Modeling and Mining Feature-Rich Networks

Rushed Kanawati
LIPN, University Paris 13
Paris, France
rushed.kanawati@lipn.fr

Martin Atzmueller
Tilburg University, Department of
Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
m.atzmuller@uvt.nl

ABSTRACT

In the field of web mining and web science, as well as data science and data mining there has been a lot of interest in the analysis of (social) networks. With the growing complexity of heterogeneous data, feature-rich networks have emerged as a powerful modeling approach: They capture data and knowledge at different scales from multiple heterogeneous data sources, and allow the mining and analysis from different perspectives. The challenge is to devise novel algorithms and tools for the analysis of such networks.

This tutorial provides a unified perspective on feature-rich networks, focusing on different modeling approaches, in particular multiplex and attributed networks. It outlines important principles, methods, tools and future research directions in this emerging field.

CCS CONCEPTS

• **Theory of computation** → **Graph algorithms analysis**; • **Human-centered computing** → *Social networks*; *Social media*; • **Information systems** → *World Wide Web*.

KEYWORDS

social media, social network analysis, social interaction networks, feature-rich networks, mining social networks

ACM Reference Format:
Rushed Kanawati and Martin Atzmueller. 2019. Modeling and Mining Feature-Rich Networks. 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/10.1145/3308560.3320098

1 INTRODUCTION

In the last years, networks have proven to be a useful tool for modeling structural complexity of a variety of complex systems in different domains including sociology, biology, ethology and computer science. In contrast to simple homogeneous static networks, networks in real world settings are often dynamic, heterogeneous, and both nodes and links can be described by a set of attributes and/or multiple relations. Thus, the mining of such *feature-rich* networks [20] is gaining increasing interest; such networks include, in particular, *node-attributed* and/or *link-attributed* networks, including multiplex networks. This tutorial provides a survey on the latest algorithmic advances for modeling and mining feature-rich networks, including both node and link attributed networks.

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

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3320098

We structure the materials as follows: First, we start with a brief introduction defining the field of feature-rich networks. Then, two main sections are presented : the first on modeling and mining *multiplex networks* and the second on mining *attributed networks*. Then, we present applications of the presented network mining methods, and brief overview on available analysis tools. We conclude with a summary and interesting research directions.

2 MODELING AND MINING MULTIPLEX NETWORKS

The concept of multiplex networks has been introduced with the goal to provide an expressive model for modeling real-world complex networks [11, 23]. A multiplex network is roughly defined as a multi-layer graph where each layer contains the same set of nodes but interconnected by different types of links. Nodes across different layers can also be coupled by special *coupling links*. Different types of coupling links can be defined allowing to model various types of networks including; multi-relational networks, dynamic networks or heterogeneous networks. Analysis of multiplex networks requires redefining most of the basic concepts and metrics usually used for plain complex networks including: nodal degree and neighborhood, paths, as well as different centrality functions. [9, 11, 27]. It also requires providing new algorithms to handle basic complex networks analysis tasks such as community detection [18] and link prediction [36]. We cover all these issues with a special focus on the central problem of community detection.

3 MODELING AND MINING ATTRIBUTED NETWORKS

Attributed networks, where nodes and/or links are labeled with additional information, allow further dimensions for network analysis and mining, e. g., [34, 43, 44]. Then, these enable novel methods and approaches for complex network analysis that directly work on the feature-rich attribute structure. These include approaches, for example, that specifically exploit the descriptive information of the labels assigned to nodes and/or edges of the graph. Nodes denoting actors, for example, can be labeled with properties of the latter for compositional subgroup analysis, which can be implemented using subgroup discovery [1, 2]. Furthermore, explicit descriptions for model building and explication can be provided, e. g.,using community detection methods e. g., [25, 33]. This also extends to link prediction [26]. Here, we provide an overview on recent algorithmic advances for modeling and mining attributed networks. We start with an introduction, providing basic definitions and introductory examples. After that, we tackle two specific methods: community detection, e. g., [3, 5, 6, 8, 10, 12, 17, 32, 35, 41, 42] and link prediction, e. g., [16, 21] on attributed networks.

4 APPLICATIONS AND TOOLS

There are different types of applications of the presented approaches for multiplex and attributed network mining. One important area is recommender systems, for which we show two applications: mining multiplex networks and attributed networks for enhancing tag recommendation in folksonomies [19], and mining multiplex networks for enhancing social interactions at social events [38, 39], respectively. Another original application of multiplex network mining includes an ensemble approach for cluster ensemble selection [22]. Furthermore, a summary of the available tools for multiplex and attributed network mining will be made. This includes a brief review of the Muna Library [14] and the VIKAMINE toolkit [4].

5 CONCLUSIONS

Network science is now mature enough to move towards analyzing and mining networks with much richer information than plain complex networks, modeled as *feature-rich complex networks*. Different extensions and future research directions can be identified: These include hybrid combinations of attributed, multiplex, and dynamic networks, e. g., [38], adaptations of Bayesian approaches for modeling and mining feature-rich networks [7, 13], extending and adapting network embedding methods and deep learning approaches for feature-rich networks, c. f., [15, 24, 37, 40] for some initial directions, and also including semantics into feature-rich network modeling and mining e. g., [28–31].

REFERENCES

[1] Martin Atzmueller. 2015. Subgroup Discovery. *WIREs DMKD* 5, 1 (2015), 35–49.

[2] Martin Atzmueller. 2018. Compositional Subgroup Discovery on Attributed Social Interaction Networks. In *Proc. DS*. Springer, Heidelberg, Germany.

[3] Martin Atzmueller, Stephan Doerfel, and Folke Mitzlaff. 2016. Description-Oriented Community Detection using Exhaustive Subgroup Discovery. *Information Sciences* 329 (2016), 965–984.

[4] Martin Atzmueller and Florian Lemmerich. 2012. VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In *Proc. ECML/PKDD (LNAI)*, Vol. 7524. Springer, Heidelberg, Germany.

[5] Martin Atzmueller and Florian Lemmerich. 2013. Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. *International Journal of Web Science* 2, 1/2 (2013), 80–112.

[6] Martin Atzmueller and Folke Mitzlaff. 2011. Efficient Descriptive Community Mining. In *Proc. FLAIRS*. AAAI, 459 – 464.

[7] Martin Atzmueller, Andreas Schmidt, Benjamin Kloepper, and David Arnu. 2017. HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses. In *New Frontiers in Mining Complex Patterns (LNAI)*, Vol. 10312. Springer, Heidelberg, Germany.

[8] Martin Atzmueller, Henry Soldano, Guillaume Santini, and Dominique Bouthinon. 2018. MinerLSD: Efficient Local Pattern Mining on Attributed Graphs. In *Proc. 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*.

[9] Federico Battiston, Vincenzo Nicosia, and Vito Latora. 2013. Metrics for the analysis of multiplex networks. *CoRR* abs/1308.3182 (2013).

[10] Cécile Bothorel, Juan David Cruz, Matteo Magnani, and Barbora Micenkova. 2015. Clustering Attributed Graphs: Models, Measures and Methods. *Network Science* 3, 3 (2015), 408–444.

[11] Peter Brodka and Przemysaw Kazienko. 2014. *Encyclopedia of Social Network Analysis and Mining*. Springer, Chapter Multi-layered social networks.

[12] David Combe, Christine Largeron, Mathias Géry, and Elöd Egyed-Zsigmond. 2015. I-Louvain: An Attributed Graph Clustering Method. In *Proc. IDA*. 181–192.

[13] Lisette Espín-Noboa, Florian Lemmerich, Markus Strohmaier, and Philipp Singer. 2017. JANUS: A Hypothesis-Driven Bayesian Approach for Understanding Edge Formation in Attributed Multigraphs. *Applied Network Science* 2, 1 (2017), 16.

[14] Issam Falih and Rushed Kanawati. 2015. Muna: A Multiplex network analysis Library. In *The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Paris, 757–760.

[15] Hongchang Gao and Heng Huang. 2018. Deep Attributed Network Embedding. In *Proc. IJCAI*. 3364–3370.

[16] Neil Zhenqiang Gong, Ameet Talwalkar, Lester Mackey, Ling Huang, Eui Chul Richard Shin, Emil Stefanov, Elaine (Runting) Shi, and Dawn Song. 2014. Joint Link Prediction and Attribute Inference Using a Social-Attribute Network. *ACM Trans. Intell. Syst. Technol.* 5, 2 (2014), 27:1–27:20.

[17] Stephan Günnemann, Ines Färber, Brigitte Boden, and Thomas Seidl. 2013. GAMer: A Synthesis of Subspace Clustering and Dense Subgraph Mining. In *Knowledge and Information Systems*. Springer.

[18] Manel Hmimida and Rushed Kanawati. 2015. Community detection in multiplex networks: a seed-centric approach. *Networks and Heterogeneous Media* 10, 1 (March 2015), 71–85.

[19] Manel Hmimida and Rushed Kanawati. 2016. A graph-coarsening approach for tag recommendation. In *25th international World Wide Web Conference*. 43–44.

[20] Roberto Interdonato, Martin Atzmueller, Sabrina Gaito, Rushed Kanawati, Christine Largeron, and Alessandra Sala. 2019. Feature-rich networks: going beyond complex network topologies. *Applied Network Science* 4, 1 (2019), 4:1–4:13. https://doi.org/10.1007/s41109-019-0111-x

[21] Maosheng Jiang, Yonxiang Chen, and Ling Chen. 2015. Link Prediction in Networks with Node Attributes by Similarity Propagation. *arXiv preprint arXiv:1502.04380* (2015).

[22] R. Kanawati. 2019. Cluster-ensemble selection: an ensemble approach. In *Proc. SCML2019*. Wuhan, China.

[23] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. 2014. Multilayer Networks. *Journal of complex networks* 2 (2014), 203–271. http://arxiv.org/abs/1309.7233

[24] Jihwan Lee and Sunil Prabhakar. 2018. A3embed: Attribute Association Aware Network Embedding. In *Proc. WWW (Companion)*. IW3C2, 1243–1251.

[25] Jure Leskovec, Kevin J. Lang, and Michael Mahoney. 2010. Empirical Comparison of Algorithms for Network Community Detection. In *WWW*. ACM, 631–640.

[26] David Liben-Nowell and Jon M. Kleinberg. 2007. The link-prediction problem for social networks. *JASIST* 58, 7 (2007), 1019–1031.

[27] Matteo Magnani and Moreno Marzolla. 2014. Path-Based and Whole-Network Measures. In *Encyclopedia of Social Network Analysis and Mining*. 1256–1269. https://doi.org/10.1007/978-1-4614-6170-8_241

[28] Spyroula Masiala and Martin Atzmueller. 2018. First Perspectives on Explanation in Complex Network Analysis. In *Proc. BNAIC*. Jheronimus Academy of Data Science, Den Bosch, The Netherlands.

[29] Spyroula Masiala and Martin Atzmueller. 2018. Towards Explainable Complex Network Analysis". In *Proc. Dutch-Belgian Database Day*. Hasselt University, Belgium.

[30] Folke Mitzlaff, Martin Atzmueller, Andreas Hotho, and Gerd Stumme. 2014. The Social Distributional Hypothesis. *Journal of Social Network Analysis and Mining* 4, 216 (2014).

[31] Folke Mitzlaff, Martin Atzmueller, Gerd Stumme, and Andreas Hotho. 2013. Semantics of User Interaction in Social Media. In *Complex Networks*. Springer, Heidelberg, Germany.

[32] Flavia Moser, Recep Colak, Arash Rafiey, and Martin Ester. 2009. Mining Cohesive Patterns from Graphs with Feature Vectors. In *SDM*, Vol. 9. SIAM, 593–604.

[33] M. E. J. Newman. 2006. Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences* 103, 23 (2006), 8577–8582.

[34] Joseph J Pfeiffer III, Sebastian Moreno, Timothy La Fond, Jennifer Neville, and Brian Gallagher. 2014. Attributed Graph Models: Modeling Network Structure with Correlated Attributes. In *Proceedings of the 23rd international conference on World wide web*. ACM, 831–842.

[35] S. Pool, F. Bonchi, and M. van Leeuwen. 2014. Description-driven Community Detection. *TIST* 5, 2 (2014).

[36] Manisha Pujari and Rushed Kanawati. 2017. *Interactions in complex systems*. Cambridge Scholar Publishing, Chapter Link prediction in Large-scale multiplex networks.

[37] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling Relational Data with Graph Convolutional Networks. In *European Semantic Web Conference*. Springer, 593–607.

[38] Christoph Scholz, Martin Atzmueller, Alain Barrat, Ciro Cattuto, and Gerd Stumme. 2013. New Insights and Methods For Predicting Face-To-Face Contacts. In *Proc. AAAI ICWSM*. AAAI Press, Palo Alto, CA, USA.

[39] Christoph Scholz, Martin Atzmueller, and Gerd Stumme. 2012. On the Predictability of Human Contacts: Influence Factors and the Strength of Stronger Ties. In *Proc. SocialCom*. IEEE Computer Society, Boston, MA, USA.

[40] Nasrullah Sheikh, Zekarias Kefato, and Alberto Montresor. 2018. GAT2VEC: Representation Learning for Attributed Graphs. *Computing* (2018), 1–23.

[41] Arlei Silva, Wagner Meira Jr, and Mohammed J Zaki. 2012. Mining Attribute-Structure Correlated Patterns in Large Attributed Graphs. *Proc. VLDB Endowment* 5, 5 (2012), 466–477.

[42] Jaewon Yang, Julian McAuley, and Jure Leskovec. 2013. Community Detection in Networks with Node Attributes. *2013 IEEE ICDM* (2013), 1151–1156. arXiv:1401.7267

[43] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. 2010. LINKREC: A Unified Framework for Link Recommendation with User Attributes and Graph Structure. In *Proc. WWW*. 1211–1212.

[44] Yang Zhou, Hong Cheng, and Jeffrey Xu Yu. 2009. Graph Clustering Based on Structural/Attribute Similarities. *Proc. VLDB Endow.* 2, 1 (2009), 718–729.