RHINE Model	53
3.4.4	Experiments	56
3.5	Network Schema Preservation	58
3.5.1	Overview	58
3.5.2	The NSHE Model	59
3.5.3	Experiments	64
3.6	Conclusions	65
	References	66
4	Attribute-Assisted Heterogeneous Graph Representation	71
4.1	Introduction	71
4.2	Heterogeneous Graph Attention Network	72
4.2.1	Overview	72
4.2.2	The HAN Model	74
4.2.3	Experiments	78
4.3	Heterogeneous Graph Propagation Network	83
4.3.1	Overview	83
4.3.2	Semantic Confusion Analysis	85
4.3.3	The HPN Model	87
4.3.4	Experiments	90
4.4	Heterogeneous Graph Structure Learning	93
4.4.1	Overview	93
4.4.2	The HGSL Model	94
4.4.3	Experiments	100
4.5	Conclusions	102
	References	103
5	Dynamic Heterogeneous Graph Representation	107
5.1	Introduction	107
5.2	Incremental Learning	108
5.2.1	Overview	108
5.2.2	The DyHNE Model	109
5.2.3	Experiments	116
5.3	Sequence Information	122
5.3.1	Overview	122
5.3.2	The SHCF Model	124
5.3.3	Experiments	129
5.4	Temporal Interaction	132
5.4.1	Overview	132

5.4.2	The THIGE Model	132
5.4.3	Experiments	138
5.5	Conclusion	140
	References	140
6	Emerging Topics of Heterogeneous Graph Representation	145
6.1	Introduction	145
6.2	Adversarial Learning	146
6.2.1	Overview	146
6.2.2	The HeGAN Model	147
6.2.3	Experiments	151
6.3	Importance Sampling	154
6.3.1	Overview	154
6.3.2	The HeteSamp Model	155
6.3.3	Experiments	160
6.4	Hyperbolic Representation	162
6.4.1	Overview	162
6.4.2	The HHNE Model	163
6.4.3	Experiments	166
6.5	Conclusion	170
	References	170
 Part II Applications		
7	Heterogeneous Graph Representation for Recommendation	175
7.1	Introduction	175
7.2	Top- <i>N</i> Recommendation	176
7.2.1	Overview	176
7.2.2	The MCRec Model	177
7.2.3	Experiments	182
7.3	Cold-Start Recommendation	185
7.3.1	Overview	185
7.3.2	The MetaHIN Model	186
7.3.3	Experiments	192
7.4	Author Set Recommendation	195
7.4.1	Overview	195
7.4.2	The ASI Model	197
7.4.3	Experiments	202
7.5	Conclusions	206
	References	206
8	Heterogeneous Graph Representation for Text Mining	209
8.1	Introduction	209
8.2	Short Text Classification	210
8.2.1	Overview	210
8.2.2	HG Modeling for Short Texts	211

8.2.3	The HGAT Model	212
8.2.4	Experiments	216
8.3	News Recommendation with Long/Short-Term Interest Modeling	220
8.3.1	Overview	220
8.3.2	Problem Formulation	220
8.3.3	The GNewsRec Model	221
8.3.4	Experiments	227
8.4	News Recommendation with Preference Disentanglement	229
8.4.1	Overview	229
8.4.2	The GNUD Model	231
8.4.3	Experiments	235
8.5	Conclusion	237
	References	238
9	Heterogeneous Graph Representation for Industry Application	241
9.1	Introduction	241
9.2	Cash-Out User Detection	242
9.2.1	Overview	242
9.2.2	Preliminaries	243
9.2.3	The HACUD Model	244
9.2.4	Experiments	247
9.3	Intent Recommendation	250
9.3.1	Overview	250
9.3.2	Problem Formulation	251
9.3.3	The MEIRec Model	252
9.3.4	Experiments	256
9.4	Share Recommendation	259
9.4.1	Overview	259
9.4.2	Problem Formulation	261
9.4.3	The HGSRec Model	262
9.4.4	Experiments	266
9.5	Friend-Enhanced Recommendation	270
9.5.1	Overview	270
9.5.2	Preliminaries	272
9.5.3	The SIAN Model	272
9.5.4	Experiments	276
9.6	Conclusions	281
	References	282
10	Platforms and Practice of Heterogeneous Graph Representation Learning	285
10.1	Introduction	285
10.2	Foundation Platforms	286
10.2.1	Deep Learning Platforms	286

- 10.2.2 Platforms of Graph Machine Learning 290
- 10.2.3 Platforms of Heterogeneous Graph Representation Learning 292
- 10.3 Practice of Heterogeneous Graph Representation Learning..... 293
 - 10.3.1 Build a New Dataset 294
 - 10.3.2 Build a New Model 299
 - 10.3.3 Practice of HAN 301
 - 10.3.4 Practice of RGCN 305
 - 10.3.5 Practice of HERec 307
- 10.4 Conclusion 309
- References 310
- 11 Future Research Directions..... 311**
 - 11.1 Introduction 311
 - 11.2 Preserving HG Structures 312
 - 11.3 Capturing HG Properties 313
 - 11.4 Deep Graph Learning on HG Data 313
 - 11.5 Making HG Representation Reliable 314
 - 11.6 Technique Deployment in Real-World Applications 315
 - 11.7 Others 315
 - References 316

About the Authors

Chuan Shi is a professor in the School of Computer Sciences of Beijing University of Posts and Telecommunications and the deputy director of the Beijing Key Lab of Intelligent Telecommunication Software and Multimedia. His main research interests include data mining, machine learning, artificial intelligence, and big data analysis. He has published more than 100 refereed papers, including top journals and conferences in data mining, such as IEEE TKDE, ACM TKDD, KDD, WWW, NeurIPS, AAAI, and IJCAI. In the meanwhile, his first monograph about heterogeneous information networks has been published by Springer. He has been honored with the best paper award in ADMA 2011 and ADMA 2018 and has guided students to the world championship in the IJCAI Contest 2015, the premier international data mining competition. He is also the recipient of “the Youth Talent Plan” and “the Pioneer of Teacher’s Ethics” in Beijing.

Xiao Wang is the associate professor in the School of Computer Sciences of Beijing University of Posts and Telecommunications. He was a postdoc in the Department of Computer Science and Technology at Tsinghua University. He got his Ph.D. in the School of Computer Science and Technology at Tianjin University and a joint-training Ph.D. at the Washington University in St. Louis. His main research interests include data mining, machine learning, artificial intelligence, and big data analysis. He has published more than 70 refereed papers, including top journals and conferences in data mining, such as IEEE TKDE, KDD, NeurIPS, AAAI, IJCAI, and WWW. He also serves as SPC/PC member and Reviewer of several high-level international conferences, e.g., KDD, AAAI, IJCAI, and journals, e.g., IEEE TKDE.

Philip S. Yu’s main research interests include big data, data mining (especially on graph or network mining), social network, privacy-preserving data publishing, data stream, database systems, and Internet applications and technologies. Dr. Yu is a distinguished professor in the Department of Computer Science at UIC and also holds the Wexler Chair in Information and Technology. Before joining UIC, he was with the IBM Thomas J. Watson Research Center, where he was the manager of

the Software Tools and Techniques department. He has published more than 1,300 papers in refereed journals and conferences with more than 149,000 citations and an H-index of 176. He holds or has applied for more than 300 U.S. patents. He is a fellow of the ACM and the IEEE. He is a recipient of the ACM SIGKDD 2016 Innovation Award and the IEEE Computer Society's 2013 Technical Achievement Award.