

Learning From Networks: Algorithms, Theory, and Applications

Xiao Huang,[†] Peng Cui,[‡] Yuxiao Dong,[§]
Jundong Li,^{¶||} Huan Liu[¶]

[†]Texas A&M University, College Station, TX, USA 77843

[‡]Tsinghua University, Beijing, 100084, P. R. China

[§]Microsoft Research, Redmond, WA 98052, USA

[¶]Arizona State University, Tempe, AZ, USA 85281

^{||}University of Virginia, Charlottesville, VA, USA 22904
xhuang@tamu.edu; cuip@tsinghua.edu.cn;
yuxdong@microsoft.com; {jundongl, huan.liu}@asu.edu

Jian Pei,^{*} Le Song,[★] Jie Tang,[‡] Fei Wang,[†]
Hongxia Yang,^τ Wenwu Zhu[‡]

^{*}Simon Fraser University, Burnaby, BC, Canada

[★]Georgia Institute of Technology, Atlanta, GA, USA 30332

[‡]Tsinghua University, Beijing, 100084, P. R. China

[†]Cornell University, New York, NY, USA 10065

^τAlibaba Group, Hangzhou, China
jpei@cs.sfu.ca; le.song@antfin.com;
{jietang, wwzhu}@tsinghua.edu.cn;
few2001@med.cornell.edu; yang.yhx@alibaba-inc.com

ABSTRACT

Arguably, every entity in this universe is networked in one way or another. With the prevalence of network data collected, such as social media and biological networks, learning from networks has become an essential task in many applications. It is well recognized that network data is intricate and large-scale, and analytic tasks on network data become more and more sophisticated. In this tutorial, we systematically review the area of learning from networks, including algorithms, theoretical analysis, and illustrative applications. Starting with a quick recollection of the exciting history of the area, we formulate the core technical problems. Then, we introduce the fundamental approaches, that is, the feature selection based approaches and the network embedding based approaches. Next, we extend our discussion to attributed networks, which are popular in practice. Last, we cover the latest hot topic, graph neural based approaches. For each group of approaches, we also survey the associated theoretical analysis and real-world application examples. Our tutorial also inspires a series of open problems and challenges that may lead to future breakthroughs. The authors are productive and seasoned researchers active in this area who represent a nice combination of academia and industry.

ACM Reference Format:

Xiao Huang, Peng Cui, Yuxiao Dong, Jundong Li, Huan Liu, Jian Pei, Le Song, Jie Tang, Fei Wang, Hongxia Yang, Wenwu Zhu. 2019. Learning From Networks: Algorithms, Theory, and Applications. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19)*, August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3292500.3332293>

1 INTRODUCTION

The ACM 25th Conference on Knowledge Discovery and Data Mining (KDD) takes place in Anchorage, on August 4–8, 2019. It targets at exploring insightful understanding and solutions for real-world tasks in data mining, data science, and applied machine learning.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD '19, August 4–8, 2019, Anchorage, AK, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6201-6/19/08.

<https://doi.org/10.1145/3292500.3332293>

Complex networks are prevalent in these real-world tasks, and learning from networks has been a hot research topic with practical significance in many areas ranging from graph analysis, social networks, recommender systems, multimedia, to bioinformatics. It thus motivates us to organize this tutorial to systematically review the state-of-the-art learning algorithms for networks, and illustrate how they theoretically tackle real-world networks' data characteristics, such as non-Euclidean topologies and data variety.

The intended audience for this tutorial mainly includes students, researchers, and professionals who are new to this area or who already have some experience with data mining, machine learning, and network science. The audience is expected to have some basic understanding of data mining, machine learning, linear algebra, graph theory, and optimization. The tutorial will be presented at a college junior/senior level and could be comfortably followed by academic researchers and practitioners from the industry.

2 THEME AND TOPICS

Learning from networks has implications for multiple disciplines and it should be of interest to many of the conference attendees from students, researchers, to practitioners. Topic areas for this tutorial include (but are not limited to) the following:

- Challenges of analyzing and learning from networks
- Traditional ways of network analysis and limitations
- Recent trends of learning from networks
- Learn from networks with feature selection
- Network embedding
- Deep network embedding
- Dynamic & heterogeneous network embedding
- Attributed network embedding
- Graph neural networks
- Human-centric network analysis
- Network analysis within an open environment
- Billion-scale systems, applications, & challenges
- Random walk based network analysis

3 OBJECTIVES

Learning From Networks is a full-day tutorial. Its goals include:

First, audiences understand the basic concepts in network analysis, network embedding, attributed networks, feature engineering, graph neural networks, etc.

Second, audiences learn recent advances of learning from networks research and the state-of-the-art embedding and feature selection algorithms; These new methodologies and innovative approaches would pave the way for breakthroughs.

Third, audiences explore the rich research opportunities in learning from networks, and learn how to leverage embedding algorithms to handle real-world networks, in combination with the latest development in machine learning and data mining.

Fourth, it serves as a platform for participants to communicate their up-to-date research on learning from networks.

4 ORGANIZERS

Xiao Huang is a Ph.D. student in the Department of Computer Science & Engineering at Texas A&M University, supervised by Dr. Xia "Ben" Hu. He received M.S. from Illinois Institute of Technology in 2015, and B.S. from Shanghai Jiao Tong University in 2012. His research interests lie in network embedding, attributed network analysis, and knowledge graphs & analytics. He serves as a PC member of EASW 2018, KDD 2019, and CIKM 2019.

Peng Cui is an Associate Professor with tenure in Tsinghua University. He got his PhD degree from Tsinghua University in 2010. His research interests include social dynamic modeling, network representation learning, as well as causal inference and stable prediction. He has published more than 100 papers in prestigious conferences and journals in data mining and multimedia. His recent research got 5 paper awards from top-level international conferences and journals. He is the Associate Editors of TKDE, IEEE TBD, ACM TIST and ACM TOMM etc. He was the recipient of CCF-IEEE CS Young Scientist Award and ACM China Rising Star Award.

Yuxiao Dong is a Senior Applied Scientist at Microsoft Research, Redmond. He received his Ph.D. from University of Notre Dame. His research focuses on data mining, network science, and applied machine learning, with an emphasis on applying computational models to addressing problems in large-scale graph systems.

Jundong Li is currently a final year Ph.D. candidate of Computer Science and Engineering at Arizona State University. He will join University of Virginia as a tenure-track assistant professor from Fall 2019. His research interests are in data mining, machine learning, and social computing. He has published more than 40 papers in high-impact venues. He leads the development of an open-source feature selection repository and serves on program committees and reviews for major international conferences and journals.

Huan Liu is a professor of Computer Science and Engineering at Arizona State University. He got his Ph.D. from University of Southern California. He worked at Telecom Australia Research Labs and was on the faculty at National University of Singapore. He was recognized for excellence in teaching and research in CSE at ASU. His research interests are in data mining, machine learning, and social computing. He serves on journal editorial boards and many conference program committees, and is a founding organizer of the International Conference Series on Social Computing, Behavioral-Cultural Modeling, and Prediction. He is an ACM Fellow, an AAAI Fellow, an AAAS Fellow, and an IEEE Fellow.

Jian Pei is a Professor in the School of Computing Science, Simon Fraser University, Canada. He is renowned for his creative and productive research in the general areas of data science, big data, data mining, and database systems. He has published prolifically

and his publications have been cited over 84 thousand times. He is recognized as an ACM Fellow and an IEEE Fellow and several prestigious awards. In his recent leave-of-absence from the university, he acted as a Vice President of JD.com and a Technical VP of Huawei Technology. He has extensive experience in industry R&D, strategy consulting, and business operation, particularly in the areas of enterprise data platform, supply chain, and fintech.

Le Song is a Principal Engineer of Ant Financial, an Associate Professor in the College of Computing, and an Associate Director of the Center for Machine Learning, Georgia Institute of Technology. Before he joined GT in 2011, he was postdoc at Carnegie Mellon University and a research scientist at Google. His principal research direction is machine learning, nonlinear models, kernel methods, deep learning, and probabilistic graphical models. He is the recipient of the NSF CAREER Award14, and many best paper awards, including AISTATS'16 Best Student Paper Award, NIPS13 Outstanding Paper Award, and ICML10 Best Paper Award. He served as the area chair or senior program committee for many leading machine learning and AI conferences such as ICML, NIPS, AISTATS, AAAI and IJCAI, and the action editor for JMLR and IEEE TPAMI.

Jie Tang is the Full Professor and the vice chair of the Department of Computer Science and Technology at Tsinghua University. His interests include social network analysis, data mining, and machine learning. He served as PC Co-chair of CIKM'16 and WSDM'15, Associate General Chair of KDD'18, and acting Editor-in-Chief of ACM TKDD. He leads the project AMiner.org for academic social network analysis and mining. He was honored with UK Royal Society-Newton Advanced Fellowship Award, NSFC Distinguished Young Scholar, and ACM SIGKDD Service Award.

Fei Wang is an Associate Professor in Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, Cornell University. His major research interest is data mining, machine learning and their applications in health data science. His papers won 6 best paper awards in international conferences. He is the recipient of the NSF CAREER Award. He is the chair of the KDDM working group in AMIA. He is the general co-chair of ICHI 2018 track chair for Medinfo 2017 and program co-chair for CHASE 2018 and ICHI 2015.

Hongxia Yang is working as the Senior Staff Data Scientist and Director in Alibaba Group. Her interests span the areas of Bayesian statistics, time series analysis, spatial-temporal modeling, and survival analysis. She worked as the Principal Data Scientist at Yahoo! Inc and Research Staff Member at IBM T.J. Watson Research Center respectively and got her PhD degree in Statistics from Duke University in 2010. She has published over 40 top conference and journal papers. She is serving as the associate editor for Applied Stochastic Models in Business and Industry. She has been elected as an Elected Member of the International Statistical Institute in 2017.

Wenwu Zhu is with Computer Science Department of Tsinghua University as Professor of "1000 People Plan" of China. He was a Senior Researcher and Research Manager at Microsoft Research Asia. He was the Chief Scientist and the Director at Intel Research China from 2004 to 2008. He is an IEEE Fellow, SPIE Fellow and ACM Distinguished Scientist. He served(s) on various editorial boards. He served as TPC Co-chair of IEEE ISCAS 2013 and serves as TPC Co-chair for ACM Multimedia 2014.