Guest Editorial: Special Issue on Network Structural Modeling and Learning in Big Data

Di Jin, Wenjun Wang, Guojie Song, Philip S. Yu, and Jiawei Han

1 Introduction

A variety of phenomena and systems are usually regraded as networks with a set of nodes and relations. Diversified methods and models have been developed for different tasks on networks, such as structural discovery, link prediction, and anomaly detection. With the growing data scale due to the information explosion and scientific development, network structural modeling [1], [2] has become an interesting and new field of research. Furthermore, building and utilizing new technologies for efficient learning, inference, and predication on different types of networks has become the trend to look forward to. In recent years, all kinds of machine learning technologies on networks have been developed rapidly, especially the network embedding and graph neural networks. They induce various methods and achieve satisfactory performance on network tasks, including node clustering or classification and link prediction [3], [4]. Furthermore, network structural modeling on real big data is also interesting. It has also contributed to complex patterns in network tasks including sematic community detection [5] network clustering [6] community hiding [7] role discovery [10] graph neutral networks [8], [9] and so on.

2 Contents of the Special Issue

This special issue focused on the most recent advances in network structural modeling and learning with applications in the big data field. We received 42 valid paper submissions for this special issue. After several rounds of rigorous reviews and revisions, we decided to publish 6 of them. It contains five regular articles and a review paper which covers different topics of this special issue.

The first paper "Identification of Communities with Multi-Semantics via Bayesian Generative Model" [5] focused on the problem of identifying communities and learning the semantics interpretation of modules jointly in an end-to-end model. It designed a novel generative model which combines two closely related parts, one for community discovery and the other for content clustering and semantics interpretation. By extracting the potential correlation between these two parts, the new method is not only robust to discovering communities, but also able to provide a community with more than one semantic topic.

The second paper "Multi-view Clustering with Self-representation and Structural Constraint" [6] is a network-based method by fusing matrix factorization and low-rank representation of various views for multi-view clustering. Specifically, it constructed a network to remove heterogeneity of multi-view data and extract the shared features of multiple views with self-representation.

The next paper "How to Protect Ourselves from Overlapping Community Detection in Social Networks" [7] is opposed with the problem of community detection. It proposed a hiding algorithm for nodes who are located in the overlapped area found by overlapping community detection algorithms and introduced evaluation methodologies to prove the effectiveness of the proposed method. It is the first work to introduce importance degree of nodes in a community, which represents the node influence level. on other nodes. The proposed hiding algorithm is not only to remove a target node from the overlapping area, but also try to find an optimal number of adding and deleting.

The paper "MAFI: GNN-based Multiple Aggregators and Feature Interactions Network for Fraud Detection over Heterogeneous Graph" [8] proposed a novel heterogeneous GNN model. It is proposed to conduct fraud detection tasks. In detailed, multiple types of aggregators are applied on different relations to aggregate neighbor information and aggregator-level attention is utilized to learn the importance of different aggregators. Also, relation-level attention is leveraged to learn the importance of each relation. Besides, conventional update operations are replaced with vector-wise implicit and explicit feature interactions.

The next paper "Active and Semi-supervised Graph Neural Networks for Graph Classification" [9] proposed a novel active and semi-supervised graph neural network (ASGNN) framework. It endeavored to complete graph classification tasks with a small number of labeled graph examples and available unlabeled graph examples with active learning and semi-supervised learning. To improve the generalization performance of the graph classification model, multiple GNNs are trained collaboratively for promoting the expressiveness of each other and increasing the reliability of graph classification results.

The last is a review paper "A Survey on Role-Oriented Network Embedding" [10]. It first clarified the differences

• Di Jin and Wenjun Wang are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China.
E-mail: {jindi, wjwang}@tju.edu.cn.
• Guojie Song is with the Department of Machine Intelligence, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China. E-mail: gjsong@pku.edu.cn.
• Philip S. Yu is with the Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607 USA.
E-mail: psyu@uic.edu.
• Jiawei Han is with the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. E-mail: hanj@illinois.edu.

Digital Object Identifier no. 10.1109/TBDATA.2022.3162000

between community-oriented and role-oriented network embedding. It also proposed a general framework for understanding role-oriented network embedding and a two-level categorization to better classify existing methods. Moreover, it conducted comprehensive experiments to empirically evaluate the popular methods on a variety of role-related tasks including node classification and clustering (role discovery), top-k similarity search, and visualization using some widely-used synthetic and real-world datasets. Finally, it further discussed the research trend of role-oriented NE from the perspective of applications and pointed out some potential future directions.

Acknowledgments

We thank the Editor-in-Chief Prof Jie Tang for his outstanding support, and Ashutosh Rawat for the editorial and administrative assistance. In addition, we are deeply grateful to the reviewers for their contributions to the decision process.

References

[1] D. Jin *et al*., "A survey of community detection approaches: From statistical modeling to deep learning," *IEEE Trans. Knowl. Data Eng*., to be published, doi: 10.1109/TKDE.2021.3104155.

[2] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 32, no. 1, pp. 4–24, Jan. 2021.

[3] D. Jin, C. Huo, C. Liang, and L. Yang, "Heterogeneous graph neural network via attribute completion," in *Proc. Web Conf*., 2021, pp. 391–400.

[4] Z. Yu *et al*., "AS-GCN: Adaptive semantic architecture of graph convolutional networks for text-rich networks," *IEEE Int. Conf. Data Mining*, pp. 837–846, 2021.

[5] D. He *et al*., "Identification of communities with multi-semantics via Bayesian generative model," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2021.3131707.

[6] X. Gao *et al*., "Multi-view clustering with self-representation and structural constraint," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2021.3128906.

[7] D. Liu, G. Yang, Y. Wang, H. Jin, and E. Chen, "How to protect ourselves from overlapping community detection in social networks," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2022.3152431.

[8] N. Jiang, F. Duan, H. Chen, W. Huang, and X. Liu, "MAFI: GNN-based multiple aggregators and feature interactions network for fraud detection over heterogeneous graph," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2021.3132672.

[9] Y. Xie, S. Lv, Y. Qian, C. Wen and J. Liang, "Active and semi-supervised graph neural networks for graph classification," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2021.3140205.

[10] P. Jiao, X. Guo, T. Pan, W. Zhang, Y. Pei, and L. Pan, "A survey on role-oriented network embedding," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2021.3131610.