Mining Dynamic Networks with Generative Models

Kevin S. Xu* James R. Foulds†

Abstract

Traditional data mining algorithms for networks typically assume that the network is a single static graph, and thus, ignore temporal dynamics. Probabilistic generative models provide a natural framework for reasoning collectively over structured data and are well-suited for analyzing dynamic network data. This tutorial presents recent advances in probabilistic generative models for dynamic networks, including both discrete-time and continuous-time networks.

1 Introduction

Traditional data mining algorithms for networks typically assume that the network is in the form of a single static graph, which either represents a single time snapshot or an aggregate view over time. However, this is often an oversimplification of the underlying network phenomenon. For example, in the age of social media, many of our social interactions are recorded digitally with timestamps, and the temporal dynamics can yield as much insight as the graph structure. Thus, it has become increasingly clear that network mining algorithms need to go beyond a single graph to include time information. *Probabilistic generative models* are well-suited for such analyzing such rich network data, as they provide a natural framework for reasoning collectively over structured data.

This tutorial presents recent advances in generative models for dynamic networks, focusing on models that encode network phenomena with latent (i.e. hidden) attributes, which are subsequently recovered from data. We consider dynamic networks in two different forms: *discrete-time* networks, where the network evolves through a set of "snapshots" observed at discrete time steps, and *continuous-time* networks, where the network is observed through relational events at arbitrary, irregularly-spaced timestamps. Our focus will be on continuous-time networks, which differ most from static networks due to the absence of graph snapshots. For such networks, model-based analysis using point process models can reveal significant insights about the frequency, burstiness, and ordering of events in networks that are not possible to uncover using snapshots.

*University of Toledo, Email: `kevin.xu@utoledo.edu`

†University of Maryland, Baltimore County, Email: `jfoulds@umbc.edu`

2 Tutorial Outline

We divide this tutorial into two parts, with the first focused on models for mining static networks, and the second focused on models for mining different types of dynamic networks. All tutorial materials are available at the tutorial website: `https://github.com/kevin-s-xu/SDM-2021-Generative-Tutorial`.

Part 1: We begin with a presentation on the fundamentals of probabilistic generative models for network data, including frequentist and Bayesian approaches for fitting such models. We will cover latent space models [7], stochastic block models [11], mixed-membership models [1], and more recent node embedding-based models [5]. We discuss the relationships between the model assumptions and sociological principles such as homophily and stochastic equivalence. We then move into a demonstration session on fitting these network models to a real online social network data set using Python.

Part 2: We introduce both discrete-time [9, 12] and continuous-time dynamic networks [2–4, 8, 10, 13], with examples of different application settings and data sets where each would apply. We also introduce the notion of indirectly-observed continuous-time networks, where edges have to be inferred from timestamped observations at the nodes, potentially including additional data such as text [6]. We then present rich generative models for these types of dynamic networks, with an emphasis on continuous-time networks where there is no defined graph snapshot time, and relate these models to typical graph mining tasks from the static network setting. We conclude with a demo fitting several dynamic network models to the same online social network data used in Part 1, but with the inclusion of time information, and compare the findings to the static network models.

3 Presenter Bios

Kevin S. Xu received the B.A.Sc. degree in Electrical Engineering from the University of Waterloo in 2007 and the M.S.E. and Ph.D. degrees in Electrical Engineering: Systems from the University of Michigan in

2009 and 2012, respectively. He was a recipient of the Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Master's and Doctorate Scholarships. He is currently an assistant professor in the EECS Department at the University of Toledo and has previously held industry research positions at Technicolor and 3M. His main research interests are in machine learning and network science with applications to human dynamics, health care, education, and wearable computing. Website: `http://kevinsxu.com`

James R. Foulds (a.k.a. Jimmy) is an assistant professor in the Department of Information Systems at the University of Maryland, Baltimore County. His research interests are broadly in the area of socially conscious machine learning and artificial intelligence, including probabilistic latent variable models and the inference algorithms to learn them from social networks and text data. His work aims to promote the practice of latent variable modeling for applied research in disciplines such as computational social science and the digital humanities, and to improve AI's role in society regarding fairness and privacy. He earned his PhD in computer science at the University of California, Irvine, and was a postdoctoral scholar at the University of California, San Diego, and the University of California, Santa Cruz. His master's and bachelor's degrees were earned with first class honours at the University of Waikato, New Zealand, where he also contributed to the Weka data mining system. Website: `http://jfoulds.informationsystems.umbc.edu`

4 Target Audience

We aim to serve a multidisciplinary audience, including researchers and practitioners from the mathematical, computer, and information sciences, and potentially the social sciences. We will assume a relatively low technical background and will not assume familiarity with statistical modeling or network science. Some knowledge of basic probability and linear algebra would be helpful in understanding the tutorial content.

Benefit to Participants: By the end of the tutorial, attendees will be able to

1. Compare and contrast the modeling approaches used by several generative models for dynamic network data.

2. Discuss differences in approaches required for modeling and mining discrete- and continuous-time network data.

3. Identify applications of generative models to their own research questions on dynamic networks.

References

[1] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, *Mixed membership stochastic blockmodels*, Journal of Machine Learning Research, 9 (2008), pp. 1981–2014.

[2] M. Arastuie, S. Paul, and K. S. Xu, *CHIP: A Hawkes process model for continuous-time networks with scalable and consistent estimation*, in Advances in Neural Information Processing Systems 33, 2020.

[3] C. Blundell, K. A. Heller, and J. Beck, *Modelling reciprocating relationships with Hawkes processes*, in Advances in Neural Information Processing Systems 25, 2012, pp. 2600–2608.

[4] C. DuBois, C. T. Butts, and P. Smyth, *Stochastic blockmodeling of relational event dynamics*, in Proceedings of the 16th International Conference on Artificial Intelligence and Statistics, 2013, pp. 238–246.

[5] A. Grover and J. Leskovec, *node2vec: Scalable Feature Learning for Networks*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.

[6] X. He, T. Rekatsinas, J. Foulds, L. Getoor, and Y. Liu, *HawkesTopic: A joint model for network inference and topic modeling from text-based cascades*, in Proceedings of the 32nd International Conference on Machine Learning, 2015, pp. 871–880.

[7] P. D. Hoff, A. E. Raftery, and M. S. Handcock, *Latent space approaches to social network analysis*, Journal of the American Statistical Association, 97 (2002), pp. 1090–1098.

[8] R. R. Junuthula, M. Haghdan, K. S. Xu, and V. K. Devabhaktuni, *The Block Point Process Model for continuous-time event-based dynamic networks*, in Proceedings of the World Wide Web Conference, 2019, pp. 829–839.

[9] C. Matias and V. Miele, *Statistical clustering of temporal networks through a dynamic stochastic block model*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79 (2017), pp. 1119–1141.

[10] X. Miscouridou, F. Caron, and Y. W. Teh, *Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data*, in Advances in Neural Information Processing Systems 31, 2018, pp. 2343–2352.

[11] K. Nowicki and T. A. B. Snijders, *Estimation and prediction for stochastic blockstructures*, Journal of the American Statistical Association, 96 (2001), pp. 1077–1087.

[12] K. S. Xu and A. O. Hero III, *Dynamic stochastic blockmodels for time-evolving social networks*, IEEE Journal of Selected Topics in Signal Processing, 8 (2014), pp. 552–562.

[13] J. Yang, V. Rao, and J. Neville, *Decoupling homophily and reciprocity with latent space network models*, in Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2017.