Representation Learning on Networks: Theories, Algorithms, and Applications

Jie Tang
Tsinghua University
Beijing, China
jietang@tsinghua.edu.cn

Yuxiao Dong
Microsoft Research
Redmond, WA
yuxdong@microsoft.com

ABSTRACT

Network representation learning offers a revolutionary paradigm for mining and learning with network data. In this tutorial, we will give a systematic introduction for representation learning on networks. We will start the tutorial with industry examples from Alibaba, AMiner, Microsoft Academic, WeChat, and XueTangX to explain how network analysis and graph mining on the Web are benefiting from representation learning. Then we will comprehensively introduce both the history and recent advances on network representation learning, such as network embedding and graph neural networks. Uniquely, this tutorial aims to provide the audience with the underlying theories in network representation learning, as well as our experience in translating this line of research into real-world applications on the Web. Finally, we will release public datasets and benchmarks for open and reproducible network representation learning research. The tutorial accompanying page is at https://aminer.org/nrl_www2019.

CCS CONCEPTS

• Information systems → Social networks; **• Computing methodologies → Unsupervised learning**; **Learning latent representations**; *Knowledge representation and reasoning*.

KEYWORDS

Network Embedding; Representation Learning; Graph Neural Networks; Feature Learning; Graph Mining; Network Science

ACM Reference Format:
Jie Tang and Yuxiao Dong. 2019. Representation Learning on Networks: Theories, Algorithms, and Applications. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308560.3320095

1 INTRODUCTION

In this tutorial, we aim to provide a comprehensive review for network representation learning. First, we will identify a wide range of important problems in networks, such as link prediction [13, 19], node attribute inference [5], and social influence modeling [21]. We then introduce the conventional network mining pipeline for addressing these problems [6]. Finally, we will show how the recent emergence of network representation learning offers a new paradigm for mining and learning with networks [2, 10].

With that, we will dive into the current development of network representation learning and introduce the important methods and milestones along each direction. The uniqueness of this tutorial lies in its emphasis on both the theoretical understanding and real applications of network representation learning. On one hand, we will try to provide the audience with the underlying theories in this topic and how different models can be unified and connected with each other. On the other hand, we will introduce how to translate the network representation learning research into online and mobile applications on Alibaba, AMiner, Microsoft Academic, MOOC, and Tencent Games and WeChat.

The main content of this tutorial is organized as follows. First, we introduce the history of (network) representation learning and how this discipline is developed from graph theory, deep learning, and natural language processing [1, 12, 14]. A brief timeline of network representation learning research is summarized in Figure 1, including both the development of network embedding and graph neural networks. Together with its history, we will also present the basic knowledge and motivation for learning network representations.

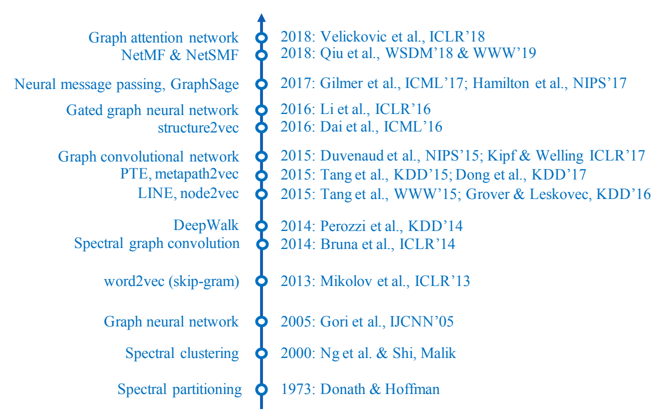

Figure 1: A timeline of network representation learning and graph neural network research as of 2018.

Second, we will introduce and categorize the recent developments of network representation learning. In specific, our focus will be on 1) skip-gram based algorithms (i.e., DeepWalk [15], LINE [20], node2vec [8]), and metapath2vec [4], 2) matrix factorization based models (i.e., NetMF [17] and NetSMF [16]), and 3) graph neural networks (i.e., Graph Convolution [3, 11], GraphSage [9], neural

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3320095

message passing [7], and Graph Attention [23]). In addition, we will show 1) how to incorporate social and network theories into representation learning, and 2) how to leverage these techniques to handle more complex networks, such as dynamic and heterogeneous networks [4, 24].

Third, we will theoretically analyze the underlying mechanism of different types of network representation learning models and see how they can be connected with graph theory [16, 17]. For example, we will show that DeepWalk is actually implicitly factorizing a low-rank transformation of a network's normalized Laplacian matrix, and LINE, in theory, is a special case of DeepWalk.

In addition, we will also introduce how these network embedding techniques help real-world applications in Alibaba, AMiner, Microsoft Academic, Tencent Games, and WeChat. The covered applications include user profiling, behavior modeling, social influence, information propagation, and online recommendation.

Finally, we will summarize the problems and models introduced for network representation learning, and discuss the challenges and future directions. To facilitate open research for learning network representations, we will release large-scale public datasets and benchmarks during the tutorial. The to-be-released dataset for network representation learning is by far the largest Open Academic Graph composing of 200 million nodes and three billion links [18, 22].

The outline of the tutorial organization and arrangement is listed below. Each part is structured to introduce important problems, recent & advanced representation learning models, and direct applications from our collaborations with online/mobile social networking sites, which also forms as a unique value of this tutorial.

- I: Preliminaries for representation learning on networks
  - Real network problems in Alibaba, Tencent, etc.
  - Basic graph theories
  - Basic deep learning
  - Basic natural language processing (skip-gram)
  - Other related concepts
- II: Network representation learning algorithms
  - Skip-gram based network embedding models
  - Matrix factorization based embedding models
  - Graph neural networks
- III: Network embedding theory
  - Theories underlying skip-gram based models
  - Theories unifying factorization and graph convolution
  - Unified network embedding models
  - Fast embedding models
- IV: Network representation learning applications
  - Applications on Web user profiling
  - Applications on social influence on the Web
  - Applications on Web user behavior modeling
  - Applications on online recommendation
- V: Summary and open network embedding challenge
  - Review on algorithms and theories
  - Release the Open Benchmarks on Network Embedding
  - Release the Open Network Embedding Datasets.

In summary, this tutorial will provide an overview of network representation learning with the latest techniques and trends covered. One of its major goals is to generate theoretical insights into the rapid development of this field. The other one is to bridge network representation learning research with industry applications. Finally, to empower open and reproducible research, we will also release large-scale network datasets and benchmarks for representation learning.

REFERENCES

[1] Yoshua Bengio, Aaron Courville, and Pierre Vincent. 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 35, 8 (2013), 1798–1828.

[2] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and locally connected networks on graphs. In *ICLR '14*.

[4] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD '17*.

[5] Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla. 2014. Inferring User Demographics and Social Strategies in Mobile Social Networks. In *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14)*. ACM, 15–24.

[6] David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press.

[7] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1263–1272.

[8] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *KDD '16*. ACM, 855–864.

[9] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS'17*. 1024–1034.

[10] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation Learning on Graphs: Methods and Applications. *IEEE Data(base) Engineering Bulletin* 40 (2017), 52–74.

[11] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR '17*.

[12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[13] David Liben-Nowell and Jon Kleinberg. 2003. The Link Prediction Problem for Social Networks. In *ACM CIKM '03*. 556–559.

[14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *ICLR Workshop '13*.

[15] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *KDD '14*. ACM, 701–710.

[16] Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Chi Wang, Kuansan Wang, and Jie Tang. 2019. NetSMF: Large-Scale Network Embedding as Sparse Matrix Factorization. In *WWW '19*. n/a.

[17] Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. 2018. Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec. In *WSDM '18*. ACM, 459–467.

[18] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. 2015. An overview of microsoft academic service (mas) and applications. In *WWW '15 Companion*. ACM, 243–246.

[19] Jie Tang, Tiancheng Lou, and Jon Kleinberg. 2012. Inferring social ties across heterogenous networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 743–752.

[20] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *WWW '15*. 1067–1077.

[21] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social Influence Analysis in Large-scale Networks. In *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09)*. 807–816.

[22] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *KDD '08*. 990–998.

[23] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[24] Xian Wu, Baoxu Shi, Yuxiao Dong, Chao Huang, and Nitesh V Chawla. 2019. Neural Tensor Factorization for Temporal Interaction Learning. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 537–545.