



UvA-DARE (Digital Academic Repository)

Machine learning tasks and representations for heterogeneous information networks

Fang, Y.

Publication date
2023

[Link to publication](#)

Citation for published version (APA):

Fang, Y. (2023). *Machine learning tasks and representations for heterogeneous information networks*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

**MACHINE LEARNING TASKS AND REPRESENTATIONS
FOR HETEROGENEOUS INFORMATION NETWORKS**

YANG FANG



**Machine Learning
Tasks and Representations
for Heterogeneous
Information Networks**

Yang Fang

Machine Learning Tasks and Representations for Heterogeneous Information Networks

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties
ingestelde commissie,
in het openbaar te verdedigen
op maandag 11 december 2023, te 16:00 uur

door

Yang Fang

geboren te Ningbo

Promotiecommissie

Promotor:	prof. dr. M. de Rijke	Universiteit van Amsterdam
Co-promotor:	dr. A. Yates	Universiteit van Amsterdam
Overige leden:	prof. dr. E. Kanoulas	Universiteit van Amsterdam
	dr. M. Alian Nejadi	Universiteit van Amsterdam
	prof. dr. P. Groth	Universiteit van Amsterdam
	dr. Q. Liao	Harbin Institute of Technology
	dr. X. Wang	Tianjin University
	prof. dr. C. Monz	Universiteit van Amsterdam
	prof. dr. Z. Ren	Shandong University

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The research was carried out at the Information Retrieval Lab of the University of Amsterdam, supported by the NSFC under grants Nos. 61872446, 62272469, U19B2024.

Copyright © 2023 Yang Fang, Changsha, China
Cover by Yang Fang
Printed by Ridderprint, Alblasterdam

ISBN: 978-94-6483-588-5

Acknowledgements

In October 2018, Maarten de Rijke agreed to have me join the Information Retrieval Lab (IRLab) at the University of Amsterdam. There was a lot of excitement for me to have the opportunity to experience a different life. However, after a long preparation and even with the flight booked, COVID-19 struck and changed everything.

I am sincerely grateful that Maarten still offered me the chance to join the group online. Nevertheless, it is a big regret for me that I was not able to receive his guidance offline in the Netherlands. Although there are many inconveniences meeting online, he has always maintained the greatest patience and encouragement for me. I am also impressed by his gentleness and erudition. Every time I communicated with him, I could always be inspired. I sincerely hope to meet Maarten in person in the Netherlands in the near future.

I want to thank my co-promotor Dr. Andrew Yates who helped me to revise the thesis. I also would like to thank my committee members, prof. dr. Evangelos Kanoulas, dr. Mohammad Alian Nejadi, prof. dr. Paul Groth, dr. Qing Liao, dr. Xin Wang, prof. dr. Christof Monz and prof. dr. Zhaochun Ren. Thanks a lot for your time and effort to review my thesis.

I thank my colleagues at the IRLab and I hope to meet them face-to-face in the future. Special thanks to Yangjun Zhang, Jin Huang, Jiahuan Pei, Jie Zhou, Yifan Chen, Wanyu Chen, and Weijia Zhang, who solved many problems for me. I also thank Petra for helping me organizing the online meetings and many other things because of my absence.

I want to thank my parents for their love and support. Last but not least, I want to thank my girlfriend Cuiwei Wu, your support and encouragement to me is irreplaceable. You fit all the good words. I look forward to building a small family with you. You are everything to me.

Yang Fang
Changsha, February 2023

Contents

Acknowledgements	iii
1 Introduction	1
1.1 Research outline and questions	2
1.2 Main contributions	3
1.2.1 Algorithmic contributions	3
1.2.2 Empirical contributions	4
1.3 Thesis overview	5
1.4 Origins	7
2 Dynamic Representation Learning	9
2.1 Introduction	9
2.2 Related work	12
2.2.1 Static network embeddings	12
2.2.2 Dynamic network embeddings	13
2.3 Preliminaries	13
2.4 Proposed model	14
2.4.1 Initial complex embedding mechanism	15
2.4.2 Dynamic embedding mechanism via triadic evolution processes	17
2.4.3 Graph prediction via LSTM-based deep autoencoders	20
2.5 Experiments	23
2.5.1 Datasets	23
2.5.2 Tasks	25
2.5.3 Baselines	25
2.5.4 Experimental setup	27
2.6 Experimental results and analysis	27
2.6.1 Link prediction	27
2.6.2 Changed link prediction	28
2.6.3 Node classification	29
2.6.4 Node prediction	30
2.6.5 Graph reconstruction	31
2.6.6 Anomaly detection	31
2.6.7 Computational costs	32
2.6.8 Ablation analysis	32
2.6.9 Parameter sensitivity	33
2.7 Conclusions	36
3 Pre-training Heterogeneous Information Networks	39
3.1 Introduction	39
3.2 Related work	41
3.2.1 Network representation learning	41
3.2.2 Graph neural networks	42
3.2.3 Graph pre-training	42
3.3 The proposed model PF-HIN	43

3.3.1	Heterogeneous node sequence generation	44
3.3.2	Input embeddings learned via Bi-LSTMs	45
3.3.3	Masked node modeling	45
3.3.4	Adjacent node prediction	46
3.3.5	Transformer architecture	47
3.3.6	Fine-tuning PF-HIN	47
3.4	Experimental setup	48
3.4.1	Datasets	48
3.4.2	Algorithms used for comparison	49
3.4.3	Parameters	50
3.5	Results and analysis	50
3.5.1	Downstream tasks	50
3.5.2	Computational costs	54
3.5.3	Ablation analysis	54
3.5.4	Parameter sensitivity	58
3.6	Conclusions	59
4	Few-shot Representation Learning	61
4.1	Introduction	61
4.2	Related work	64
4.2.1	Graph neural networks	64
4.2.2	Contrastive learning	64
4.2.3	Few-shot learning	65
4.3	The proposed model META-HIN	66
4.3.1	Preliminaries and overview	66
4.3.2	Sampling strategy	67
4.3.3	Structure module	68
4.3.4	Meta-learning module	69
4.4	Experimental setup	73
4.4.1	Datasets	73
4.4.2	Algorithms used for comparison	75
4.4.3	Parameters	75
4.5	Results and analysis	76
4.5.1	Results of tasks	76
4.5.2	Ablation analysis	78
4.5.3	Parameter sensitivity analyses	81
4.6	Conclusion	82
5	Textual Representation Learning	85
5.1	Introduction	85
5.2	Related work	88
5.3	The proposed model P-HIN	88
5.3.1	Text encoder	88
5.3.2	Graph encoder	89
5.3.3	Training	91
5.3.4	Few-shot inference	91

5.3.5	Optimization	92
5.3.6	Residual connection	92
5.4	Experiments	92
5.4.1	Experimental setup	93
5.4.2	Node classification	94
5.4.3	Link prediction	94
5.4.4	Keyword generation	95
5.4.5	Ablation study	96
5.4.6	Parameter analysis	97
5.5	Conclusion	98
6	Conclusions	99
6.1	Results	99
6.2	Future work	101
6.2.1	Model compression	101
6.2.2	Multi-modal heterogeneous information network attributes	101
6.2.3	Application-oriented heterogeneous information network representation learning	101
6.2.4	Interpretability of heterogeneous information network representation learning	102
	Bibliography	103
	Summary	111
	Samenvatting	113