A Survey of Statistical Network Models

A Survey of Statistical Network Models

Anna Goldenberg

University of Toronto
Canada
anna.goldenberg@utoronto.ca

Alice X. Zheng

Microsoft Research
USA
alicez@microsoft.com

Stephen E. Fienberg

Carnegie Mellon University
USA
fienberg@stat.cmu.edu

Edoardo M. Airoldi

Harvard University
USA
airoldi@fas.harvard.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Goldenberg, A. X. Zheng, S. E. Fienberg and E. M. Airoldi, A Survey of Statistical Network Models, Foundation and Trends® in Machine Learning, vol 2, no 2, pp 129–233, 2009

ISBN: 978-1-60198-320-6

Foundations and Trends® in Machine Learning

Volume 2 Issue 2, 2009

Editorial Scope

Foundations and Trends® in Machine Learning will publish survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians
Foundations and Trends® in Machine Learning, 2009, Volume 2, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends® in
Machine Learning
Vol. 2, No. 2 (2009) 129–233

DOI: 10.1561/2200000005

A Survey of Statistical Network Models

Anna Goldenberg[1], Alice X. Zheng[2], Stephen E. Fienberg[3] and Edoardo M. Airoldi[4]

[1] *Center for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, M5S 3G4, Canada, anna.goldenberg@utoronto.ca*
[2] *Microsoft Research, One Microsoft Way, Redmond, WA, 98052, USA, alicez@microsoft.com*
[3] *Department of Statistics, Machine Learning Department, Cylab, and iLab Carnegie Mellon University, Pittsburgh, PA, 15213-3890, USA, fienberg@stat.cmu.edu*
[4] *Department of Statistics & FAS Center for Systems Biology, Harvard University, 1 Oxford Street, Cambridge, MA, 02138, USA, airoldi@fas.harvard.edu*

Abstract

Networks are ubiquitous in science and have become a focal point for discussion in everyday life. Formal statistical models for the analysis of network data have emerged as a major topic of interest in diverse areas of study, and most of these involve a form of graphical representation. Probability models on graphs date back to 1959. Along with empirical studies in social psychology and sociology from the 1960s, these early works generated an active "network community" and a substantial literature in the 1970s. This effort moved into the statistical literature in the late 1970s and 1980s, and the past decade has seen a burgeoning network literature in statistical physics and computer science. The growth

of the World Wide Web and the emergence of online "networking communities" such as *Facebook*, *MySpace*, and *LinkedIn*, and a host of more specialized professional network communities has intensified interest in the study of networks and network data.

Our goal in this review is to provide the reader with an entry point to this burgeoning literature. We begin with an overview of the historical development of statistical network modeling and then we introduce a number of examples that have been studied in the network literature. Our subsequent discussion focuses on a number of prominent static and dynamic network models and their interconnections. We emphasize formal model descriptions, and pay special attention to the interpretation of parameters and their estimation. We end with a description of some open problems and challenges for machine learning and statistics.

Contents

1

Introduction

Many scientific fields involve the study of networks in some form. Networks have been used to analyze interpersonal social relationships, communication networks, academic paper co-authorships and citations, protein interaction patterns, and much more. Popular books on networks and their analysis began to appear a decade ago (see, e.g., [24, 50, 68, 318, 319]), and online "networking communities" such as *Facebook*, *MySpace*, and *LinkedIn* are an even more recent phenomenon.

In this work, we survey selective aspects of the literature on statistical modeling and analysis of networks in social sciences, computer science, physics, and biology. Given the volume of books, papers, and conference proceedings published on the subject in these different fields, a single comprehensive survey would be impossible. Our goal is far more modest. We attempt to chart the progress of statistical modeling of network data over the past 70 years and to outline succinctly the major schools of thought and approaches to network modeling and to describe some of their interconnections. We also attempt to identify major statistical gaps in these modeling efforts. From this overview one might then synthesize and deduce promising future research directions. Kolaczyk [177] provides a complementary statistical overview.

The existing set of statistical network models may be organized along several major axes. For this monograph, we choose the axis of static vs. dynamic models. Static network models concentrate on explaining the observed set of links based on a single snapshot of the network, whereas dynamic network models are often concerned with the mechanisms that govern changes in the network over time. Most early examples of networks were single static snapshots. Hence static network models have been the main focus of research for many years. However, with the emergence of online networks, more data are available for dynamic analysis, and in recent years there has been growing interest in dynamic modeling.

In the remainder of this chapter we provide a brief historical overview of network modeling approaches. In subsequent chapters we introduce some examples studied in the network literature and give a more detailed comparative description of select modeling approaches.

1.1 Overview of Modeling Approaches

Almost all of the "statistically" oriented literature on the analysis of networks derives from a handful of seminal papers. In social psychology and sociology there is the early work of Simmel and Woff [268] at the turn of the last century and Moreno [222] in the 1930s as well as the empirical studies of Stanley Milgram [216, 298] in the 1960s; in mathematics/probability there is the Erdös–Rényi paper on random graph models [94]. There are other papers that dealt with these topics contemporaneously or even earlier. But these are the ones that appear to have had lasting impact.

Moreno [222] invented the sociogram — a diagram of points and lines used to represent relations among persons, a precursor to the graph representation for networks. Luce and others developed a mathematical structure to go with Moreno's sociograms using incidence matrices and graphs (see, e.g., [11, 200, 201, 202, 203, 244, 282]), but the structure they explored was essentially deterministic. Milgram gave the name to what is now referred to as the "Small-World" phenomenon — short paths of connections linking most people in social spheres — and his experiments had provocative results: the shortest path between any two people for completed chains has a median length of around 6; however, the majority of chains initiated in his experiments were never completed! (His studies provided the title for the play and movie *Six Degrees of Separation*, ignoring the complexity of his results due to the censoring.) White [321] and Fienberg and Lee [100] gave a formal Markov chain like model and analysis of the Milgram experimental data, including information on the uncompleted chains. Milgram's data were gathered in batches of transmission, and thus these models can be thought of as representing early examples of generative descriptions of dynamic network evolution. Recently, Dodds et al. [86] studied a global "replication" variation on the Milgram study in which more than 60,000 e-mail users attempted to reach one of 18 target persons in 13 countries by forwarding messages to acquaintances. Only 384 of 24,163 chains reached their targets but they estimate the median length for completions to be 7, by assuming that attrition occurs at random.

The social science network research community that arose in the 1970s was built upon these earlier efforts, in particular the Erdös–Rényi–Gilbert model. Research on the Erdös–Rényi–Gilbert model (along with works by Katz et al. [166, 167, 168]) engendered the field of random graph theory. In their papers, Erdös and Rényi worked with fixed number of vertices, N, and number of edges, E, and studied the properties of this model as E increases. Gilbert studied a related two-parameter version of the model, with N as the number of vertices and p the fixed probability for choosing edges. Although their descriptions might at first appear to be static in nature, we could think in terms of adding edges sequentially and thus turn the model into a dynamic one. In this alternative binomial version of the Erdös–Rényi–Gilbert model, the key to asymptotic behavior is the value $\lambda = pN$. There is a "phase change" associated with the value of $\lambda = 1$, at which point we shift from seeing many small connected components in the form of trees to the emergence of a single "giant connected component." Probabilists such as Pittel [243] imported ideas and results from stochastic processes into the random graph literature.

Holland and Leinhardt's [150] p_1 model extended the Erdös–Rényi–Gilbert model to allow for differential attraction (popularity) and expansiveness, as well as an additional effect due to reciprocation.

The p_1 model was log-linear in form, which allowed for easy computation of maximum likelihood estimates using a contingency table formulation of the model [103, 104]. It also allowed for various generalizations to multidimensional network structures [101] and stochastic blockmodels. This approach to modeling network data quickly evolved into the class of p^* or exponential random graph models (ERGMs) originating in the work of Frank and Strauss [110] and Strauss and Ikeda [287]. A trio of papers demonstrating procedures for using ERGMs [241, 254, 316] led to the widespread use of ERGMs in a descriptive form for cross-sectional network structures or cumulative links for networks — what we refer to here as static models. Full maximum likelihood approaches for ERGMs appeared in the work of Snijders and Handcock and their collaborators, some of which we describe in Section 3.

Most of the early examples of networks in the social science literature were relatively small (in terms of the number of nodes) and involved the study of the network at a fixed point in time or cumulatively over time. Only a few studies (e.g., Sampson's 1968 data on novice monks in the monastery [259]) collected, reported, and analyzed network data at multiple points in time so that one could truly study the evolution of the network, i.e., network dynamics. The focus on relatively small networks reflected the state-of-art of computation, but it was sufficient to trigger the discussion of how one might assess the fit of a network model. Should one focus on "small sample" properties and exact distributions given some form of minimal sufficient statistic, as one often did in other areas of statistics, or should one look at asymptotic properties, where there is a sequence of networks of increasing size? Even if we have "repeated cross-sections" of the network, if the network is truly evolving in continuous-time we need to ask how to ensure that the continuous-time parameters are estimable. We return to many of these question in subsequent sections.

In the late 1990s, physicists began to work on network models and study their properties in a form similar to the macro-level descriptions of statistical physics. Barabási, Newman, and Watts, among others, produced what we can think of as variations on the Erdös–Rényi–Gilbert model which either controlled the growth of the network or allowed for differential probabilities for edge addition and/or deletion. These variations were intended to produce phenomena such as "hubs," "local clustering," and "triadic closures." The resulting models gave us fixed degree distribution limits in the form of power-laws — variations on preferential attachment models ("the rich get richer") that date back to Yule [329] and Simon [269] (see also [219]) — as well as what became known as "small-world" models. The small-world phenomenon, which harks back to Milgram's 1960s studies, usually refers to two distinct properties: (1) small average distance and (2) the "clustering" effect, where two nodes with a common neighbor are more likely to be adjacent. Many of these authors claim that these properties are ubiquitous in realistic networks. To model networks with the small-world phenomenon, it is natural to utilize randomly generated graphs with a power-law degree distribution, where the fraction of nodes with degree k is proportional to k^{-a} for some positive exponent a. Many of the most relevant papers are included in an edited collection by Newman et al. [204]. More recently this style of statistical physics models has been used to detect community structure in networks, e.g., see Girvan and Newman [122] and Backstrom et al. [20], a phenomenon which has its counterpart description in the social science network modeling literature.

The probabilistic literature on random graph models from the 1990s made the link with epidemics and other evolving stochastic phenomena. Picking up on this idea, Watt and Strogatz [320] and others used epidemic models to capture general characteristics of the evolution of these new variations on random networks. Durrett [91] has provided us with a book-length treatment on the topic with a number of interesting variations on the theme. The appeal of stochastic processes as descriptions of dynamic network models comes from being able to exploit the extensive literature already developed, including the existence and the form of stationary distributions and other model features or properties. Chung and Lu [69] provide a complementary treatment of these models and their probabilistic properties.

One of the principal problems with this diverse network literature that we see is that, with some notable exceptions, the statistical tools for estimation and assessing the fit of "statistical physics" or stochastic process models are lacking. Consequently, no attention is paid to the fact that real data may often be biased and noisy. What authors in the network literature have often relied upon is the extraction of key features of the related graphical network representation, e.g., the use of power-laws to represent degree distributions or measures of centrality and clustering, without any indication that they are either necessary or sufficient as descriptors for the actual network data. Moreover, these summary quantities can often be highly misleading as the critique by Stouffer et al. [285, 286] of methods used by Barabási [25] and Vázquez et al. [304] suggest. Barabási claimed that the dynamics of a number of human activities are scale-free, i.e., he specifically reported that the probability distribution of time intervals between consecutive e-mails sent by a single user and time delays for e-mail replies follow a power-law with exponent -1, and he proposed a priority-queuing process as an explanation of the bursty nature of human activity. Stouffer et al. [286] demonstrated that the reported power-law distribution was solely an artifact of the analysis of the empirical data and used Bayes factors to show that the proposed model is not representative of e-mail communication patterns. See a related discussion of the poor fit of power-laws in Clauset et al. [74]. There are several works, however, that try to address model fitting and model comparison. For example, the work of Williams and Martinez [323] showed how a simple two-parameter model predicted "key structural properties of the most complex and comprehensive food webs in the primary literature". Another good example is the work of Middendorf et al. [215] where the authors used network motif counts as input to a discriminative systematic classification for deciding which configuration model the actual observed network came from; they looked at power-law, small-world, duplication-mutation and duplication-mutation-complementation and other models (seven in total) and concluded that the duplication-mutation-complementation model described the protein–protein interaction data in *Drosophila melanogaster* species best.

Machine learning approaches emerged in several forms over the past decade with the empirical studies of Faloutsos et al. [97] and Kleinberg [172, 173, 174], who introduced a model for which the underlying graph is a grid — the graphs generated do not have a power-law degree distribution, and each vertex has the same expected degree. The strict requirement that the underlying graph be a cycle or grid renders the model inapplicable to webgraphs or biological networks. Durrett [91] treats variations on this model as well. More recently, a number of authors have looked to combine the stochastic blockmodel ideas from the 1980s with latent space models, model-based clustering [137] or mixed-membership models [7], to provide generative models that scale in reasonable ways to substantial-sized networks. The class of mixed membership models resembles a form of soft clustering [95] and includes the latent Dirichlet allocation model [41] from machine learning as a special case. This class of models offers much promise for the kinds of network dynamical processes we discuss here.

1.2 What This Survey Does Not Cover

This survey focuses primarily on statistical network models and their applications. As a consequence there are a number of topics that we touch upon only briefly or essentially not at all, such as:

- *Probability theory associated with random graph models.* The probabilistic literature on random graph models is now truly extensive and the bulk of the theorems and proofs, while interesting in their own right, are largely unconnected with the present exposition. For excellent introductions to this literature, see Chung and Lu [69] and Durrett [91]. For related results on the mathematics of graph theory, see Bollobás [43].
- *Efficient computation on networks.* There is a substantial computer science literature dealing with efficient calculation of quantities associated with network structures, such as shortest paths, network diameter, and other measures of connectivity, centrality, clustering, etc. The edited volume by Brandes and Erlebach [48] contains good overviews of a number of these topics as well as other computational issues associated with the study of graphs.
- *Use of the network as a tool for sampling.* Adaptive sampling strategies modify the sampling probabilities of selection based on observed values in a network structure.

This strategy is beneficial when searching for rare or clustered populations. Thompson and Seber [296] and Thompson [293] discuss adaptive sampling in detail. There is also related work on target sampling [294] and respondent-driven sampling [258, 305].

- *Neural networks.* Neural networks originated as simple models for connections in the brain but have more recently been used as a computational tool for pattern recognition (e.g., Bishop [38]), machine learning (e.g., Neal [229]), and models of cognition (e.g., Rogers and McClelland [257]).
- *Networks and economic theory.* A relatively new area of study is the link between network problems, economic theory, and game theory. Some useful entrees to this literature are Even-Dar and Kearns [96], Goyal [131], Kearns et al. [169], and Jackson [160], whose book contains an excellent semi-technical introduction to network concepts and structures.
- *Relational networks.* This is a very popular area in machine learning. It uses probabilistic graphical models to represent uncertainty in the data. The types of "networks" in this area, such as Bayes nets, dependency diagrams, etc., have a different meaning than the networks we consider in this review. The main difference is that the networks in our work are considered to "be given" or arising directly from properties of the network under study, rather than being representative of the uncertainty of the relationships between nodes and node attributes. There is a multitude of literature on relational networks, e.g., see Friedman et al. [112], Getoor et al. [116], Neville and Jensen [230], Neville et al. [231], and Getoor and Taskar [117].
- *Bipartite graphs.* These are graphs that represent measurement on two populations of objects, such as individuals and features. The graphs in this context are seldom the best representation of the data, with exception perhaps of binary measurements or when the true populations have comparable sizes. Recent work on exchangeable Rasch matrices is related to this topic and potentially relevant for network analysis.

Lauritzen [186, 187], Bassetti et al. [29] suggest applications to bipartite graphs.

- *Agent-based modeling.* Building on older ideas such as cellular automata, agent-based modeling attempts to simulate the simultaneous operations of multiple agents, in an effort to re-create and predict the actions of complex phenomena. Because the interest is often on the interaction among the agents, this domain of research has been linked with network ideas. With the recent advances in high-performance computing, simulations of large-scale social systems have become an active area of research, e.g., see [46]. In particular, there is a strong interest in areas that revolve around national security and the military, with studies on the effects of catastrophic events and biological warfare, as well as computational explorations of possible recovery strategies [56, 58]. These works are the contemporary counterparts of more classical work at the interface between artificial intelligence and the social sciences [54, 55, 57].

References

[1] E. M. Airoldi, "Bayesian mixed membership models of complex and evolving networks," PhD thesis, School of Computer Science, Carnegie Mellon University, 2006.

[2] E. M. Airoldi, "Getting started in probabilistic graphical models," *PLoS Computational Biology*, vol. 3, no. 12, p. e252, 2007.

[3] E. M. Airoldi, "Model-based clustering for social networks: Discussion," *Journal of the Royal Statistical Society, Series A*, vol. 170, no. 2, pp. 330–331, 2007.

[4] E. M. Airoldi, "The exchangeable graph model," Technical report 1, Department of Statistics, Harvard University, 2009.

[5] E. M. Airoldi, "A family of distributions on the unit hypercube," Technical report 2, Department of Statistics, Harvard University, 2009.

[6] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership analysis of high-throughput interaction studies: Relational data," http://arXiv.org/abs/0706.0294, 2007.

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

[8] E. M. Airoldi, D. M. Blei, E. P. Xing, and S. E. Fienberg, "A latent mixed-membership model for relational data," in *Proceedings of the 3rd International Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD '05),* in conjunction with *the 11th International ACM SIGKDD Conference*, pp. 82–89, New York: ACM Press, 2005.

[9] E. M. Airoldi and K. M. Carley, "Sampling algorithms for pure network topologies: A study on the stability and the separability of metric embeddings," *ACM SIGKDD Explorations*, vol. 7, no. 2, pp. 13–22, 2005.

[10] L. Akoglu and C. Faloutsos, "RTG: A recursive realistic graph generator using random typing," in *Data Mining and Knowledge Discovery,* 19(2):194–209, Springer Netherlands, 2009.

[11] R. D. Alba, "A graph-theoretic definition of a sociometric clique," *Journal of Mathematical Sociology*, vol. 3, pp. 113–126, 1973.

[12] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[13] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world wide web," *Nature*, vol. 401, pp. 130–131, 1999.

[14] D. L. Alderson, "Catching the 'Network Science' Bug: Insight and opportunity for the operations researcher," *Operations Research*, vol. 56, no. 5, pp. 1047–1065, 2008.

[15] D. J. Aldous, "Exchangeability and related topics," in *Lecture Notes in Mathematics*, vol. 1117, pp. 1–198, Springer Berlin/Heidelberg, 1985. (Also in Ecole d'Ete St Flour 1983).

[16] S. Allesina, D. Alonso, and M. Pascual, "A general model for food web structure," *Science*, vol. 320, no. 5876, pp. 658–661, 2008.

[17] U. Alon, "Network motifs: Theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, pp. 450–461, 2007.

[18] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11149–11152, 2000.

[19] P. Arabie, S. A. Boorman, and P. R. Levitt, "Constructing blockmodels: How and why," *Journal of Mathematical Psychology*, vol. 17, no. 1, pp. 21–63, 1978.

[20] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: Membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 44–54, New York: ACM Press, 2006.

[21] D. Banks and K. M. Carley, "Metric inference for social networks," *Journal of Classification*, vol. 11, no. 1, pp. 121–149, 1994.

[22] D. Banks and K. M. Carley, "Models for network evolution," *Journal of Mathematical Sociology*, vol. 21, pp. 173–196, 1996.

[23] E. Banks, E. Nabieva, R. Peterson, and M. Singh, "NetGrep: Fast network schema searches in interactomes," *Genome Biology*, vol. 9, no. 9, p. R:138, http://genomebiology.com/content/9/9/R138, 2008.

[24] A.-L. Barabási, *Linked: The New Science of Networks.* Cambridge, MA: Perseus, 2002.

[25] A.-L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, 2005.

[26] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[27] A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaboration," *Physica A*, vol. 311, no. 3–4, pp. 590–614, 2002.

[28] A.-L. Barabási and Z. Oltvai, "Network biology: Understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.

[29] F. Bassetti, M. C. Lagomarsino, and S. Mandra, "Exchangeable random networks," *Internet Mathematics*, vol. 4, no. 4, pp. 357–400, 2007.

[30] J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. A. Wallace, "Discovering hidden groups in communication networks," in *Lecture Notes in Computer Science*, vol. 3073, pp. 378–389, Springer Berlin/Heidelberg, 2004.

[31] P. S. Bearman, J. Moody, and K. Stovel, "Chains of affection: The structure of adolescent romantic and sexual networks," *American Journal of Sociology*, vol. 110, no. 1, pp. 44–91, 2004.

[32] A. Bernard, D. S. Vaughn, and A. J. Hartemink, "Reconstructing the topology of protein complexes," in *Research in Computational Molecular Biology 2007 (RECOMB07)*, (T. Speed and H. Huang, eds.), pp. 32–46, Springer Berlin/Heidelberg, 2007.

[33] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, no. 2, pp. 192–236, 1974.

[34] I. Bezáková, A. Kalai, and R. Santhanam, "Graph model selection using maximum likelihood," in *Proceedings of the 23rd International Conference on Machine Learning*, vol. 148*, ACM International Conference Proceeding Series*, pp. 105–112, New York: ACM Press, 2006.

[35] S. Bhamidi, G. Bresler, and A. Sly, "Mixing time of exponential random graphs," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 803–812, Washington, D.C.: IEEE Computer Society, 2008.

[36] I. Bhattacharya, "Collective entity resolution in relational data," PhD thesis, University of Maryland, 2006.

[37] P. J. Bickel and A. Chen, "A nonparametric view of network models and Newman–Girvan and other modularities," *Proceedings of the National Academy of Sciences*, (to appear), 2009.

[38] C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[39] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: MIT Press, 1975. Reprinted by Springer-Verlag, 2007.

[40] D. M. Blei and S. E. Fienberg, "Model-based clustering for social networks: Discussion," *Journal of the Royal Statistical Society, Series A*, vol. 170, no. 2, p. 332, 2007.

[41] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[42] J. Blitzstein and P. Diaconis, "A Sequential importance sampling algorithm for generating random graphs with prescribed degrees," Technical report, Stanford University, 2006.

[43] B. Bollobás, *Random Graphs.* New York: Cambridge University Press, 2nd ed., 2001.

[44] B. Bollobás and F. R. K. Chung, "The diameter of a cycle plus a random matching," *SIAM Journal on Discrete Mathematics*, vol. 1, no. 3, pp. 328–333, 1988.

[45] B. Bollobás, S. Janson, and O. Riordan, "The phase transition in inhomogeneous random graphs," *Random Structures & Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.

[46] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences*, vol. 99, no. Suppl. 3, pp. 7280–7287, 2002.

[47] D. Botstein, S. A. Chervitz, and J. M. Cherry, "Yeast as a model organism," *Science*, vol. 277, no. 5330, pp. 1259–1260, 1997.

[48] U. Brandes and T. Erlebach, eds., *Network Analysis: Methodological Foundations*, volume 3418 of *Lecture Notes in Computer Science*. Springer Berlin/Heidelberg, 2005.

[49] M. Braun and J. McAuliffe, "Variational inference for large-scale models of discrete choice," http://arXiv.org/abs/0712.2526, 2007.

[50] M. Buchanan, *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: W. W. Norton & Company, 2002.

[51] M.-L. G. Buot and D. S. P. Richards, "Counting and locating the solutions of polynomial systems of maximum likelihood equations, I," *Journal of Symbolic Computation*, vol. 41, no. 2, pp. 234–244, 2006.

[52] M.-L. G. Buot and D. S. P. Richards, "Counting and locating the solutions of polynomial systems of maximum likelihood equations, II: The Behrens–Fisher problem," http://arXiv.org/abs/0709.0957, 2007.

[53] R. S. Burt, "Models of network structure," *Annual Review of Sociology*, vol. 6, pp. 79–141, 1980.

[54] K. M. Carley, "Group stability: A socio-cognitive approach," in *Advances in Group Processes*, (E. Lawler, B. Markovsky, C. Ridgeway, and H. Walker, eds.), pp. 1–44, Greenwich, CT: JAI Press, 1990.

[55] K. M. Carley, "Smart agents and organizations of the future," in *The Handbook of New Media*, (L. Lievrouw and S. Livingstone, eds.), pp. 206–220, Thousand Oaks, CA: Sage, 2002.

[56] K. M. Carley, D. B. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave, "BioWar: Scalable agent-based model of bioattacks," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 36, no. 2, pp. 252–265, 2006.

[57] K. M. Carley and A. Newell, "The nature of the social agent," *Journal of Mathematical Sociology*, vol. 19, no. 4, pp. 221–262, 1994.

[58] K. M. Carley and J. Reminga, "ORA: Organizational Risk Analyzer," http://www.casos.cs.cmu.edu/projects/ora/, 2004.

[59] K. M. Carley and D. Skillicorn, "Special Issue on Analyzing Large Scale Networks: The Enron Corpus," *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pp. 179–181, Springer Netherlands, 2005.

[60] D. Chakrabarti, Y. Zhan, and C. Faloutsos, "R-MAT: A recursive model for graph mining," in *Proceedings of the 4th SIAM International Conference on Data Mining*, 2004.

[61] J. Chang and D. M. Blei, "Relational Topic Models for Document Networks," in *Proceedings of the 12th International Conference on Artifical Intelligence and Statistics (AISTATS '09)*, 2009.

[62] H. Chen, E. Reid, J. Sinai, A. Silke, and B. Ganor, eds., *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*. New York: Springer-Verlag, 2008.

[63] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in internet topologies revisited," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communication Societies*, vol. 2, pp. 608–617, 2002.

[64] J. M. Cherry, C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, R. K. Mortimer, and D. Botstein, "Genetic and physical maps of *Saccharomyces cerevisiae*," *Nature*, vol. 387, no. 6632 Suppl., pp. 67–73, 1997.

[65] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, vol. 357, no. 370–379, 2007.

[66] N. A. Christakis and J. H. Fowler, "The collective dynamics of smoking in a large social network," *New England Journal of Medicine*, vol. 358, pp. 2249–2258, 2008.

[67] N. A. Christakis and J. H. Fowler, "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study," *British Medical Journal*, vol. 337, p. a2338, 2008.

[68] N. A. Christakis and J. H. Fowler, *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. New York: Little, Brown and Co., 2009.

[69] F. Chung and L. Lu, *Complex Graphs and Networks*. Providence, RI: American Mathematical Society, 2006.

[70] F. Chung, L. Lu, and V. Vu, "The spectra of random graphs with given expected degrees," *Proceedings of the National Academy of Sciences*, vol. 100, no. 11, pp. 6313–6318, 2003.

[71] A. Clauset, "Finding local community structure in networks," *Physical Review E*, vol. 72, no. 2, p. 026132, 2005.

[72] A. Clauset and C. Moore, "How do networks become navigable?," http://arXiv.org/abs/cond-mat/0309415, 2003.

[73] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, pp. 98–101, 2008.

[74] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.

[75] R. Clegg, R. Landa, U. Harder, and M. Rio, "Evaluating and optimising models of network growth," http://arXiv.org/abs/0904.0785, 2009.

[76] P. Clifford, "Markov random fields in statistics," in *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, (G. R. Grimmett and D. J. A. Welsh, eds.), pp. 19–32, Oxford University Press, 1990.

[77] E. Cohen-Cole and J. M. Fletcher, "Detecting implausible social network effects in acne, height and headaches: Longitudinal analysis," *British Medical Journal*, vol. 337, p. a2533, 2008.

[78] E. Cohen-Cole and J. M. Fletcher, "Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic," *Journal of Health Economics*, vol. 27, pp. 1382–1387, 2008.
[79] J. Copic, M. O. Jackson, and A. Kirman, "Identifying community structures from network data via maximum likelihood methods," *The B.E. Journal of Theoretical Economics*, vol. 9, no. 1, 2009.
[80] A. Davis, B. B. Gardner, M. R. Gardner, and J. J. Wallach, *Deep South: A Social Anthropological Study of Caste and Class*. University of Chicago Press, 1941. Reprinted by University of South Carolina Press, 2009.
[81] G. B. Davis and K. M. Carley, "Clearing the FOG: Fuzzy, overlapping groups for social networks," *Social Networks*, vol. 30, no. 3, pp. 201–212, 2008.
[82] B. de Finetti, *Theory of Probability, Vol. 1–2*. New York: John Wiley & Sons, 1990. Reprint of the 1974–1975 translation.
[83] D. J. de Solla Price, "Networks of scientific Papers: The pattern of bibliographic references indicates the nature of the scientific research front," *Science*, vol. 149, no. 3683, pp. 510–515, 1965.
[84] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," Technical report, Department of Statistics, Stanford University, 2008.
[85] P. Diaconis and B. Sturmfels, "Algebraic algorithms for sampling from conditional distributions," *Annals of Statistics*, vol. 26, no. 1, pp. 363–397, 1998.
[86] P. S. Dodds, R. Muhamad, and D. J. Watts, "An experimental study of search in global social networks," *Science*, vol. 301, no. 5634, pp. 827–829, 2003.
[87] P. Domingos, "Mining social networks for viral marketing," *IEEE Intelligent Systems*, vol. 20, no. 1, pp. 80–82, 2005.
[88] P. Doreian, V. Batagelj, and A. Ferligoj, "Generalized blockmodeling of two-mode network data," *Social Networks*, vol. 26, pp. 29–53, 2004.
[89] P. Doreian, V. Batagelj, and A. Ferligoj, *Generalized Blockmodeling (Structural Analysis in the Social Sciences)*. Cambridge University Press, 2004.
[90] S. N. Dorogovtsev and J. F. F. Mendes, "Scaling behavior of developing and decaying networks," *Europhysics Letters*, vol. 52, no. 1, p. 33, 2000.
[91] R. Durrett, *Random Graph Dynamics*. Cambridge University Press, 2006.
[92] N. Eagle, A. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15274–15278, 2009.
[93] P. Erdös and A. Rényi, "On random graphs, I," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.
[94] P. Erdös and A. Rényi, "The Evolution of Random Graphs," *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 5, pp. 17–61, 1960.
[95] E. A. Erosheva, S. E. Fienberg, and J. Lafferty, "Mixed-membership Models of Scientific Publications," *Proceedings of the National Academy of Sciences*, vol. 101, no. Suppl. 1, pp. 5220–5227, 2004.
[96] E. Even-Dar and M. Kearns, "A small world threshold for economic network formation," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 19, pp. 385–392, Cambridge, MA: MIT Press, 2007.
[97] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proceedings of the Conference on Applications,*

Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '99), pp. 251–261, New York: ACM Press, 1999.

[98] S. Fields and O. Song, "A novel genetic system to detect protein-protein interactions," *Nature*, vol. 340, no. 6230, pp. 245–246, 1989.

[99] S. E. Fienberg, *Analysis of Cross-Classified Categorical Data*. Cambridge, MA: MIT Press, 2nd ed., 1980. Reprinted by Springer-Verlag, 2007.

[100] S. E. Fienberg and S. K. Lee, "On small world statistics," *Psychometrika*, vol. 40, no. 2, pp. 219–228, 1975.

[101] S. E. Fienberg, M. M. Meyer, and S. S. Wasserman, "Statistical analysis of multiple sociometric relations," *Journal of the American Statistical Association*, vol. 80, pp. 51–67, 1985.

[102] S. E. Fienberg, S. Petrović, and A. Rinaldo, "Algebraic statistics for p_1 random graph models: Markov bases and their uses," in *Papers in Honor of Paul W. Holland*, (S. Sinharay and N. J. Dorans, eds.), Educational Testing Service, 2009.

[103] S. E. Fienberg and S. S. Wasserman, "Categorical data analysis of single sociometric relations," *Sociological Methodology*, pp. 156–192, 1981.

[104] S. E. Fienberg and S. S. Wasserman, "An exponential family of probability distributions for directed graphs: Comment," *Journal of the American Statistical Association*, vol. 76, no. 373, pp. 54–57, 1981.

[105] J. Flannick, A. Novak, B. S. Srinivasan, H. H. McAdams, and S. Batzoglou, "Græmlin: General and robust alignment of multiple large interaction networks," *Genome Research*, vol. 16, no. 9, pp. 1169–1181, 2006.

[106] A. D. Flaxman, A. M. Frieze, and J. Vera, "A geometric preferential attachment model of networks," *Internet Mathematics*, vol. 3, no. 2, pp. 187–206, 2006.

[107] A. D. Flaxman, A. M. Frieze, and J. Vera, "A geometric preferential attachment model of networks II," *Internet Mathematics*, vol. 4, no. 1, pp. 87–112, 2007.

[108] J. Fowler and N. Christakis, "Estimating peer effects on health in social networks," *Journal of Health Economics*, vol. 27, no. 5, pp. 1400–1405, 2008.

[109] O. Frank, "Network sampling and model fitting," in *Models and Methods in Social Network Analysis*, (P. J. Carrington, J. Scott, and S. S. Wasserman, eds.), pp. 31–56, Cambridge University Press, 2005.

[110] O. Frank and D. Strauss, "Markov graphs," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 832–842, 1986.

[111] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol. 303, no. 5659, pp. 799–805, 2004.

[112] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning probabilistic relational models," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, pp. 1300–1309, 1999.

[113] M. T. Gastner and M. E. J. Newman, "Shape and efficiency in spatial distribution networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 1, p. P01015, 2006.

[114] A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier,

V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga, "Proteome survey reveals modularity of the yeast cell machinery," *Nature*, vol. 440, no. 7084, pp. 631–636, 2006.

[115] A.-C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.-A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga, "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, pp. 141–147, 2002.

[116] L. Getoor, N. Friedman, D. Koller, and B. Taskar, "Learning probabilistic models of link structure," *Journal of Machine Learning Research*, vol. 3, pp. 679–707, 2003.

[117] L. Getoor and B. Taskar, eds., *Introduction to Statistical Relational Learning*. Cambridge, MA: MIT Press, 2007.

[118] C. J. Geyer and E. A. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 54, pp. 657–699, 1992.

[119] E. N. Gilbert, "Random graphs," *Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

[120] K. J. Gile and M. S. Handcock, "Model-based assessment of the impact of missing data on inference for networks," CSSS Working paper No. 66, 2006.

[121] P. S. Gill and T. B. Swartz, "Bayesian analysis of directed graphs data with application to social networks," *Applied Statistics*, vol. 53, no. 2, pp. 249–260, 2004.

[122] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[123] K. S. Gleditsch, "Expanded trade and GDP data," *Journal of Conflict Resolution*, vol. 46, no. 5, pp. 712–724, 2002.

[124] A. Globerson, G. Chechik, F. Pereira, and N. Tishby, "Euclidean embedding of co-occurrence data," *Journal of Machine Learning Research*, vol. 8, pp. 2265–2295, 2007.

[125] S. Goel and M. J. Salganik, "Respondent-driven sampling as Markov chain Monte Carlo," *Statistics in Medicine*, vol. 28, no. 17, pp. 2202–2229, 2009.

[126] A. Goldenberg and A. Moore, "Tractable learning of large Bayes net structures from sparse data," in *Proceedings of the 21st International Conference on Machine Learning*, p. 44, New York: ACM Press, 2004.

[127] A. Goldenberg and A. Moore, "Bayes net graphs to understand coauthorship networks," in *KDD Workshop on Link Discovery: Issues, Approaches and Applications*, 2005.

[128] A. Goldenberg and A. Zheng, "Exploratory study of a new model for evolving networks," in *Statistical Network Analysis: Models, Issues and New Directions,* volume 4503 in *Lecture Notes in Computer Science*, (E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. X. Zheng, eds.), Springer Berlin/Heidelberg, 2007.

[129] S. M. Goodreau, M. S. Handcock, D. R. Hunter, C. T. Butts, and M. Morris, "A *statnet* tutorial," *Journal of Statistical Software*, vol. 24, no. 9, pp. 1–26, 2008.

[130] S. M. Goodreau, J. A. Kitts, and M. Morris, "Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks," *Demography*, vol. 46, no. 1, pp. 103–125, 2009.

[131] S. Goyal, *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2007.

[132] B. Grún and F. Leisch, "Dealing with label switching in mixture models under genuine multimodality," *Journal of Multivariate Analysis*, vol. 100, no. 5, pp. 851–861, 2008.

[133] A. Guetz and P. Constantine, "Lecture notes for course on Information Networks," http://www.stanford.edu/class/msande337/notes/Lec1.pdf, 2007.

[134] S. J. Haberman, "An exponential family of probability distributions for directed graphs: Comment," *Journal of the American Statistical Association*, vol. 76, no. 373, pp. 60–61, 1981.

[135] M. S. Handcock and K. J. Gile, "Modeling networks from sampled data," *Annals of Applied Statistics*, vol. 4, no. 1, 2010.

[136] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris, "statnet: Software tools for the representation, visualization, analysis and simulation of network data," *Journal of Statistical Software*, vol. 24, no. 1, pp. 12–25, 2008.

[137] M. S. Handcock, A. E. Raftery, and J. Tantrum, "Model-based clustering for social networks (with discussion)," *Journal of the Royal Statistical Society, Series A*, vol. 170, pp. 301–354, 2007.

[138] M. S. Handcock, G. L. Robins, T. A. B. Snijders, J. Moody, and J. Besag, "Assessing degeneracy in statistical models of social networks," *Journal of the American Statistical Association*, vol. 76, pp. 33–50, 2003.

[139] S. Hanneke and E. P. Xing, "Discrete temporal models of social networks," in *Statistical Network Analysis: Models, Issues and New Directions,* volume 4503 of *Lecture Notes in Computer Science*, (E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. X. Zheng, eds.), Springer Berlin/Heidelberg, 2007.

[140] L. P. Hansen and J. A. Scheinkman, "Back to the future: Generating moment implications for continuous-time Markov processes," *Econometrica*, vol. 63, no. 4, pp. 767–804, 1995.

[141] K. M. Harris, F. Florey, J. Tabor, P. S. Bearman, J. Jones, and R. J. Udry, "The national longitudinal study of adolescent health: Research Design," Technical report, Carolina Population Center, University of North Carolina, Chapel Hill, 2003.

[142] S. Hill, F. Provost, and C. Volinsky, "Network-based marketing: Identifying likely adopters via consumer networks," *Statistical Science*, vol. 21, no. 2, pp. 256–276, 2006.
[143] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sørensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. V. Hogue, D. Figeys, and M. Tyers, "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry," *Nature*, vol. 415, pp. 180–183, 2002.
[144] P. D. Hoff, "Random effects models for network data," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, (R. Breiger, K. M. Carley, and P. E. Pattison, eds.), pp. 303–312, Washington, D.C.: The National Academies Press, 2003.
[145] P. D. Hoff, "Modeling homophily and stochastic equivalence in symmetric relational data," in *Advances in Neural Information Processing Systems*, vol. 20, (J. C. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), pp. 657–664, MIT Press, 2008.
[146] P. D. Hoff, A. E. Raftery, and M. S. Handcock, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1090–1098, 2002.
[147] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
[148] P. W. Holland and S. Leinhardt, "Local structure in social networks," *Sociological Methodology*, vol. 7, pp. 1–45, 1976.
[149] P. W. Holland and S. Leinhardt, "A dynamic model for social networks," *Journal of Mathematical Sociology*, vol. 5, no. 1, pp. 5–20, 1977.
[150] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs (with discussion)," *Journal of the American Statistical Association*, vol. 76, no. 373, pp. 33–65, 1981.
[151] P. Holme, J. Karlin, and S. Forrest, "An integrated model of traffic, geography and economy in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 3, pp. 7–15, 2008.
[152] B. A. Huberman and L. A. Adamic, "Growth dynamics of the world-wide web," *Nature*, vol. 401, p. 131, 1999.
[153] S. Huh and S. E. Fienberg, "Temporally-evolving mixed membership stochastic blockmodels: Exploring the Enron e-mail database," in *Proceedings of the NIPS Workshop on Analyzing Graphs: Theory & Applications*, Whistler, British Columbia, 2008.
[154] M. Huisman and C. Steglich, "Treatment of non-response in longitudinal network studies," *Social Networks*, vol. 30, no. 4, pp. 297–308, 2008.
[155] D. R. Hunter, S. M. Goodreau, and M. S. Handcock, "Goodness of fit of social network models," *Journal of the American Statistical Association*, vol. 103, no. 481, pp. 248–258, 2008.

[156] D. R. Hunter and M. S. Handcock, "Inference in curved exponential family models for networks," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 565–583, 2006.

[157] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, "ergm: A package to fit, simulate and diagnose exponential-family models for networks," *Journal of Statistical Software*, vol. 24, no. 3, http://www.jstatsoft.org/v24/i03/paper, 2008.

[158] M. Huss and P. Holme, "Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks," *IET Systems Biology*, vol. 1, pp. 280–285, 2007.

[159] T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki, "Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins," *Proceedings of the National Academy of Sciences*, vol. 97, no. 3, pp. 1143–1147, 2000.

[160] M. O. Jackson, *Social and Economic Networks*. Princeton University Press, 2008.

[161] S. Janson, T. Luczak, and A. Ruciński, *Random Graphs*. New York: John Wiley & Sons, 2000.

[162] L. J. Jensen and P. Bork, "Biochemistry: Not comparable, but complementary," *Science*, vol. 322, no. 5898, pp. 56–57, 2008.

[163] J. H. Jones and M. S. Handcock, "An assessment of preferential attachment as a mechanism for human sexual network formation," in *Proceedings of the Royal Society, Series B*, vol. 270, no. 1520, pp. 1123–1128, 2003.

[164] J. H. Jones and M. S. Handcock, "Social networks (communication arising): Sexual contacts and epidemic thresholds," *Nature*, vol. 423, pp. 605–606, 2003.

[165] O. Kallenberg, "Probabilistic symmetries and invariance principles," in *Probability and its Applications*, New York: Springer, 2005.

[166] L. Katz, "The distribution of the number of isolates in a social group," *Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 271–276, 1952.

[167] L. Katz and J. H. Powell, "Probability distributions of random variables associated with a structure of the sample space of sociometric investigations," *Annals of Mathematical Statistics*, vol. 28, no. 2, pp. 442–448, 1957.

[168] L. Katz and T. R. Wilson, "The variance of the number of mutual choices in sociometry," *Psychometrika*, vol. 21, no. 3, pp. 299–304, 1956.

[169] M. Kearns, S. Suri, and N. Montfort, "An experimental study of the coloring problem on human subject networks," *Science*, vol. 313, no. 5788, pp. 824–827, 2006.

[170] D. Kempe, J. Kleinberg, and E. Tardos, "Influential nodes in a diffusion model for social networks," in *Automata, Languages and Programming*, vol. 3580 of *Lecture Notes in Computer Science*, pp. 1127–1138, Springer Berlin/Heidelberg, 2005.

[171] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[172] J. M. Kleinberg, "Navigation in a small world — It is easier to find short chains between points in some networks than others," *Nature*, vol. 406, p. 845, 2000.

[173] J. M. Kleinberg, "The small-world phenomenon: An algorithmic perspective," in *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pp. 163–170, New York: ACM Press, 2000.

[174] J. M. Kleinberg, "Small-world phenomena and the dynamics of information," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 14, Cambridge, MA: MIT Press, 2001.

[175] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, "The web as a graph: Measurements, models and methods," in *Computing and Combinatorics*, vol. 1627 of *Lecture Notes in Computer Science*, pp. 1–17, Springer Berlin/Heidelberg, 1999.

[176] A. S. Klovdahl, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth, and W. W. Darrow, "Social networks and infectious disease: The Colorado Springs study," *Social Science & Medicine*, vol. 38, no. 1, pp. 79–88, 1994.

[177] E. D. Kolacyzk, *Statistical Analysis of Network Models.* New York: Springer, 2009.

[178] J. Koskinen, G. L. Robins, and P. E. Pattison, "Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation," Technical report, Department of Psychology, School of Behavioural Science, University of Melbourne, Austrailia, 2008.

[179] J. H. Koskinen and T. A. B. Snijders, "Bayesian inference for dynamic social network data," *Journal of Statistical Planning and Inference*, vol. 137, no. 12, pp. 3930–3938, 2007.

[180] G. Kossinets, "Effects of missing data in social networks," *Social Networks*, vol. 28, no. 3, pp. 247–268, 2006.

[181] D. Krackhardt, "The ties that torture: Simmelian tie analysis in organizations," *Research in the Sociology of Organizations*, vol. 16, pp. 183–210, 1999.

[182] V. E. Krebs, "Mapping networks of terrorist cells," *Connections*, vol. 24, no. 3, pp. 43–52, 2002.

[183] P. N. Krivitsky, M. S. Handcock, A. E. Raftery, and P. D. Hoff, "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models," *Social Networks*, vol. 31, no. 3, pp. 204–213, 2009.

[184] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, and J. F. Greenblatt, "Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 440, no. 7084, pp. 637–643, 2006.

[185] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph," in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pp. 57–65, 2000.

[186] S. L. Lauritzen, "Rasch models with exchangeable rows and columns," in *Bayesian Statistics 7*, (J. M. Bernardo et al., ed.), pp. 215–232, Oxford University Press, 2003.

[187] S. L. Lauritzen, "Exchangeable Rasch matrices," *Rendiconti di Matematica, Serie VII*, vol. 28, no. 1, pp. 83–95, 2008.

[188] S. Lee and C. F. Stevens, "General design principle for scalable neural circuits in a vertebrate retina," *Proceedings of the National Academy of Sciences*, vol. 104, no. 31, pp. 12931–12935, 2007.

[189] R. T. A. J. Leenders, "Models for network dynamics: A Markovian framework," *Journal of Mathematical Sociology*, vol. 20, pp. 1–21, 1995.

[190] E. A. Leicht, G. Clarkson, K. Shedden, and M. Newman, "Large-scale structure of time evolving citation networks," *European Physics Journal B*, vol. 59, no. 1, pp. 75–83, 2007.

[191] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos, "Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication," in *Knowledge Discovery in Databases: PKDD 2005*, vol. 3721 of *Lecture Notes in Computer Science*, pp. 133–145, Springer Berlin/Heidelberg, 2005.

[192] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," http://arXiv.org/abs/0812.4905v2, 2009.

[193] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 631–636, New York: ACM Press, 2006.

[194] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 177–187, New York: ACM Press, 2005.

[195] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, p. 2, 2007.

[196] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: Definition, proper ties and implications," *Internet Mathematics*, vol. 2, no. 4, pp. 431–523, 2005.

[197] W. Li and A. McCallum, "Pachinko allocation: DAG-structured mixture models of topic correlations," in *Proceedings of the 23rd International Conference on Machine Learning*, vol. 148 of *ACM International Conference Processing Series*, pp. 577–584, New York: ACM Press, 2006.

[198] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM '03)*, pp. 556–559, ACM Press: New York, 2003.

[199] F. Lorrain and H. C. White, "Structural equivalence of individuals in social networks," *Journal of Mathematical Sociology*, vol. 1, pp. 49–80, 1971.

[200] R. D. Luce, "Connectivity and generalized cliques in sociometric group structure," *Psychometrika*, vol. 15, no. 2, pp. 169–190, 1950.

[201] R. D. Luce, “Networks satisfying minimality conditions,” *American Journal of Mathematics*, vol. 75, no. 4, pp. 825–838, 1953.

[202] R. D. Luce, J. Macy, Jr., and R. Tagiuri, “A statistical model for relational analysis,” *Psychometrika*, vol. 20, no. 4, pp. 319–327, 1955.

[203] R. D. Luce and A. D. Perry, “A method of matrix analysis of group structure,” *Psychometrika*, vol. 14, no. 2, pp. 95–116, 1949.

[204] A.-L. B. M. Newman and D. J. Watts, eds., *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[205] G. S. Mann, D. Mimno, and A. McCallum, “Bibliometric impact measures leveraging topic analysis,” in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digial Libraries*, pp. 65–74, New York: ACM Press, 2006.

[206] J.-M. Marin, K. Mengersen, and C. P. Robert, “Bayesian modelling and inference on mixtures of distributions,” in *Handbook of Statistics 25*, (D. Dey and C. R. Rao, eds.), pp. 15840–15845, Elsevier Sciences, 2005.

[207] T. F. Mayer, “Parties and networks: Stochastic models for relationship networks,” *Journal of Mathematical Sociology*, vol. 10, pp. 51–103, 1984.

[208] A. McCallum, A. Corradda-Emmanuel, and X. Wang, “Topic and role discovery in social networks,” in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 786–791, 2005.

[209] A. McCallum, X. Wang, and N. Mohanty, “Joint group and topic discovery from relations and text,” in *Statistical Network Analysis: Models, Issues and New Directions*, vol. 4503 of *Lecture Notes in Computer Science*, (E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. Zheng, eds.), pp. 28–44, Springer Berlin/Heidelberg, 2007.

[210] K. McComb, C. Moss, S. M. Durant, L. Baker, and S. Sayialel, “Matriarchs as repositories of social knowledge in African elephants,” *Science*, vol. 292, no. 5516, pp. 491–494, 2001.

[211] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. Chapman & Hall/CRC, 2nd ed., 1989.

[212] J. W. McDonald, P. W. F. Smith, and J. J. Forster, “Markov chain Monte Carlo exact inference for social networks,” *Social Networks*, vol. 29, no. 1, pp. 127–136, 2007.

[213] M. McGlohon, L. Akoglu, and C. Faloutsos, “Weighted graphs and disconnected components: Patterns and a generator,” in *Proceedings of the 14th international conference on knowledge discovery and data mining*, pp. 524–532, New York: ACM Press, 2008.

[214] M. M. Meyer, “Transforming contingency tables,” *Annals of Statistics*, vol. 10, no. 4, pp. 1172–1181, 1982.

[215] M. Middendorf, E. Ziv, and C. H. Wiggins, “Inferring network mechanisms: The Drosophila melanogaster protein interaction network,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 9, pp. 3192–3197, 2005.

[216] S. Milgram, “The small world problem,” *Psychology Today*, vol. 1, no. 1, pp. 60–67, 1967.

[217] D. Mimno and A. McCallum, “Mining a digital library for influential authors,” in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 105–106, New York: ACM Press, 2007.

[218] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan, "Finding strongly-knit clusters in social networks," *Internet Mathematics*, vol. 5, no. 1-2, pp. 155–174, 2008.

[219] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[220] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures and Algorithms*, vol. 6, no. 2–3, pp. 161–180, 1995.

[221] M. Molloy and B. Reed, "The size of the largest component of a random graph on a fixed degree sequence," *Combinatorics, Probability and Computing*, vol. 7, pp. 295–306, 1998.

[222] J. Moreno, *Who Shall Survive?* Washington, D.C.: Nervous and Mental Disease Publishing Company, 1934.

[223] M. Morris, M. S. Handcock, and D. R. Hunter, "Specification of exponential-family random graph models: terms and computational aspects," *Journal of Statistical Software*, vol. 24, no. 4, http://www.jstatsoft.org/v24/i04, 2008.

[224] M. Morris, M. S. Handcock, W. C. Miller, C. A. Ford, J. L. Schmitz, M. M. Hobbs, M. S. Cohen, K. M. Harris, and J. R. Udry, "Prevalence of HIV infection among young adults in the United States: Results from the Add Health Study," *American Journal of Public Health*, vol. 96, no. 6, pp. 1091–1097, 2006.

[225] M. Morris and M. Kretzschmar, "Concurrent partnerships and transmission dynamics in networks," *Social Networks*, vol. 17, no. 3–4, pp. 299–318, 1995.

[226] Q. Morris, B. Frey, and C. Paige, "Denoising and untangling graphs using degree priors," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 16, Cambridge, MA: MIT Press, 2003.

[227] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, and Q. Morris, "GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function," *Genome Biology*, vol. 9, no. Suppl. 1, p. S4, 2008.

[228] E. Nabieva, K. Jim, A. Agarwal, B. Chazelle, and M. Singh, "Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps," *Bioinformatics*, vol. 21, no. Suppl. 1, pp. i302–i310, 2005.

[229] R. M. Neal, *Bayesian Learning for Neural Networks*, vol. 118 of Lecture Notes in Statistics. New York: Springer-Verlag, 1996.

[230] J. Neville and D. Jensen, "Collective classification with relational dependency networks," in *Proceedings of the 2nd Multi-Relational Data Mining Workshop, 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[231] J. Neville, D. Jensen, L. Friedland, and M. Hay, "Learning relational probability trees," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 625–630, New York: ACM Press, 2003.

[232] M. E. J. Newman, "Detecting community structure in networks," *European Physics Journal B*, vol. 38, no. 2, pp. 321–330, 2004.

[233] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, no. 3, p. 036104, 2006.

[234] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[235] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proceedings of the National Academy of Science*, vol. 99, no. Suppl. 1, pp. 2566–2572, 2002.

[236] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 1077–1087, 2001.

[237] M. Nunkesser and D. Sawitzki, "Blockmodels," in *Network Analysis*, vol. 3418 of *Lecture Notes in Computer Science*, (U. Brandes and T. Erlebach, eds.), pp. 253–292, Springer Berlin/Heidelberg, 2005.

[238] J. O'Madadhain, P. Smyth, and L. Adamic, "Learning predictive models for link formation," in *Proceedings of the International Sunbelt Social Network Conference*, 2005.

[239] J. Park and M. E. J. Newman, "Solution of the two-star model of a network," *Physics Reviews E*, vol. 70, p. 066146, 2004.

[240] J. Park and M. E. J. Newman, "Statistical Mechanics of Networks," *Physical Review E*, vol. 70, p. 066117, 2004.

[241] P. E. Pattison and S. S. Wasserman, "Logit models and logistic regressions for social networks: II. Multivariate relations," *British Journal of Mathematical and Statistical Psychology*, vol. 52, no. 2, pp. 169–193, 1999.

[242] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles, "Winners don't take all: Characterizing the competition for links on the web," *Proceedings of the National Academy of Sciences*, vol. 99, no. 8, pp. 5207–5211, 2002.

[243] B. Pittel, "On tree census and the giant component in sparse random graphs," *Random Structures Algorithms*, vol. 1, no. 3, pp. 311–342, 1990.

[244] R. Radner and A. Tritter, "Communication in Networks," Technical report Ec2098, Cowles Commission, University of Chicago, 1954.

[245] O. Ratmann, O. Jørgensen, T. Hinkley, M. P. H. Stumpf, S. Richardson, and C. Wiuf, "Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*," *PLoS Computational Biology*, vol. 3, no. 11, pp. 2266–2278, 2007.

[246] O. Ratmann, C. Wiuf, and J. W. Pinney, "From evidence to inference: Probing the evolution of protein interaction networks," *HFSP Journal*, vol. 3, no. 5, pp. 290–306, 2009.

[247] P. Ravikumar, "Approximate inference, structure learning and feature estimation in Markov random fields," PhD thesis, Machine Learning Department, School of Computer Science, Carnegie Mellon University, 2007.

[248] T. Reguly, A. Breitkreutz, L. Boucher, B.-J. Breitkreutz, G. C. Hon, C. L. Myers, A. Parsons, H. Friesen, R. Oughtred, A. Tong, C. Stark, Y. Ho, D. Botstein, B. Andrews, C. Boone, O. G. Troyanskya, T. Ideker, K. Dolinski, N. N. Batada, and M. Tyers, "Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*," *Journal of Biology*, vol. 5, no. 4, p. 11, 2006.

[249] E. Reid and H. Chen, "Mapping the contemporary terrorism research domain: Researchers, publications, and institutions analysis," in *Intelligence and Security Informatics*, vol. 3495 of *Lecture Notes in Computer Science*, pp. 322–339, Springer Berlin/Heidelberg, 2005.

[250] E. Reid, J. Qin, W. Chung, J. Xu, Y. Zhou, R. Schumaker, M. Sageman, and H. Chen, "Terrorism knowledge discovery project: A knowledge discovery approach to addressing the threats of terrorism," in *Intelligence and Security Informatics*, vol. 3073 of *Lecture Notes in Computer Science*, pp. 125–145, Springer Berlin/Heidelberg, 2004.

[251] A. Rinaldo, S. E. Fienberg, and Y. Zhou, "On the geometry of discrete exponential families with application to exponential random graph models," *Electronic Journal of Statististics*, vol. 3, pp. 446–484, 2009.

[252] J. M. Roberts, Jr., "Simple methods for simulating sociomatrices with given marginal totals," *Social Networks*, vol. 22, no. 3, pp. 273–283, 2000.

[253] G. L. Robins and P. E. Pattison, "Random graph models for temporal processes in social networks," *Journal of Mathematical Sociology*, vol. 25, pp. 5–41, 2001.

[254] G. L. Robins, P. E. Pattison, and S. S. Wasserman, "Logit models and logistic regressions for social networks: III. Valued relations," *Psychometrika*, vol. 64, no. 3, pp. 371–394, 1999.

[255] G. L. Robins, P. E. Pattison, and J. Woolcock, "Missing data in networks: Exponential random graph (p*) models for networks with non-respondents," *Social Networks*, vol. 26, no. 3, pp. 257–283, 2004.

[256] G. L. Robins, T. A. B. Snijders, P. Wang, M. S. Handcock, and P. E. Pattison, "Recent developments in exponential random graph (p^*) models for social networks," *Social Networks*, vol. 29, no. 2, pp. 192–215, 2007.

[257] T. T. Rogers and J. L. McClelland, *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press, 2004.

[258] M. J. Salganik and D. D. Heckathorn, "Sampling and estimation in hidden populations using respondent-driven sampling," *Sociological Methodology*, vol. 34, pp. 193–239, 2004.

[259] F. S. Sampson, "A novitiate in a period of change: An experimental and case study of social relationships," PhD thesis, Cornell University, 1968.

[260] O. Sandberg, "Searching in a small world," PhD thesis, Division of Mathematical Statistics, Department of Mathematical Sciences, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2005.

[261] O. Sandberg, "Neighbor selection and hitting probability in small-world graphs," *Annals of Applied Probability*, vol. 18, no. 5, pp. 1771–1793, 2008.

[262] O. Sandberg and I. Clarke, "The evolution of navigable small-world networks," http://arXiv.org/abs/cs/0607025, 2006.

[263] P. Sarkar and A. W. Moore, "Dynamic social network analysis using latent space models," *SIGKDD Explorations: Special Edition on Link Mining*, vol. 7, no. 2, pp. 31–40, 2005.

[264] P. Sarkar and A. W. Moore, "Dynamic social network analysis using latent space models," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 18, pp. 1145–1152, Cambridge, MA: MIT Press, 2005.

[265] P. Sarkar, S. M. Siddiqi, and G. J. Gordon, "A latent space approach to dynamic embedding of co-occurrence data," in *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AI-STATS '07)*, 2007.

[266] C. R. Shalizi, M. F. Camperi, and K. L. Klinkner, "Discovering functional communities in dynamical networks," in *Statistical Network Analysis: Models, Issues and New Directions*, vol. 4503 of *Lecture Notes in Computer Science*, (E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. Zheng, eds.), pp. 140–157, Springer Berlin/Heidelberg, 2007.

[267] B. Shneiderman and A. Aris, "Network visualization by semantic substrates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 733–740, 2006.

[268] G. Simmel and K. H. Wolff, *The Sociology of Georg Simmel*. New York: The Free Press, 1950.

[269] H. A. Simon, "On a class of skew distribution functions," *Biometrika*, vol. 42, no. 3–4, pp. 425–440, 1955.

[270] B. Singer and S. Spilerman, "Social mobility models for heterogenous populations," *Sociological Methodology*, vol. 5, pp. 356–401, 1973–1974.

[271] B. Singer and S. Spilerman, "The representation of social processes by Markov models," *The American Journal of Sociology*, vol. 82, no. 1, pp. 1–54, 1976.

[272] T. A. B. Snijders, "The transition probabilities of the reciprocity model," *Journal of Mathematical Sociology*, vol. 23, no. 4, pp. 241–253, 1999.

[273] T. A. B. Snijders, "The statistical evaluation of social network dynamics," *Sociological Methodology*, vol. 31, pp. 361–395, 2001.

[274] T. A. B. Snijders, "Accounting for degree distributions in empirical analysis of network dynamics," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, (R. L. Breiger, K. M. Carley, and P. E. Pattison, eds.), pp. 146–161, Washington, D.C.: The National Academies Press, 2003.

[275] T. A. B. Snijders, "Models for longitudinal network data," in *Models and Methods in Social Network Analysis*, (P. J. Carrington, J. Scott, and S. S. Wasserman, eds.), ch. 11, New York: Cambridge University Press, 2005.

[276] T. A. B. Snijders, "Statistical methods for network dynamics," in *Proceedings of the XLIII Scientific Meeting, Italian Statistical Society*, (S. R. Luchini et al., ed.), pp. 281–296, Padova: CLEUP, 2006.

[277] T. A. B. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, vol. 14, no. 1, pp. 75–100, 1997.

[278] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," *Sociological Methodology*, vol. 36, pp. 99–153, 2006.

[279] T. A. B. Snijders and M. A. J. van Duijin, "Simulation for statistical inference in dynamic network models," in *Simulating Social Phenomena*, (R. Conte, R. Hegselmann, and P. Terna, eds.), pp. 493–512, Berlin: Springer, 1997.

[280] T. A. B. Snijders and M. A. J. van Duijn, "Conditional maximum likelihood estimation under various specifications of exponential random graph

models," in *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank*, (J. Hagberg, ed.), pp. 117–134, Stockholm, Sweden: Department of Statistics, University of Stockholm, 2002.

[281] R. Solomonoff and A. Rapoport, "Connectivity of random nets," *Bulletin of Mathematical Biology*, vol. 13, no. 2, pp. 107–117, 1951.

[282] S. Spilerman, "Structural analysis and the generation of sociograms," *Behavioral Science*, vol. 11, pp. 312–318, 1966.

[283] M. Stephens, "Bayesian analysis of mixtures with an unknown number of components — an alternative to reversible jump methods," *Annals of Statistics*, vol. 28, no. 1, pp. 40–74, 2000.

[284] D. Stork and W. Richards, "Nonrespondents in communication network studies," *Group & Organization Management*, vol. 17, no. 2, pp. 193–209, 1992.

[285] D. B. Stouffer, R. D. Malmgren, and L. A. N. Amaral, "Comment on Barabási," *Nature*, vol. 435, p. 207, http://arXiv.org/abs/physics/ 0510216, 2005, 2005.

[286] D. B. Stouffer, R. D. Malmgren, and L. A. N. Amaral, "Log-normal statistics in e-mail communication patterns," http://arXiv.org/abs/physics/0605027, 2008.

[287] D. Strauss and M. Ikeda, "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, vol. 85, no. 409, pp. 204–212, 1990.

[288] M. P. H. Stumpf and T. Thorne, "Multi-model Inference of Network Properties from Incomplete Data," *Journal of Integrative Bioinformatics*, vol. 3, no. 2, p. 32, http://journal.imbio.de/index.php?paper_id=32, 2006.

[289] M. P. H. Stumpf, C. Wiuf, and R. M. May, "Subnets of scale-free networks are not scale-free: Sampling properties of networks," *Proceedings of the National Academy of Sciences*, vol. 102, no. 12, pp. 4221–4224, 2005.

[290] S. Swasey, "Netflix awards $1 Million Netflix prize and announces second $1 million challenge," *Wall Street Journal*, September 21 2009.

[291] K. Tarassov, V. Messier, C. R. Landry, S. Radinovic, M. M. Serna Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey, and S. W. Michnick, "An in vivo map of the yeast protein interactome," *Science*, vol. 320, no. 5882, pp. 1465–1470, 2008.

[292] H. M. Taylor and S. Carlin, *An introduction to stochastic modeling*. New York: Academic Press, 3rd ed., 1998.

[293] S. K. Thompson, "Adaptive web sampling," *Biometrics*, vol. 62, no. 4, pp. 1224–1234, 2006.

[294] S. K. Thompson, "Targeted random walk designs," *Survey Methodology*, vol. 32, no. 1, pp. 11–24, 2006.

[295] S. K. Thompson and O. Frank, "Model-based estimation with link-tracing sampling designs," *Survey Methodolology*, vol. 26, no. 1, pp. 87–98, 2000.

[296] S. K. Thompson and G. A. F. Seber, *Adaptive Sampling*. New York: Wiley, 1996.

[297] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley & Sons, 1986.

[298] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.

[299] R. J. Udry, "The national longitudinal study of adolescent health: (Add health) Waves I and II, 1994–1996; Wave III, 2001–2002," Technical report, Carolina Population Center, University of North Carolina, Chapel Hill, 2003.

[300] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg, "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.

[301] M. A. J. van Duijn, K. J. Gile, and M. S. Handcock, "A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models," *Social Networks*, vol. 31, no. 1, pp. 52–62, 2009.

[302] M. A. J. van Duijn, T. A. B. Snijders, and B. J. H. Zijlstra, "p_2: A Random Effects Model with Covariates for Directed Graphs," *Statistica Neerlandica*, vol. 58, no. 2, pp. 234–254, 2004.

[303] E. A. Vance, E. A. Archie, and C. J. Moss, "Social networks in African elephants," *Computational & Mathematical Organization Theory*, http://www.springerlink.com/content/enpk5g428272927m, 2008. To appear in print, 2009.

[304] A. Vázquez, J. G. Oliveira, Z. Dezsö, K. Goh, I. Kondor, and A.-L. Barabási, "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, p. 036127, 2006.

[305] E. Volz and D. D. Heckathorn, "Probability based estimation theory for respondent driven sampling," *Journal of Official Statistics*, vol. 24, no. 1, pp. 79–97, 2008.

[306] E. Volz and L. A. Meyers, "Epidemic thresholds in dynamic contact networks," *Journal of the Royal Society Interface*, vol. 6, no. 32, pp. 233–241, 2009.

[307] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork, "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature*, vol. 417, no. 6887, pp. 399–403, 2002.

[308] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[309] A. M. Walczak, A. Mugler, and C. H. Wiggins, "A stochastic spectral analysis of transcriptional regulatory cascades," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6529–6534, 2009.

[310] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS '03)*, pp. 25–34, 2003.

[311] Y. Y. Wang and G. Y. Wong, "Stochastic Blockmodels for Directed Graphs," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 8–19, 1987.

[312] S. S. Wasserman, "Stochastic models for directed graphs," PhD thesis, Department of Statistics, Harvard University, 1977.

[313] S. S. Wasserman, "Analyzing social networks as stochastic processes," *Journal of the American Statistical Association*, vol. 75, no. 370, pp. 280–294, 1980.

[314] S. S. Wasserman and C. Anderson, "Stochastic a posteriori blockmodels: Construction and assessment," *Social Networks*, vol. 9, no. 1, pp. 1–36, 1987.

[315] S. S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge University Press, 1994.

[316] S. S. Wasserman and P. E. Pattison, "Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*.," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[317] S. S. Wasserman, G. L. Robins, and D. Steinley, "Statistical models for networks: A brief review of some recent research," in *Statistical Network Analysis: Models, Issues and New Directions*, vol. 4503 of *Lecture Notes in Computer Science*, (E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. X. Zheng, eds.), Springer Berlin/Heidelberg, 2007.

[318] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.

[319] D. J. Watts, *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton & Company, 2003.

[320] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[321] H. C. White, "Search parameters for the small world problem," *Social Forces*, vol. 49, no. 2, pp. 259–264, 1970.

[322] H. C. White, S. A. Boorman, and R. L. Breiger, "Social structure from multiple networks. I. Blockmodels of roles and positions," *The American Journal of Sociology*, vol. 81, no. 4, pp. 730–780, 1976.

[323] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180–183, 2000.

[324] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the internet: A source of enormouse confusion and great potential," *Notices of the American Mathematical Society*, vol. 56, no. 5, pp. 586–599, 2009.

[325] C. Wiuf, M. Brameier, O. Hagberg, and M. P. H. Stumpf, "A likelihood approach to analysis of network data," *Proceedings of the National Academy of Sciences*, vol. 103, no. 20, pp. 7566–7570, 2006.

[326] C. Wiuf and M. P. H. Stumpf, "Binomial subsampling," *Journal of the Royal Society, Series A*, vol. 462, no. 2068, pp. 1181–1195, 2006.

[327] S. L. Wong, L. V. Zhang, A. H. Y. Tong, Z. Li, D. S. Goldberg, O. D. King, G. Lesage, M. Vidal, B. Andrews, H. Bussey, C. Boone, and F. P. Roth, "Combining biological networks to predict genetic interactions," *Proceedings of the National Academy of Sciences*, vol. 101, no. 44, pp. 15682–15687, 2004.

[328] H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J. F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A. S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabási,

J. Tavernier, D. E. Hill, and M. Vidal, "High-quality binary protein interaction map of the yeast interactome network," *Science*, vol. 322, no. 5898, pp. 104–110, 2008.
[329] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S," *Philosophical Transactions of the Royal Society of London, Series B, Containing Papers of a Biological Character*, vol. 213, pp. 21–87, 1925.
[330] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, pp. 452–473, 1977.
[331] A. Zheng and A. Goldenberg, "A generative model for dynamic contextual friendship networks," Technical report, Machine Learning Department, Carnegie Mellon University, 2006.
[332] X. Zhu, M. Gerstein, and M. Snyder, "Getting connected: Analysis and principles of biological networks," *Genes Development*, vol. 21, no. 9, pp. 1010–1024, 2007.
[333] B. J. H. Zijlstra, M. A. J. van Duijn, and T. A. B. Snijders, "The multilevel p_2 model: A random effects model for the analysis of multiple social networks," *Methodology*, vol. 2, no. 1, pp. 42–47, 2006.