A Survey of Statistical Network Models

Anna Goldenberg
University of Toronto

Alice X. Zheng
Microsoft Research

Stephen E. Fienberg
Carnegie Mellon University

Edoardo M. Airoldi
Harvard University

December 2009

Contents

Preface

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 fromd [(N8p(.21960sy)8 -1)-312(thp(.2eaiolrlog)1(y)t)27(w)2sgyolv

Chapter 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 coauthorships 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; 318; 319; 68] 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 seventy 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 article, 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 is 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 Wolff [268] at the turn of the last century and Moreno [221] in the 1930s as well as the empirical studies of Stanley Milgram [215; 298] in the 1960s; in mathematics/probability there is the Erdős-Renyi 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 [221] 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., [202; 200; 201; 203; 244; 282; 11]), 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 compleity 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 Erdős-Renyi-Gilbert model. Research on the Erdős-Renyi-Gilbert model (along with works by Katz et al. [166; 168; 167]) engendered the field of random graph theory. In their papers, Erdős and Renyi 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 Erdős-Renyi-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 [149]'s p_1 model extended the Erdős-Renyi-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 [101; 102]. It also allowed for various generalizations to multidimensional network structures [103] and stochastic blockmodels. This approach to modeling network data quickly evolved into the class of p^* or exponential random graph models (ERGM) originating in the work of Frank and Strauss [110] and Strauss and Ikeda [287]. A trio of papers demonstrating procedures for using ERGMs [316; 241; 254] led to the wide-spread 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 chapter 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 chapters.

In the late 1990s, physicists began to work on network models and study their properties

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, Watts 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 t of \statistical physics" or stochastic process models is 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 su cient as descriptors for the actual network data. Moreover, these summary quantities can often be highly misleading as the critique by Stou er et al. [285, 286] of methods used by Barabasi [25] and Vazquez et al. [304] suggest. Barabasi claimed that the dynamics of a number of human activities are scale-free, i.e., he speci cally 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 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 [9], 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 o ers 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 brie y 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 Bollobas [43].

E cient computation on networks. There is a substantial computer science literature dealing with e cient 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.

].

[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 di erent meaning than the networks we consider in this review. The main di erence 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

Chapter 2

Motivation and Dataset Examples

2.1 Motivations for Network Analysis

Why do we analyze networks? The motivation behind network analysis is as diverse as the origin of network problems within differing academic fields. Before we delve into details of the “how” of statistical network modeling, we start with some examples of the “why.” This chapter also includes descriptions of popular datasets for interested readers who may wish to exercise their modeling muscles.

Social scientists are often interested in questions of interpretation such as the meanings of edges in a social network [181]. Do they arise out of friendliness, strategic alliance, obligation, or something else? When the meaning of edges are known, the object is often to characterize the structure of those relations (e.g., whether friendships or strategic alliances are hierarchical or transitive). A large volume of statistically-oriented social science literature is dedicated to modeling the mechanisms and relations of network properties and testing hypotheses about network structure, see, e.g., [280].

Physicists, on the other hand, tend to be interested in understanding parsimonious mechanisms for network formation [28; 235]. For example, a common modeling goal is to explain how a given network comes to have its particular degree distribution or diameter at time t.

Several network analysis concepts have found niches in computational biology. For example, work on protein function classification can be thought of as finding hidden groups in the protein-protein interaction network [7; 8] to gain better understanding of underlying biological processes. Label propagation (node similarity) in networks can be harnessed to help with functional gene annotation [226]. Graph alignment can be used to locate subgraphs that are common among species, thus advancing our understanding of evolution [105]. Motif finding, or more generally the search for subgraph patterns, also has many applications [17]. Combining networks from heterogeneous data sources helps to improve the accuracy of predicted genetic interactions [327]. Heterogeneity of network data sources in biology introduces a lot of noise into the global network structure, especially when networks created for different purposes (such as protein co-regulation and gene co-expression) are combined. [225] addresses network de-noising via degree-based structure priors on graphs. For a review of biological applications of networks, please see [332].

The task of nding hidden groups is also relevant in analyzing communication networks, e.g., in detecting possible latent terrorist cells [30]. The related task of discovering the \roles" of individual nodes is useful for identity disambiguation [36] and for business organization analysis [207]. These applications often take the machine learning approach of graph partitioning, a topic previously known in social science and statistics literature as blockmodeling [199; 89]. A related question is *functional* clustering, where the goal is not to statistically cluster the network, but to discover members of dynamic communities with similar functions based on existing network connectivity [122; 232; 234; 266].

In the machine learning community, networks are often used to predict missing information, which can be edge related, e.g., predicting missing links in the network [238; 73; 198], or attribute related, e.g., predicting how likely a movie is to be a box o ce hit [229]. Other applications include locating the crucial missing link in a business or a terrorist network, or calculating the probability that a customer will purchase a new product, given the pattern of purchases of his friends [142]. The latter question can more generally be stated as predicting individual's preferences given the preferences of her \friends". This research direction has evolved into an area of its own under the name of *recommender systems*, which has recently received a lot of media attention due to the competition by the largest online movie rental company Net ix. The company has awarded a prize of one million dollars to a team of researchers that were able to predict customer ratings of movies with higher than 10% accuracy than their own in-house system [290].

The concept of information propagation also nds many applications in the network domain, such as virus propagation in computer networks [310], HIV infection networks [222; 163; 164], viral marketing [87] and more generally gossiping [170]. Here some work focuses on nding network con gurations optimal for routing, while other research assumes that the network structure is given and focus on suitable models for disease or information spread.

2.2 Sample Datasets

A plethora of data sets are available for network analysis, and more are emerging every year. We provide a quick guided tour of the most popular datasets and applications in each eld.

In his ground-breaking paper, Milgram [215] experimented with the construction of interpersonal social networks. His result that the median length of completed chains was approximately 6 led to the pop-culture coining of the phrase \six degrees of separation." Subjects of subsequent studies ranged from social interactions of monks [259], to hierarchies of elephants [209; 303], to sexual relationships between adults of Colorado [176], to friendships amongst elementary school students [141; 299].

While a lot of biological applications focus on the study of protein-protein interaction networks [114; 115; 184; 248; 328], metabolic networks [158], functional and co-expression gene similarity networks and gene regulatory networks [111; 309], computer science applications revolve around e-mail [207], the internet [97; 63; 151], the web [152; 13], academic paper co-authorship [127] and citation networks [204; 216]. Citation networks have a long history of modeling in different areas of research starting with the seminal paper of de Solla Price [83] and more recently in physics [190]. With the recent rise of online networks, computer science and social science researchers are also starting to examine blogger networks such as *LiveJournal*, social networks found on *Friendster*, *Facebook*, *Orkut*, and dating networks such as *Match.com*.

Terrorist networks (often simulated) and telecommunication networks have come under

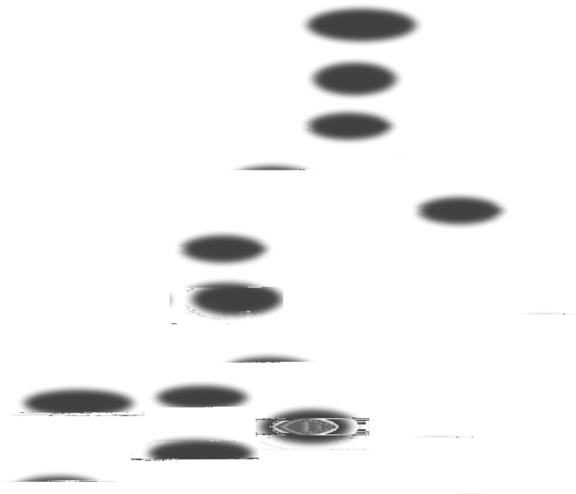

Figure 2.1: Network derived from \whom do you like" sociometric relations collected by Sampson.

shortly after these events. About a year after leaving the monastery, Sampson surveyed all of the novices, and asked them to rank the other novices in terms of four sociometric relations: like/dislike, esteem, personal in uence, and alignment with the monastic credo, retrospectively, at four di erent epochs spanning his stay at the monastery.

The presence of a well de ned social structure within the monastery (the factions) that can be inferred from responses to the survey, as well as the social dynamics of subtle ideological con icts that led to the dissolution of the monastic order, have much intrigued both statisticians and social scientists for the past four decades. Researchers typically consider the faction labels assigned by Sampson to the novices as the anthropological ground truth in their analysis. For example analyses, we refer to [103; 137; 81; 9].

2.2.2 The Enron Email Corpus

The Enron email corpus has been widely studied in recent machine learning network literature. Enron Corporation was an energy and trading company specializing in the marketing of electricity and gas. In 2000 it was the seventh largest company in the United States with re-

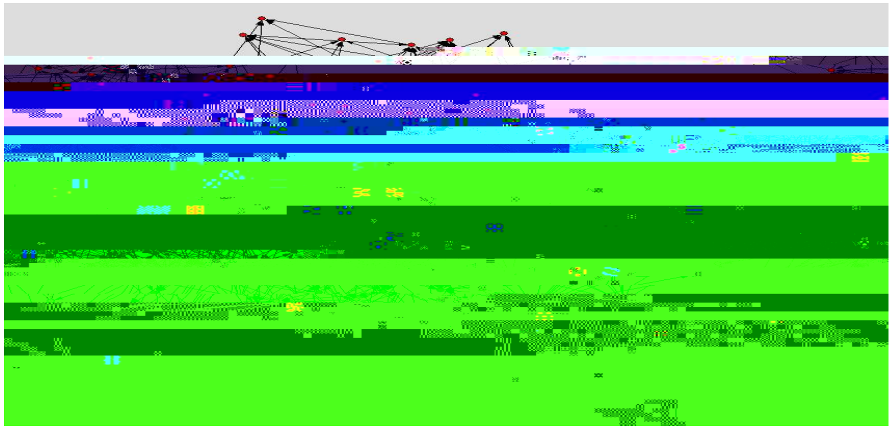

Figure 2.2: E-mail exchange data among 151 Enron executives, using a threshold of a minimum of 5 messages for each link. Source: [153].

in the CALO (Cognitive Assistant that Learns and Organizes) project corrected integrity problems in the dataset.[6] The original FERC dataset contains 619,446 email messages (about 92% of Enron's sta emails), and the cleaned-up CALO dataset contains 200,399 messages from 158 users. Another version of the data consists of the contents of the mail folders of the top 151 executives, containing about 225,000 messages covering a period from 1997 to 2004.[7] Figure 2.2 and Figure 2.3 give network snapshots of the e-mail tra c among these network analysis to visualization. A collection of papers working with the Enron corpus were gathered together in a special 2005 issue of *Computational & Mathematical Organization Theory*, see [58].

2.2.3 The Protein Interaction Network in Budding Yeast

The budding yeast is a unicellular organism that has become a de-facto model organism for the study of molecular and cellular biology [47]. There are about 6,000 proteins in the budding yeast, which interact in a number of ways [64]. For instance, proteins bind together to form protein complexes, the physical units that carry out most functions in the cell [184]. In recent years, a large amount of resources has been directed to collect experimental evidence of physical proteins binding, in an e ort to infer and catalogue protein complexes and their multifaceted functional roles [e.g. 98; 159; 300; 114; 143]. Currently, there are

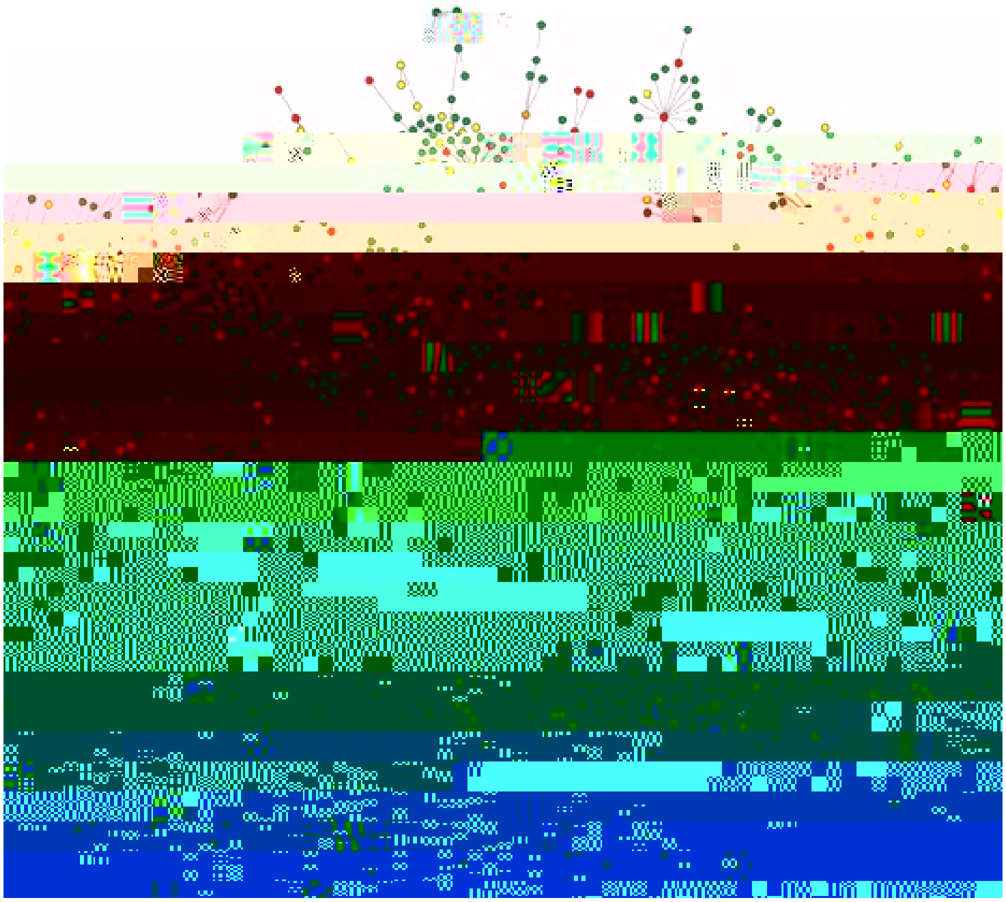

Figure 2.4: A popular image of the protein interaction network in *Saccharomyces cerevisiae*, also known as the budding yeast. The gure is reproduced with permission. Source: [27].

snowball samples collected from past studies; it allows for the construction of relationship networks with more accurate global characteristics. The fully observed friendship networks in all the schools are also a valuable resource and an important contribution of this work.

Wave II data collection occurred 18-months after Wave I in 1996 and followed up on the in-home interviews. The dataset covered 14,738 adolescents and 128 school administrators. Based on the data collected from Wave I and II, Bearman et al. [31] constructed the timed sequence of relationship networks amongst students from the two large schools with saturated sampling. The resulting sexual relationship network bears strong resemblance to a spanning tree as opposed to previously hypothesized core or inverse-core structures[8] (See Figure 2.5.)

Wave III interviews were conducted in 2001 and 2002 with topics including marriage,

[8] A core is a group of inter-connected individuals who sit at the center of the graph and interact with individuals on the periphery. An inverse core is a group of central individuals who are connected to those on the periphery but not to each other.

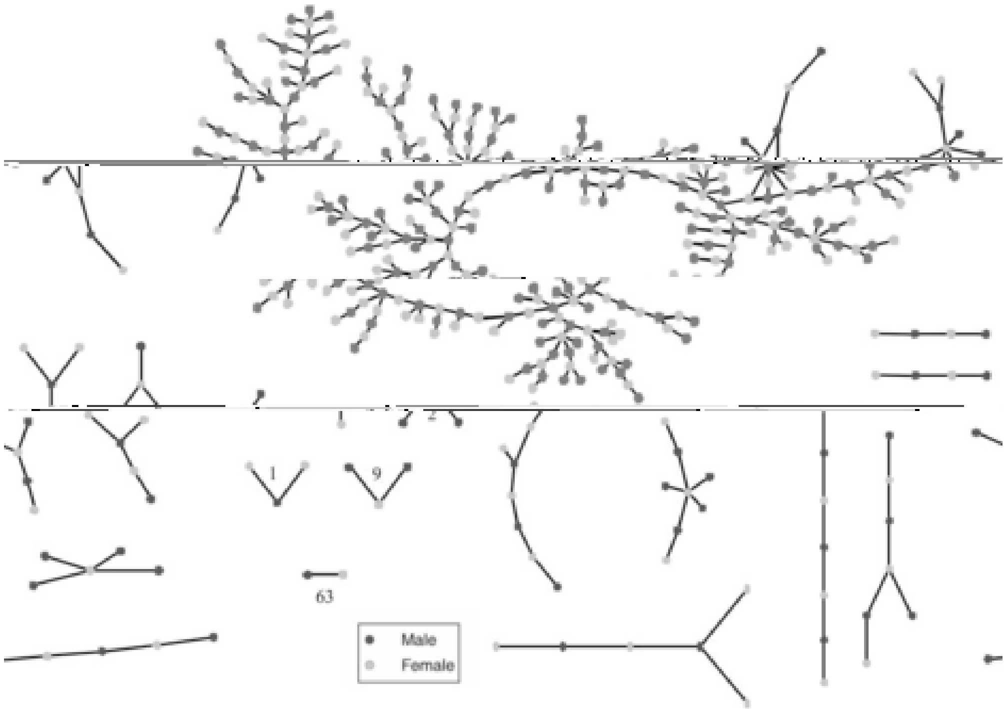

Figure 2.5: The Add Health sexual relationships network of US highschool adolescents. This gure is reproduced with permission. Source: Bearman et al. [31]

childbearing, and sexually transmitted diseases. Of the original Wave I in-home respondents, 15,170 were interviewed again for Wave III. Of these, 13,184 participants provided oral uid specimens for HIV testing. Morris et al. [223 year periods beginning 1973, 1981, 1985, 1989, 1992, 1997, 1999. Christakis and Fowler [65] derive body mass index information on a total of 12,067 individuals who appeared in any of the Framingham Heart cohorts (one \close friend" for each cohort member).[9] There were 38,611 observed family and social ties (edges) to the core 5,124 cohort members.

Through a series of network snapshots and statistical analyses, Christakis and Fowler described the evolution of the \clustering" of obesity in this social network. In particular they claim to have examined whether the data conformed to \small-world," \scale-free," and \hierarchical" types of of random graph network models. Figure 2.6 depicts data on the largest connected subcomponent (the so-called giant component) for the network in 2000, which consists of 2200 individuals. Other analyses in their paper explore attributions of the individuals via longitudinal logistic-regression models with lagged e ects. Subsequently, they have published similar papers focused on the dynamics of smoking behavior over time [66] and on happiness [67], both using the structure of Framingham \o spring" cohort.

This work has come under criticism by others. For example Cohen-Cole and Fletcher note that there are plausible alternative explanations to the network structure based on contextual factors [77], and in a separate paper demonstrate that the same methodology detects \implausible" social network e ects for such medical conditions as acne and headaches as well as for physical height [78]. The authors answer to these criticisms can be found in [108]. The question of the magnitude and signi cance of social network e ects is still a subject of an ongoing debate.

2.2.6 The NIPS Paper Co-Authorship Dataset

The NIPS dataset contains information on publications that appeared in the *Neural Information Processing Systems* (NIPS) conference proceedings, volumes 1 through 12, corresponding to years 1987-1999 | the pre-electronic submission era. The original collection contained scanned full papers made available by Yann LeCunn. Sam Roweis subsequently processed the data to glean information such as title, authorship information, and word counts per document. In total, there are 2,037 authors and 1,740 papers with an average of 2.29 authors per paper and 1.96 papers per author. The NIPS database is available from Sam Roweis' website[10] in raw and *MATLAB* formats along with a detailed description and information on its construction.

Various authors have used the NIPS data to analyze author-to-author connectivity in static [126] as well as dynamic settings [264]. Li and McCallum [197] modeled the text of the documents and Sarkar et al. [265] analyzed the two-mode network (author-word-author) in a dynamic context. In Figure 2.7 we reproduce a graphic illustration of the inferred dynamic evolution of the network from [263].

Figure 2.6: Obesity network from Framingham o spring cohort data. Each node represents one person in the dataset (a total of 2200 in this picture). Circles with red borders denote women, with blue borders { men. The size of each circle is proportional to the body-mass index. The color inside the circle denotes obesity status - yellow is obese (body-mass index 30, green is non-obese. The colors of ties between nodes indicate relationships - purple denotes a friendship or marital tie and orange is a familial tie. This gure is reproduced with permission. Source: [65].

Chapter 3

Static Network Models

A number of basic network models are essentially static in nature. The statistical activities associated with them focus on certain local and global network statistics and the extent to which they capture the main elements of actual realized networks. In this chapter, we briefly summarize two lines of research. The first originates in the mathematics community with the Erdős-Renyi-Gilbert model and led to two types of generalizations: (i) the "statistical physics" generalizations that led to power laws for degree distributions — the so-called scale-free graphs, and (ii) the exchangeable graph models that introduce weak dependences among the edges in a controlled fashion, which ultimately lead to a range of more structured connectivity patterns and enable model comparison strategies rooted in information theory. A second line of research originated in the statistics and social sciences communities in response to a need for models of social networks. The p_1 model of Holland and Leinhardt, which in some sense generalizes the Erdős-Renyi-Gilbert model, and the more general descriptive family of exponential random graph models effectively initiate this line of modeling. Some of these models also have a *generative* interpretation that allows us to think about their use in to types of links, relationships, or interactions between the units, and they may be directed, as in the Holland-Leinhardt model, or undirected, as in the Erdős-Renyi-Gilbert model.

A note about terminology: in computer science, graphs contain nodes and edges; in social sciences, the corresponding terminology is usually actors and ties. We largely follow the computer science terminology in this review.

3.2 The Erdős-Renyi-Gilbert Random Graph Model

The mathematical biology literature of the 1950s contains a number of papers using what we now know as the network model $G(N, p)$, which for a network of N nodes sets the probability of an edge between each pair of nodes equal to p, independently of the other edges, e.g., see Solomonoff and Rapoport [281] who discuss this model as a description of a neural network. But the formal properties of simple random graph network models are usually traced back to Gilbert [119], who examined $G(N, p)$, and to Erdős and Renyi [93]. The Erdős-Renyi-Gilbert random graph model, $G(N, E)$, describes an undirected graph involving N nodes and a fixed number of edges, E, chosen randomly from the $\binom{N}{2}$ possible edges in the graph; an equivalent interpretation is that all $\binom{\binom{N}{2}}{E}$ graphs are equally likely.[1] The $G(N, p)$ model has a binomial likelihood where the probability of E edges is

$$\ell(G(N, p) \text{ has } E \text{ edges} \mid p) = p^E (1 - p)^{\binom{N}{2} - E},$$

P3. If tends to a constant $c > 1$, then a graph in $G(N; p)$ will have a unique "giant" component containing a positive fraction of the nodes, a.s. as $N \to \infty$. No other component will contain more than $O(\log N)$ nodes, a.s. as $N \to \infty$.

A summary of a proof using branching processes is given in the appendix of this chapter. Some of the proof concepts will be useful for discussion of exchangeable graph models in section 3.3.

The Erdős-Renyi-Gilbert model has spawned an enormous number of mathematical papers that study and generalize it, e.g., see [43]. But few of them are especially relevant for the actual statistical analysis of network data. In essence, the model dictates that every node in a graph has approximately the same number of neighbors. Empirically there are few observed networks with such simple structure, but we still need formal tools for deciding on how poor a fit the model provides for a given observed network, and what kinds of generalized network models appear to be more appropriate. This has led to two separate literatures, one of which has focused on formal statistical properties associated with estimating parameters of network models — the p_1 and exponential random graph models described below — and a second that identifies selected predicted features of models and empirically checks observed networks for those features. The latter is largely associated with papers emanating from statistical physics and computer science, several of which are described in detail in chapter 4.

3.3 The Exchangeable Graph Model

The exchangeable graph model provides the simplest possible extension of the original random graph model by introducing a weak form of dependence among the probability of sampendent given the binary string representations of the incident nodes. They are *exchangeable* in the sense of De Finetti [82].

From a statistical perspective, the exchangeable graph model we survey here [1; 5] provides perhaps the simplest step-up in complexity from the random graph model [93; 119]. In the data generation process, the bit strings are equally probable but the induced probabilities of observing edges are different. A class of random graphs with such a property has been recently rediscovered and further explored in the mathematics literature, where the class of such graphs is referred to as *inhomogeneous* random graphs [45]. An alternative and arguably more interesting set of specifications can be obtained by imposing dependence among the bits at each node. This can be accomplished by sampling sets of dependent probabilities from a family of distributions on the unit hypercube, $\vec{p}_n \in [0, 1]^K$, and then sampling the bits independently given these dependent probabilities.

1. Sample node-specific K-bit binary strings for each node $n \in N$

   $\vec{p}_n$ hypercube $(\vec{\ };\ ;\ )$, where $> (K \ 1) \ > 0;$

   b_{nk} Bern (p_{nk}), for $k = 1, \dots, K$

2. Sample directed edges for all node pairs $n, m \in N \ N$

   Y_{nm} Bern $q(\vec{b}_n, \vec{b}_m)$,

In the hypercube distribution[3], $\vec{\ };\ ;$ control the frequency, variability and correlation of the bits within a string, respectively; and q maps binary pairs of strings into the unit interval.

In the exchangeable graph model, the number of bits, K, captures the complexity of the graph. For instance, for $K < N$ the model provides a compression of the graph. For directed graphs the function q

giant component emerges because a number of smaller components must intersect with high probability. In exchangeable graph models however, the giant component has a peculiar structure; connected components are themselves connected to form the giant component as soon as bit strings that match on two bits appear with high probability. Figure 3.1 provides a graphical illustration of this intuition. Nodes that *bridge* two connected components are

Figure 3.1: *Left panel.* An example adjacency matrix that correspond to a fully connected component among 100 nodes. *Right panel.* The clustering coefficient as a function of   on a sequence of graphs with 100 nodes. Here   = 12, and $\log(\ _i) = \frac{1}{K}$ for every $i = 1 \ldots K$.

evident in the left panel. Note that there are no nodes that bridge three components, as bit strings that match on three bits is an unlikely event in a graph with 100 nodes.

Given a graph, we can infer the corresponding set of binary strings from data. The likelihood that correspond to an exchangeable graph model is simple to write,

$$\ell(Y|\ \ ) = \int d\vec{b}_{1:N} \prod_{n,m} \Pr(Y_{n,m}|\vec{b}_n,\vec{b}_m,q) \prod_n \Pr(\vec{b}_n|\ \ ) \ ;$$

where $= (\tilde{\ };\ ;\ )$ or an appropriate set of parameters. We can apply standard inference techniques [2; 9]. Fitting an exchangeable graph model allows us to assess the complexity of an observed graph, leveraging notions from information theory. For instance, we can use the minimum description length (MDL) principle to decide how many bits we need to explain the observed connectivity patterns with high probability. We can also quantify how much *information* is retained at different bit-lengths, and plot the corresponding information profile for $K < N$ and an entropy histogram for any given value of K.

The exchangeable graph model allows for algorithmic comparison of any set of statistical models that are proposed to summarize an observed graph. As an illustration, consider an observed graph G and two alternative models A and B. Rather than comparing how well models A and B recover the degree distribution of G or other graph statistics, and independently of whether it makes sense to directly compare the two likelihoods of A and B (in fact, these models need not have a likelihood), we can proceed as follows.

1. Given a graph G, fit models $A(\;_a)$ and $B(\;_b)$ to obtain an estimate of their parameters $\;_a^{Est}$ and $\;_b^{Est}$ respectively.

2. Sample M graphs at random from the support of $A(\;_a^{Est})$ and $B(\;_b^{Est})$.

3. Compute the distributions of summary statistics based on notion from information theory, such as information profile and entropy histogram, corresponding to the $2M$ graphs sampled from A and B.

4. Compare models in terms of the distribution on the statistics above, such as the complexity of the two models' supports and their similarity to the complexity of G.

The exchangeable graph model also allows for evaluation of the distribution of the number of bit strings with l matching bits, for any integer $l < K$. In theory this distribution leads to expectations on the number of nodes that bridge l communities, where the members of each community have only one out of l matching bits. In practice, we may want to specify K in advance so that each bit corresponds to a well defined property. For instance, in applications to biology, nodes may correspond to proteins and the K bits encode presence or absence of specific protein domains. The distribution on the number of l matchings leads to p-values that summarize how unexpected it is to observe binding events among a set of proteins that share a certain combination of domains.

Overall, the exchangeable graph model introduces weak dependences among the edges of a random graph in a controlled fashion, which ultimately lead to a range of more structured connectivity patterns and enable model comparison strategies rooted in notions from information theory. The focus here is not on modeling per se. In fact, the model is kept as simple as possible. Rather, the focus is on modeling as a means to establish a technical link between graph connectivity and node attributes. This technical link is useful to address some of the issues listed in Chapter 5

3.4 The p_1 Model for Social Networks

A conceptually separate thread of research developed in parallel in the statistics and social sciences literature, starting with the introduction of the p_1 model. Consider a directed graph on the set of n nodes. Holland and Leinhardt's p_1 model focuses on dyadic pairings and keeps track of whether node i links to j, j to i, neither, or both. It contains the following parameters:

: a base rate for edge propagation,

${}_i$ (expansiveness): the e ect of an outgoing edge from i,

${}_j$ (popularity): the e ect of an incoming edge into j,

${}_{ij}$ (reciprocation/mutuality): the added e ect of reciprocated edges.

Let $P(0;0)$ be the probability for the absence of an edge between i and j, $P_{ij}(1;0)$ the probability of i linking to j (\1" indicates the outgoing node of the edge), $P_{ij}(1;1)$ the probability of i linking to j *and* j linking to i. The p_1 model posits the following probabilities (see [149]):

$$\log P_{ij}(0;0) = {}_{ij}; \tag{3.1}$$
$$\log P_{ij}(1;0) = {}_{ij} + {}_i + {}_j + ; \tag{3.2}$$
$$\log P_{ij}(0;1) = {}_{ij} + {}_j + {}_i + ; \tag{3.3}$$
$$\log P_{ij}(1;1) = {}_{ij} + {}_i + {}_j + {}_j + {}_i + 2 + {}_{ij}: \tag{3.4}$$

In this representation of p_1, ${}_{ij}$ is a normalizing constant to ensure that the probabilities for each dyad $(i;j)$ add to 1. For our present purposes, assume that the dyad is in one and only one of the four possible states. The reciprocation e ect, ${}_{ij}$, implies that the odds of observing a mutual dyad, with an edge from node i to node j and one from j to i, is enhanced by a factor of $\exp({}_{ij})$ over and above what we would expect if the edges occured independently of one another.

The problem with this general p_1 representation is that there is a lack of identi cation of the reciprocation parameters. The following special cases of p_1 are identi able and of special interest:

1. ${}_i = 0$, ${}_j = 0$, and ${}_{ij} = 0$. This is basically an Erdős-Renyi-Gilbert model for directed graphs: each directed edge has the same probability of appearance.

2. ${}_{ij} = 0$, *no reciprocal e ect*. This model e ectively focuses solely on the degree distributions into and out of nodes.

3. ${}_{ij} =$, *constant reciprocation*. This was the version of p_1 studied in depth by Holland and Leinhardt using maximum likelihood estimation.

4. $\quad_{ij} = \quad + \quad_i + \quad_j$, *edge-dependent reciprocation*. Fienberg and Wasserman [101, 102] described this model and how to nd maximum likelihood estimate for the parameters.

In the constant reciprocation setting, the elevated probability of reciprocal edges does not depend on the dyad, whereas edge-dependent reciprocation dictates multiplicative increases of the reciprocation probability based on node-speci c parameters.

The likelihood function for the *p*

lik estimates. Fien 1 7b et al. [

3.6 Exponential Random Graph Models

Under the assumption that two possible edges are dependent only if they share a common node,[6] Frank and Strauss [110] proved the following characterization for the probability distribution of undirected Markov graphs:

$$\Pr\{Y = y\} = \exp$$

these models where the major problem of double-counting is mitigated but not overcome. Hunter and Handcock [155] estimate likelihood ratios for nearby $\{\theta_i\}$ using a MCMC procedure related to the work of Geyer and Thompson [118]. Their estimation procedure can be used for models based on distributions in the curved exponential family.

Robins et al. [256] describe problems associated with the estimation of parameters in many ERGMs, involving near degeneracies of the likelihood function and thus of methods used to estimate parameters using maximum likelihood. For example, for a certain combination of ERGM statistics, the likelihood function may have multiple, clearly distinct modes, and there are very few network configurations — often radically different from each other — that have non-zero probabilities. This is a topic of current theoretical and empirical investigation rooted in the theory of discrete exponential families [136; 251]. For a discussion of mixing times of MCMC methods for ERGMs and the relevance to convergence and degeneracies, see [35].

There are two carefully constructed packages of routines that are available for analyzing network data using ERGMs: *statnet*[7] and *SIENA*[8]. These packages focus on the use of MCMC methods for estimating the parameters in ERGMs.

Remark. It is possible to express the current formulation of exponential random graphs using the formalism of undirected graphical models and the Hammersley-Clifford theorem [76; 33]. We can write the likelihood of an arbitrary undirected graph as

$$\Pr(\mathbf{y}|\theta) = \frac{\prod_{c \in C} \psi(\mathbf{y}_c|\theta_c)}{z}, \tag{3.8}$$

where $\mathbf{y}_c$ denotes the nodes in clique c, θ_c denotes the corresponding set of parameters, ψ are non-normalized potentials over the cliques, and $z = \sum \prod_{c \in C} \psi(\mathbf{y}_c|\theta_c)$ is the normalization constant. If the likelihood is in the exponential family, then the log potentials are linear in θ_c and "features" $u(\mathbf{y}_c)$, and we can write:

$$\begin{aligned}
\Pr(\mathbf{y}|\theta) &= \exp\Big\{\sum_{c \in C} \log \psi(\mathbf{y}_c|\theta_c) - \log z\Big\} \\
&= \exp\Big\{\sum_{c \in C} \theta_c^{\top} u(\mathbf{y}_c) - \log z\Big\} \\
&= \exp\Big\{\theta^{\top} u(\mathbf{y}) - \log z\Big\}.
\end{aligned}$$

Within the exponential family, the advantage is that computing derivatives and likelihood and deriving the corresponding EM algorithm are feasible, although possibly computationally expensive, by using variational approximation strategies and Monte Carlo methods. A lot of methodology on the subject has been developed in the area of machine learning. There,

[7] A package written for the R statistical environment described at

undirected graphs appear primarily in the context of relational learning and imaging. For an in-depth discussion on exact and approximation methods and for references see [247; 308].

3.7 Random Graph Models with Fixed Degree Distribution

The Erdős-Renyi-Gilbert random graph model is fully symmetric and the expected degree (the number of edges associated with a node) is the same for all nodes in the graph, following a binomial distribution. A number of natural extensions of the Erdős-Renyi-Gilbert model result in varying node degrees. For example,

- the preferential attachment model [26] captures the formation of hubs in a graph (see section 4.1);

- the one-parameter “small-world” model [320] interpolates between an ordered finite-dimensional lattice and an Erdős-Renyi-Gilbert random graph in order to produce local clustering and triadic closures (see section 4.2).

Albert and Barabasi [12] describe a number of variants on these themes. Many of the investigators exploring the use of such models often focus on the empirical degree distribution, claiming for example that it follows a power-law in many real world networks (cf. [26; 232; 69; 91]). The papers utilizing these “statistical physics” style models often talk about fixed-degree distributions [e.g., 239], and they either fix the degree-distribution parameters or compute distributions that are conditional on some function of the degree distributions or sequences, such as their expectations (cf. [235; 70]). Software is available to largely as a mechanism for avoiding the degeneracies and near degeneracies observed when unconditional maximum likelihood is used, cf. section 3.6 and [256]. Snijders [274] does

266; 217]. This literature is now voluminous and seemingly unconnected to the statistical blockmodel work.

The basic idea, in both the model-based and algorithmic approaches as well as the community detection literature, is that nodes that are heavily interconnected should form a block or community. The nodes are reordered to display the blocks down the diagonal of the adjacency matrix representing the network. Moreover, the connections between nodes in di erent blocks appear in much sparser o -diagonal blocks. In model-based approaches, the partition of the nodes maximizes a statistical criterion linked to the model, e.g., a likelihood function, whereas most algorithmic solutions maximize ad hoc criteria related to the \density" of links within and between blocks.

More formally, a blockmodel is a model of network data that relies on the intuitive notion of *structural equivalence*: two nodes are de ned to be structurally equivalent if their connectivity with similar nodes is similar | this is a \soft" de nition.[9]

Also note that the pairs of group memberships that underlie interactions need not be equal; this fact is useful for characterizing asymmetric interaction networks. Equality may be enforced when modeling symmetric interactions.

Inference in the blockmodel is challenging, as the integrals that need to be solved to compute the likelihood cannot be evaluated analytically. For simplicity, the likelihood is

$$\ell(Y \mid \vec{\alpha}, B) = \int_{\Pi} \int_{Z} \Pr(Y \mid Z, B)\, \Pr(Z \mid \Pi)\, \Pr(\Pi \mid \vec{\alpha})\, dZ\, d\Pi\,.$$

While the inner integral is easily solvable[10], the outer integral is not. Exact inference is thus not an option. To complicate things, the number of observations scales as the square of the number of nodes, $O(N^2)$. Sampling algorithms such as Monte Carlo Markov chains are typically too slow for real-size problems in the natural, social, and computational sciences. Airoldi et al. [9] suggest a nested variational inference strategy to approximate the posterior distribution on the latent variables, (Π, Z). (Variational methods scale to large problems without loosing much in terms of accuracy [3; 49; 308].)

Bickel and Chen [37], the most recent contribution to this literature, brings new twists to the model-based approach of community discovery. They use a blockmodel to formalize a given network in terms of its community structure. The main result of this work implies that community detection algorithms based on the modularity score of Newman and Girvan [122] are (asymptotically) biased. It shows that using modularity scores can lead to the discovery of an incorrect community structure even in the favorable case of large graphs, where communities are substantial in size and composed of many individuals. This work also proves that blockmodels and the corresponding likelihood-based algorithms are (asymptotically) unbiased and lead to the discovery of the correct community structure. The proof relies on the exchangeability results developed in the statistics community [15; 165] applied to paired measurements [84].

3.9 Latent Space Models

The intuition at the core of latent space models is that each node $i \in N$ can be represented as a point z_i in a "low dimensional" space, say $\mathbb{R}^k$. The existence of an edge in the adjacency matrix, $Y(i, j) = 1$, is determined by the distance among the corresponding pair of nodes in the low dimensional space, $d(z_i, z_j)$, and by the values of a number of covariates measured on each node individually. The latent space model was first introduced by Hoff et al. [146] with applications to social network analysis, and has been recently extended in a number of directions to include treatment of transitivity, homophily on node-specific attributes, clustering, and heterogeneity of nodes [144; 137; 183].

[10] The inner integral resolves into a series of sums, each one over the support of an individual $\vec{z}$ variable. The support is the same for all such $\vec{z}$ variables, and it is given by the N vertices of the K-dimensional unit hypercube. In other words, the inner integral is a series of sums, each over the same N elements.

Note that it is possible to re-parametrize $Z_i = \;_i !_i$ to separate the position in a latent reference space, , from its magnitude, $_i$

variational methods for a computationally e cient approximation to the posterior. These methods can scale to large matrices (e.g., millions of nodes) because of the simpli ed approximation, but at an unknown cost to accuracy. It would be interesting to explore computational tradeo s for the latent space cluster model [

giant component, G, in which each node can be reached from every other node.

The following formal argument comes from lecture notes by Guetz and Constantine [133] based on proofs given by Janson et al. [161]. Pick a node $v\ 2\ N$. If v is connected to all of the nodes in G, then we say that v is *saturated* in G

Chapter 4

Dynamic Models for Longitudinal Data

In chapter 3 we focused on models for static networks, that consider a cross-section of a real network at a given point in time. However, real networks often contain a dynamic component. In the language of networks, dynamics can be translated into the birth and death of edges and nodes. For example, in a friendship network, new nodes may be introduced at any time and old nodes may drop out due to inactivity; links of friendships and alliances may be even more brittle. Dynamic network modeling has been a neglected sibling of static network properties to observed data. For this reason, we view them as "pseudo-dynamic" models and discuss three examples here: the Erdős-Renyi-Gilbert model, preferential attachment model, and small-world models.

For example, we can view the Erdős-Renyi-Gilbert model $G(N; E)$, itself as a dynamic process used to generate a random graph:

- start from the graph of N unconnected nodes at time 0;

- at each subsequent time step, add a different edge to the network with probability $p = E / \binom{N}{2}$.

By convention, we usually fix the number of nodes at N, although we can extend the process to allow for addition of nodes. This model assumes that edges (and nodes) are not removed once they are added. The degree distribution for $G(N; E)$ is binomial. But as N gets large, Np tends to a constant, so it is approximately Poisson. Durrett [91] provides a rich discussion for situating this dynamic description with the tradition of discrete time random walks and branching processes. In particular, he uses this representation to explore the emergence of the giant component described in section 3.2 (see appendix of chapter 3).

The Erdős-Renyi-Gilbert model is simple and easy to study but does not address many issues present in real network dynamics. One of the major criticisms [26] of this model centers on the fact that it does not produce a scale-free network, i.e., the resulting node degree distribution does not follow a power law. The network literature is replete with claims that many real networks exhibit the power-law phenomenon, (cf. [12]), and much subsequent research has focused on how various generalizations of the Erdős-Renyi-Gilbert model conform to the power law degree distribution. Molloy and Reed [219] were the first to describe how to construct graphs with a general degree distribution and they went on to describe the emergence of the giant component in that context as well [220].

Barabasi and Albert [26] described a dynamic preferential attachment (PA) model specifically designed to generate scale-free networks. At time 0, the model starts out with N_0 unconnected nodes. At each subsequent time step, a new node is added with $m \leq N_0$ edges. The probability that the new node is connected to an existing node is proportional to the degree of the latter. In other words, the new node picks m nodes out of the existing network according to the multinomial distribution

$$p_i = \frac{\kappa_i}{\sum_j \kappa_j},$$

where κ_i denotes the (undirected) degree of node i. This model, which was described much earlier in the statistical literature by Yule [329] and Simon [269], is intended to describe networks that grow from a small nucleus of nodes and follow a "rich-get-richer" scheme. The assumption is that, for instance, a new web page will more likely link via a URL to a well-known web page as opposed to a little-known one. Mitzenmacher [218] gives a brief history of generative models for power law distributions.

The preferential attachment model of Barabasi and Albert results in a network with

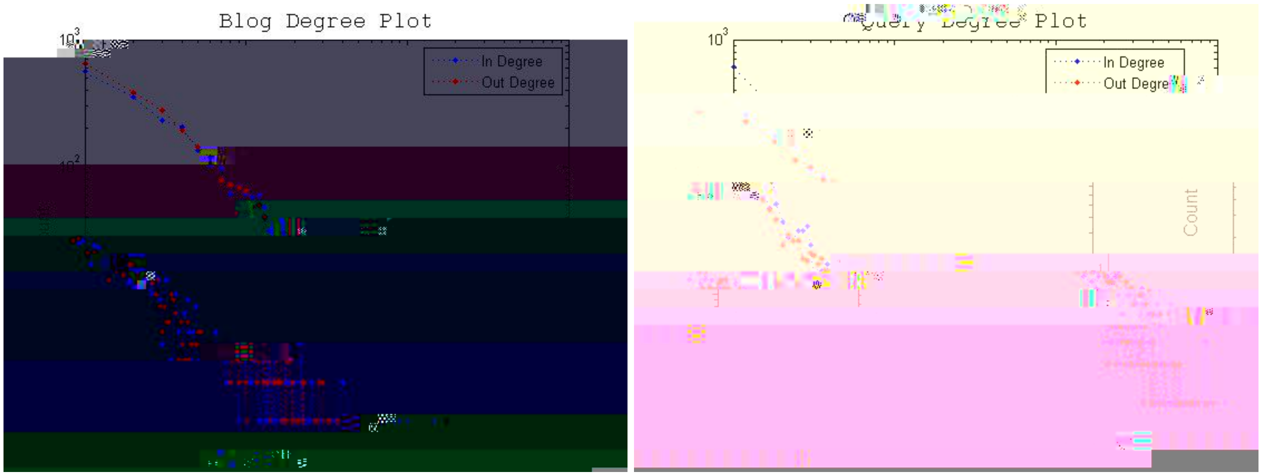

Figure 4.1: Log-log plots of degree distributions for a query data bases and a blog data base from a company database. Left: Blog indegree and outdegree distributions. Right: Query indegree and outdegree distributions. Source: Data from an unnamed large company, stored in iLab, Carnegie Mellon University.

has turned into a well analyzed methodology [195] with an e cient algorithm for model tting, analysis of the parameter space, and model selection. This work goes further in understanding real network structure and provides a way for principled graph sampling.

4.2 Small-World Models

Watts and Strogatz [320] proposed a small-world model which can be thought of as a \pseudo-dynamic" model in the sense we described in section 4.1. This one-parameter \small-world" model interpolates between an ordered nite-dimensional lattice and an Erdős-Renyi-Gilbert random graph in order to produce local clustering and triadic closures. Bollobas and Chung [44] had previously noted that adding random edges to a ring of N nodes drastically reduces the diameter of the network. The Watts-Strogatz model begins with a ring lattice with N nodes and k edges per node, and randomly rewires each edge with probability p. As p goes from 0 to 1, the construction moves toward an Erdős-Renyi-Gilbert model. They and others who followed, studied the behavior of such small-world networks when $0 < p < 1$. This model is not dynamic although it is often used to describe networks that evolve over time. Figure 4.2 shows a small-world graph for $n = 25$ nodes and 2 rewirings per node.

Kleinberg [174] introduced a variation on the small-world model where random edges are added to a xed grid. Starting with an underlying nite-dimensional grid, he added shortcut edges, where the probability that two nodes are connected by a long edge depends on the

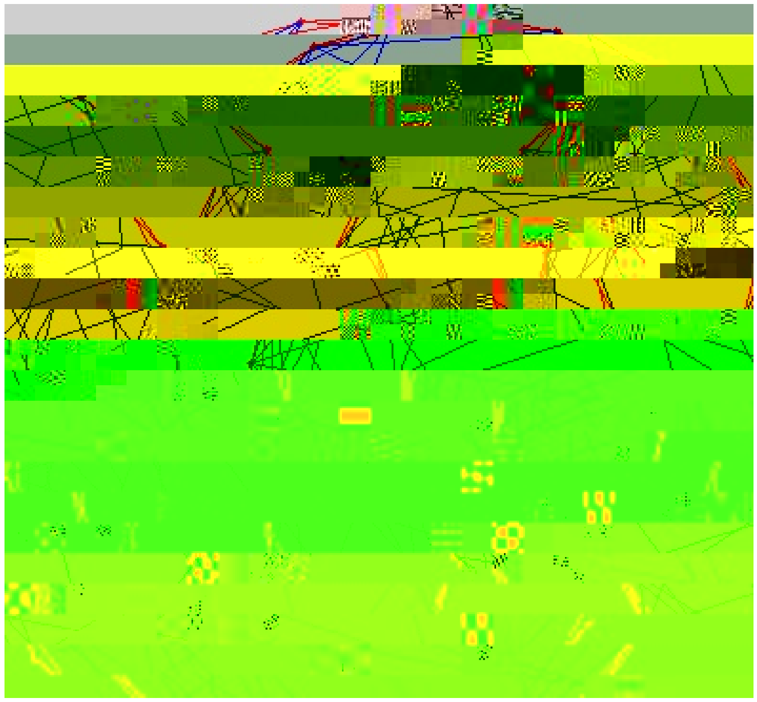

Figure 4.2: Small-world graph for $N = 25$ nodes and 2 rewirings per node. The red edges form the ring lattice and the blue edges the rewiring. This graph was generated using the Java applet at http://cs.gmu.edu/~astavrou/smallworld.html

Several follow-up works have made adjustments to Kleinberg's rewiring procedure in attempt to improve the understanding and efficiency of the navigability of networks. For example, Clauset and Moore [72] suggested to rewire a long distance edge from node x, if while performing a greedy walk over to y, the original topology of the network did not allow to reach y within $T_{\mathtt{thresh}}$ steps. The edge was rewired to the place where the search gave up (the node reached after $T_{\mathtt{thresh}}$ steps of the walk).They show that through this rewiring procedure the network degree distribution converges to a power law, where $\; = \;_{\mathtt{rewired}}$. Their work also studied finite size effects and showed that $_{\mathtt{opt}} \to d$, as $n \to \infty$ rather slowly.

Sandberg [260, 261] and Sandberg and Clarke [262] introduced a different rewiring scheme with the end goal to make the network more amenable to statistical analysis. Starting with N nodes on a ring, each with two neighbor links and a long range link, the model of Sandberg [260] randomly rewires a graph in the following steps:

> at each time step $j = 1, 2, 3, \ldots,$ choose a random starting node x and a target node y

This defines a Markov chain on a collection of labeled graphs. Sandberg and Clarke [262] conjecture that when the chain achieves stationarity, the distribution of distances spanned by long-range links is (close to) theoretical optimum for search and the expected length of searches is polylogarithmic. They support the conjecture by a series of simulations. This methodology has been applied to the study of peer-to-per (P2P) networks.

Durrett [91] discusses links between small-world models and stochastic processes. Typical usage of small-world models include empirical analyses involving aggregate summary statistics (see, e.g., [18; 231]). There are as yet no formal statistical methods for examining the evolution of small-world network models and for assessing their fit to network data measured over time.

4.3 Duplication-Attachment Models

Duplication-Attachment models were originally developed in the computer science theory community to study the world wide web as a directed graph [175; 185]. These models aim at describing properties of a snapshot of the web graph at a specific time, that is, a static directed graph. The data generating process underlying these models, however, is explicitly dynamic. The following example demonstrates some basic assumptions behind the dynamics. Consider a newly added web page A, which provides a new node in the web graph. The creator of web page A will then add *hyper-links* to it, which provide new directed edges in the web graph. In particular, *some* of these hyper-links will point to other web pages regardless of whether their topical content matches the topical content of web page A, but *most* of these hyper-links will point to web pages with a topical content that closely matches the topical content of web page A.

Technically, there are many possible specifications and variants. The basic duplication-attachment model proposed and analyzed by Kumar et al. [185] is as follows. Denote the graph at time t as $G_t = (N_t; E_t)$. At each step, say $t + 1$, one new node N is added to G_t. The new node is connected to a *prototype* node m, chosen uniformly at random among those in N_t. Then d out-links are added to node N. The ith out-link is chosen as follows: with probability α the destination node is chosen uniformly at random among those in N_t, and with probability $1 - \alpha$ the destination node is taken to be the ith out-link of the prototype node m. Note that this is possible since the algorithm generates a constant degree graph. Rather than proposing estimation strategies for the two parameters $(\alpha; d)$ of this particular duplication-attachment model, the goal of the analysis of Kumar et al. [185] is on deriving results about topological properties of duplication-attachment graphs, described as functions of the two parameters $(\alpha; d)$. Recent extensions of this model include a model where fractions of both out-links and in-links of the prototype node m are *copied* by the newly added node N [193]. The goal of the analyses in this line of research, however, remains that of replicating properties of observed graphs, with a few exceptions. In the biological context, duplication-attachment models have appeared to be useful in modeling protein-protein interaction networks. For example, Ratmann et al. [245] proposed a mixture of preferential attachment and duplication divergence with parent-child attachment model to assess evolutionary dynamics of protein interaction networks of *H. pylori* and *P. falciparum*. They proposed a likelihood-free MCMC-based routine to estimate posterior of network summary statistics. A more general review of work in modeling dynamics (evolution) on the basis of protein-protein interaction data is available in [246].

Wiuf et al. [326] have developed a recursive construction of the likelihood for duplication-attachment models, effectively enabling principled statistical data analysis, estimation and inference.

4.4 Continuous Time Markov Chain Models

The use of continuous Markov processes to model dynamic networks was first proposed by Holland and Leinhardt [148] and Wasserman [312] and most recently studied by Snijders and colleagues [275; 276]. As shall become clear in this section, continuous Markov process models (CMPM) are intimately tied to the ERGM models described in section 3.6. Within the CMPM family, network edges are taken to be binary (either absent or present, but not weighted), and the evolution occurs one edge at a time. Model variants arise due to the many possible specifications of edge change probability. Some exceptions to this general approach include the party model of Mayer [206], where multiple edges are allowed to change at the same time, and the work of Koskinen and Snijders [179], which deals with Bayesian parameter inference methods for the case where not all edge modifications are observed.

We begin by providing a quick reminder of continuous Markov processes, borrowing notation from [275]. Define $\{Y(t) \mid t \in \mathcal{T}\}$ to be a stochastic process, where $Y(t)$ has a finite outcome space $\mathcal{Y}$ and Y a binary vector of length $\binom{N}{2}$. We use the shorthand $q_{ij}(\mathbf{y})$ to denote the propensity for the edge between node i and j to flip into its opposite value under configuration $\mathbf{y}$. The function $q_{ij}(\mathbf{y})$ completely specifies the dynamics of the network model. We now review several variants of CMPM which differ only in their definition of $q_{ij}(\mathbf{y})$.

Independent arc, reciprocity, and popularity models. The *independent arc* model employs the simplest definition of $q_{ij}(\mathbf{y})$:

$$\text{Independent arc model:} \quad q_{ij}(\mathbf{y}) = {}_{y_{ij}}; \tag{4.4}$$

i.e., Y_{ij} changes from 0 to 1 at a rate ${}_0$, and from 1 to 0 at rate ${}_1$. In this model, modification to one edge does not depend on the setting of other edges. The model is simple enough that the transition probabilities $\Pr(t)$ can be derived in closed form (see, e.g., Taylor and Carlin [292] p. 362-364). Maximum likelihood parameter estimation for this model was discussed in [278].

In the *reciprocity* model, the rate of change in y_{ij} depends only on the reciprocal edge y_{ji}:

$$\text{Reciprocity model:} \quad q_{ij}(\mathbf{y}) = {}_{y_{ij}} + {}_{y_{ij}} y_{ji}. \tag{4.5}$$

Thus, if no link currently exists between nodes i and j, then the propensity for adding either directed edge is ${}_0$; if one directed edge exists, then the reciprocal edge is added with propensity ${}_0 + {}_0$. If one directed edge exists, then it is deleted with rate ${}_1$. If both edges exist, then the deletion propensity for either is ${}_1 + {}_1$. The transition matrix $\Pr(t)$ can be derived but has a complicated form [189; 272].

Along the same line of development, the *popularity* model and the *expansiveness* model [312; 313] define the change rate for edge y_{ij} to be dependent on y_{+j}, the in-degree of node j, or y_{i+}, the out-degree of node i:

$$\text{Popularity model:} \quad q_{ij}(\mathbf{y}) = {}_{y_{ij}} + {}_{y_{ij}} y_{+j}; \tag{4.6}$$

$$\text{Expansiveness model:} \quad q_{ij}(\mathbf{y}) = {}_{y_{ij}} + {}_{y_{ij}} y_{i+}. \tag{4.7}$$

Edge-oriented dynamics. Snijders [276] outlines two categories of transition dynamics: edge-oriented and node-oriented. In both cases, the intensity matrix is factored into two components: one controls the *opportunity* for change, and the other specifies the propensity of change. More precisely, the continuous time Markov process is now split into two sub-processes; the first operating in the continuous time domain and dictating *when* a change should occur; the second dealing with the probability of the discrete event of individual edge flips. Both edge-oriented and node-oriented dynamics can be interpreted as stochastic optimizations of a potential function $f(\mathbf{y})$ on the network configuration. The difference is that, in the edge-oriented case, f is based on global statistics of the network, ef Tf 9.721 3.008 Td [(+)]TJ

Using $y(i, j, z)$ to denote the configuration where the edge e_{ij} has the value $z \in \{0, 1\}$, edge-oriented dynamics can be written in the following general form:

$$q_{ij}(\mathrm{y}) = \lambda p_{ij}(\mathrm{y}), \tag{4.8}$$

where

$$p_{ij}(\mathrm{y}) = \frac{\exp(f(y(i, j, 1 - y_{ij})))}{\exp(f(y(i, j, 0))) + \exp(f(y(i, j, 1)))}. \tag{4.9}$$

Thus, in edge-oriented dynamics each edge follows an independent Poisson process, so that the time until the next event has an exponential distribution with parameter λ. When an event occurs for edge $i \to j$, the edge flips to its opposite value with probability $p_{ij}(\mathrm{y})$.

The potential function $f(\mathrm{y})$ is usually defined as a linear combination of network statistics:

$$f(\mathrm{y}) = \sum_k \theta_k s_k(\mathrm{y}). \tag{4.10}$$

This should start to look familiar. Indeed the CMPM process with edge-oriented dynamics is equivalent to the Gibbs sampling process for ERGMs (where the next edge to be updated is selected randomly). The statistics $s_k(\mathrm{y})$ for node k take on the usual forms (see Table 4.1).

Number of directed arcs:	$s_1(\mathrm{y}) = \sum_{i,j} y_{ij}$
Number of reciprocated arcs:	$s_2(\mathrm{y}) = \sum_{i,j} y_{ij} y_{ji}$
Number of pairs of arcs with the same target:	$s_3(\mathrm{y}) = \sum_{i,j,k} y_{kj} y_{ji}$
Number of pairs of arcs with the same origin:	$s_4(\mathrm{y}) = \sum_{i,j,k} y_{ik} y_{ij}$
Number of paths of length two:	$s_5(\mathrm{y}) = \sum_{i,j,k} y_{ij} y_{jk}$
Number of transitive triplets:	$s_6(\mathrm{y}) = \sum_{i,j,k} y_{ij} y_{ik} y_{jk}$

Table 4.1: The table of network statistics for a directed social network.

The statistics in Table 4.1 assume directed graphs, however it is easy to come up with the corresponding statistics for undirected graphs. For example, in the undirected case all the edges are "reciprocal" and thus s_1 and s_2 are combined into $s'(\mathrm{y}) = \sum_{i,j>i \in N} y_{ij}$.

Due to their close relations to ERGMs, edge-oriented models suffer the same fate of degeneracy. For example, if the parameter θ_6 for transitive triplets is not too small, then with high probability the simulated network will be a complete graph. However, compared to static networks, degeneracy in the longitudinal case is not as much a concern, as the complete graph will only emerge at some distant time in the future.

Node-oriented dynamics. Fully node-oriented dynamics [275] defines the intensity matrix as

$$q_{ij}(\mathbf{y}) = \lambda_i p_{ij}(\mathbf{y}), \tag{4.11}$$

where

$$p_{ij}(\mathbf{y}) = \frac{\exp(f_i(\mathbf{y}(i, j, 1 - y_{ij})))}{\sum_{h \neq i} \exp(f_i(\mathbf{y}(i, h, 1 - y_{ih})))}. \tag{4.12}$$

Thus the independent Poisson processes for determining edge change *opportunity* are now defined for each node (with intensity λ_i) as opposed to each edge. Given the opportunity for edge change, each node seeks to optimize its own potential function as defined by

$$f_i(\mathbf{y}) = \sum_k \beta_k s_{ik}(\mathbf{y}). \tag{4.13}$$

The function $f_i(\mathbf{y})$ is similar to the global potential $f(\mathbf{y})$ in Equation 4.10 but only aggregates over the local neighborhood of node i. Node i favors changing the incident edge that would lead to the biggest increase in its potential.

Edge-node mixed dynamics. Snijders [276] also suggested a form of mixed dynamics where the opportunity for change is edge-oriented, but the potential functions are node-oriented:

$$q_{ij}(\mathbf{y}) = \rho \frac{\exp(f_i(\mathbf{y}(i, j, 1 - y_{ij})))}{\sum_{h \neq i} \exp(f_i(\mathbf{y}(i, h, 1 - y_{ih})))}. \tag{4.14}$$

Thus the opportunity to modify each edge $i \to j$ follows independent Poisson processes with parameter ρ. But given the opportunity for change, the probability of an actual flip depends on node i's local network configuration.

Remark. Parameter estimation in CPCM models has until recently been done via method of moments, where the expected values are obtained through MCMC on simulated networks [273]. Koskinen and Snijders [179] proposed a Bayesian inference method that allows for computation of the posterior distribution of the parameters and treats missing values more adequately. For details of the procedure, please refer to Koskinen and Snijders [179].

4.5 Discrete Time Markov Models

In this section, we outline three recent proposals of dynamic network models operating in the discrete time domain (see also [22]). All three models have the Markov property and represent the likelihood as a sequence of factored conditional probabilities

$$\Pr(Y^1, Y^2, \ldots, Y^T) = \Pr(Y^T \mid Y^{T-1}) \Pr(Y^{T-1} \mid Y^{T-2}) \cdots \Pr(Y^2 \mid Y^1)), \tag{4.15}$$

where f^1)$_{ih}$))).7(Pr()]TJ/F43 11.299-2936TJ/F4g8d [(Y)mn.247 TdJ/F45(Y)mn665TJ/F18-4.5 Discrete

4.5.2 Dynamic Latent Space Model

Sarkar and Moore [264] extended the static latent space model of Ho et al. [146] (cf. section 3.9

known machine learning researchers over time. The dynamics of the researchers' latent positions allowed for an insight into the evolution of the machine learning community.

Sarkar et al. [265] also proposed a richer model based on [124], which improved upon previous work in two ways. One of the differentiating features of this work was the ability to simultaneously embed words and authors into the latent space, which allowed for representation of a two-mode network. The major advantage, however, was the inference method — the authors proposed a Kalman-filter like dynamic procedure, which allowed for estimation of the posterior distributions over the positions of the authors in the latent space. Proposed procedure was applied to a simulated NIPS dataset.

The impact of this line of work is dichotomous: first, it offers an explanation of the network at every time step, and second, it enables an accurate and efficient prediction of the state of the network at a time step in the future. The proposed inference procedures made it possible for network modeling to scale to large dynamic collections of data. The drawback of this approach is the lack of an explicit mechanism that could explain the dynamics behind the real networks.

Another latent model for citation networks was developed in the physics community. Leicht et al. [190] proposed to use latent variables to capture the grouping of papers that have similar citation profiles over time. The network in this case is a directed acyclic graph and the nodes are papers rather than authors. Using as example a set of opinions from the US Supreme Court and their citations between the years of 1789 and 2007, the authors showed how a simple latent model was able to recover, in a completely unsupervised manner, the different eras in US Supreme court opinion references. The parameters of the model, except for the number of latent classes, were estimated using an EM algorithm. Different numbers of latent classes were tested and each revealed something new about the underlying data. The authors also compared the latent method to a clustering based on network *modularity* [233]. Even with the information about time (directionality in the graph) removed, the latent variable model was still able to discover the same split between two groups of opinions that happened around 1937. The network modularity clustering in a way validated the outcome of the latent model.

In a separate experiment, Leicht et al. [190] showed that deterministic approaches such as “hubs and authorities” and eigenvector centrality [171] discovered interesting network properties that were not revealed by the statistical models. The deterministic analyses showed several significant drops in the age of authorities sited, meaning that once in a while, the younger set of opinions became the new authorities and that the process happened in a “decisive” manner, rather than gradually. In this way, deterministic network analysis approaches complement statistical models.

4.5.3 Dynamic Contextual Friendship Model (DCFM)

The dynamic contextual friendship model (DCFM) of Goldenberg and Zheng [128] represents an attempt to capture several aspects of the complexity of the evolution of real social networks over time. In a real-life friendship network, people may meet and interact with each other under different contexts (e.g., school, work projects, social outings, etc.), and the strength of interpersonal relationships change over time based on these interactions. DCFM o ers such a mechanism for network evolution, where edges have weights that indicate the strength of the relationship, and each node is given a distribution over social interaction spheres (contexts). Context is de ned to be any activity where people may interact with each other. At each given time step, each node chooses a random context according to the node's distribution over contexts. Nodes that appear in the same context update the weights of the links between them. The probability of a weight increase (or decrease) depends on whether the pair had a chance to meet (a coin toss in a model) and the \friendliness" parameter of the individuals involved. The possibility of both positive and negative weight

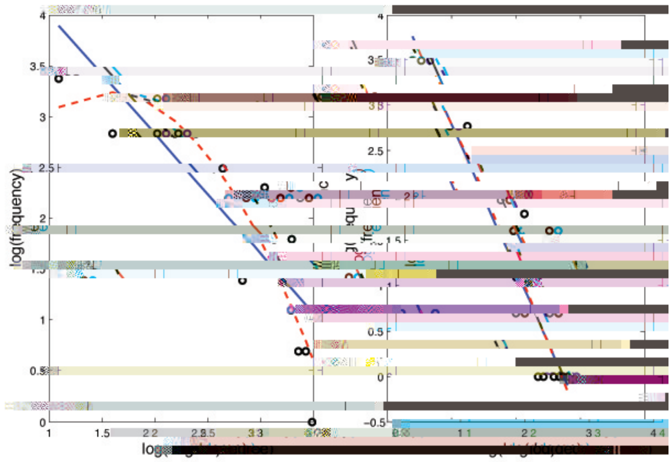

Figure 4.3: Log-log plot of the degree distributions of a network with 200 people. $_i$ is drawn from Beta(1; 3) for the plot on the left, and from Beta(1; 8) for the right hand side. Solid lines represent a linear t and dashed lines quadratic t to the data. Contexts are drawn every 4W1 2t]3simestepsple.every

peopl2t]3siowe2t]3si6200t]37imest2t]3siepsple.-4305(Con)27(texts)-3sie(wicn)27hes2t]3siow4.44wCo2(st7 cr changes contexts and is very friendly or because the contexts themselves tend to be large. Also, weighted network data are hard to come by and thus pseudo-weights often have to be used.

The DCFM model is important in its own right: the life-mimicking, rich generative

Chapter 5

Issues in Network Modeling

There are a number of major statistical modeling and inferential challenges in the analysis of network data that go well beyond those described in previous sections of this article. These relate to both the quality and the ease of statistical inference and we mention a few of them here:

Network Visualization. With the rise of online social networks and network modeling, we have seen a proliferation of visualization tools, especially those based on variations of constraint-based spring model algorithms, e.g., see the discussion and references in Shneiderman and Aris [267]. The automated algorithms often use node degrees or some form of distance metric between nodes to arrange their placement. For example, *SoNIA*[1] is a popular package for visualizing dyna5e(and)rnisbnetwkalda;.,b their own drawbacks such as sensitivity to the starting point, are not realizable for networks on a really large scale. The key to network modeling and parameter estimation is to take selected subgraphs. For details, see the many papers by Ove Frank [109; 295] and others [125; 135; 258]. Wiuf and Stumpf [325] and Stumpf and Thorne [288] recently adopted known links | information that is incomplete and available only for a few organisms. In the sociological literature on organizations, there is often interest in distinguishing among organizations on the basis of their network structure, so there would clearly be interest in utilizing methodology for prediction based on network structure. Because making predictions of various sorts from dynamic network models ts well within the machine learning paradigm, we expect to see many more papers on the topic in the not too distant future.

Embeddability. Underlying most dynamic network models is a continuous time stochastic process even though the data used to study the models and their implications may come in the form of repeated snapshots at discrete time points (epochs) | a form of time sampling as opposed to node sampling referred to above | or cumulative network links. In such circumstances we need to take special care in how we represent and estimate the continuous-time parameters in the actual data realizations used to t models. This is known in the statistical literature as the

Chapter 6

Summary

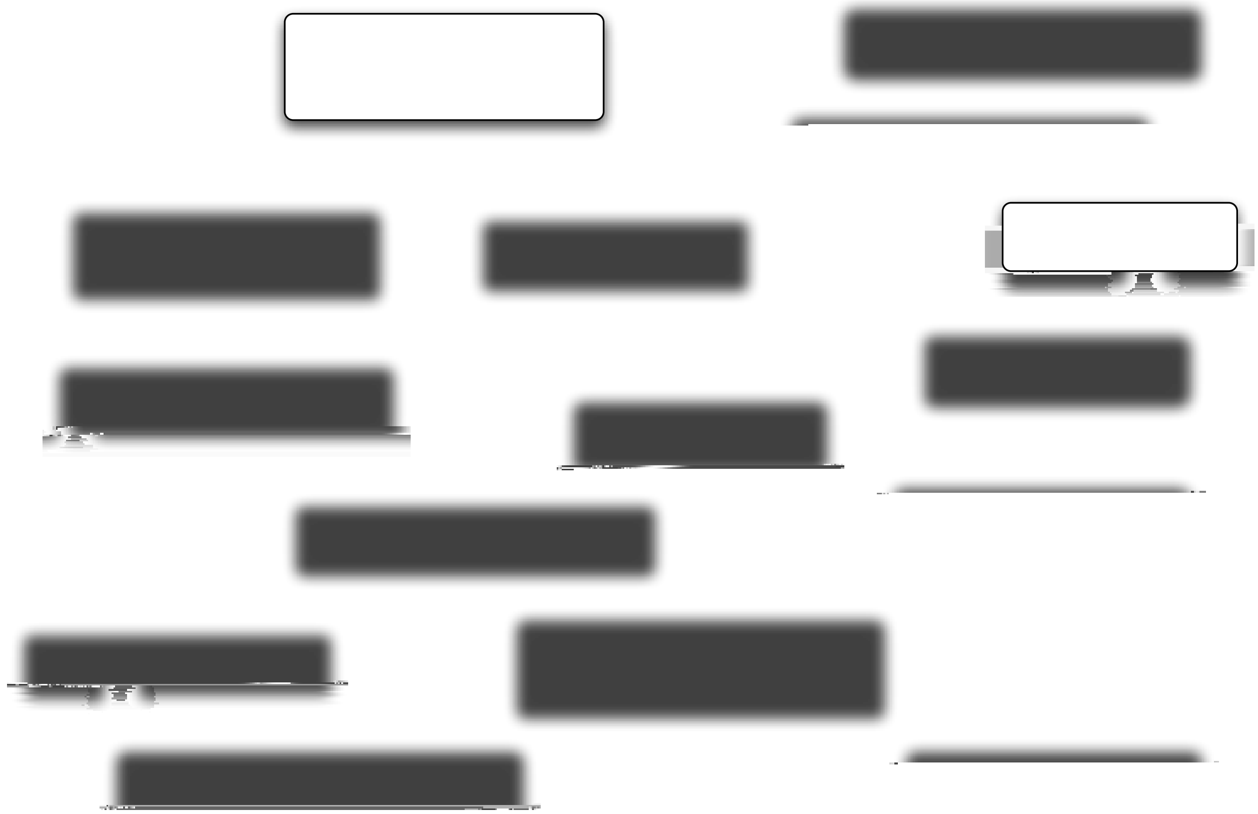

Figure 6.1: Network summarizing the relations between models discussed in our review. White nodes denote static models, yellow nodes { \pseudo-dynamic" and green { dynamic models. Arrows indicate inspiration or in uence of the model at the source on the model at the target.

equivalence of the nodes, whereas latent space models assume the existence of an embedding of the network in a low dimensional space. These models allow for better understanding of the data in cases where it is believed to contain hidden structure.

We divided the category of dynamic models into continuous time Markov models and discrete time Markov models. CMPM (section 4.4) assumes that the adjacency matrix evolves according to a continuous Markov chain whose intensity matrix can depend on various edge and node dynamics. Discrete time Markov network models deal with a set of network snapshots observed at various time po28me pnusE6li7To-p 0 -14,-8me pn2uou6crete timecasstatistics or machine learning perspective, the biggest breakthroughs are to be made in the areas of inference and dynamic modeling. Creating a model or perhaps xing an existing one in such a way that provides realistic generative and inference mechanisms which can identi ably infer parameters of a large real world network would make a great contribution to the statistical network modeling community.

Acknowledgments

This research was partly supported by United States National Institute of General Medical Sciences Center of Excellence grant P50 GM071508, by National Science Foundation grants DBI-0546275, IIS-0513552, by National Institutes of Health grant R01 GM071966 to Princeton University, by National Science Foundation grant DMS-0907009 to Harvard University, and by National Science Foundation grant DMS-0631589 and partial support from U.S. Army Research O ce Contract W911NFo910360 to the Department of Statistics, Carnegie Mellon University. Edoardo M. Airoldi was a postdoctoral fellow in the Department of Computer Science and the Lewis-Sigler Institute for Integrative Genomics at Princeton University when a large portion of this work was carried out. We thank three anonymous reviewers for their valuable comments, as well as their helpful additions and corrections to our citation list. We thank Joseph Blitzstein and Pavel Krivitsky for a careful reading and the correction of a number of infelicities. We nally wish to thank Laszlo Barabasi and Zoltan Oltvai; Peter Bearman, James Moody, and Katherine Stovel; James Fowler and Nicholas Christakis; Purnamrita Sarkar and Andrew Moore for giving permission to re-print gures from their original papers [27; 31; 65; 263].

Bibliography

[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. Model-based clustering for social networks: Discussion. *Journal of the Royal Statistical Society, Series A*, 170(2):330{331, 2007.

[3] E. M. Airoldi. Getting started in probabilistic graphical models. *PLoS Computational Biology*, 3(12):e252, 2007.

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

[5] E. M. Airoldi. The exchangeable graph model. Technical Report 1, Department of Statistics, Harvard University, 2009.

[6] 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*, 7(2):13{22, 2005.

[7] 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*, pages 82{89. ACM Press, New York, 2005.

[8] 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.

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

[10]

[39] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA, 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*, 170(2):332, 2007.

[41] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993{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. Bollobas. *Random Graphs*. Cambridge University Press, New York, 2nd edition, 2001.

[44] B. Bollobas and F. R. K. Chung. The diameter of a cycle plus a random matching. *SIAM Journal on Discrete Mathematics*, 1(3):328{333, 1988.

[45] B. Bollobas, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3{122, 2007.

[46] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Suppl. 3):7280{7287, 2002.

[47] D. Botstein, S. A. Chervitz, and J. M. Cherry. Yeast as a model organism. *Science*, 277(5330):1259{1260, 1997.

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

[49] M. Braun and J. McAuli e. 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*. W. W. Norton & Company, New York, 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*, 41(2): 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*, 6:79{141, 1980.

[54] K. M. Carley. Group stability: A socio-cognitive approach. In E. Lawler, B. Markovsky, C. Ridgeway, and H. Walker, editors, *Advances in Group Processes,* pages 1{44. JAI Press, Greenwich, CT, 1990.

[55] K. M. Carley. Smart agents and organizations of the future. In L. Lievrouw and S. Livingstone, editors, *The Handbook of New Media,* pages 206{220. Sage, Thousand Oaks, CA, 2002.

[56] K. M. Carley and A. Newell. The nature of the social agent.

[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*, 357(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*, 358:2249{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*, 337:a2338, 2008.

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

[69] F. Chung and L. Lu. *Complex Graphs and Networks*. American Mathematical Society, Providence, RI, 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*, 100(11):6313{6318, 2003.

[71] A. Clauset. Finding local community structure in networks. *Physical Review E*, 72(2): 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*, 453:98{101, 2008.

[74] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):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. Cli ord. Markov random elds in statistics. In G. R. Grimmett and D. J. A. Welsh, editors, *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, pages 19{32. Oxford University Press, 1990.

[77] E. Cohen-Cole and J. M. Fletcher. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, 27:1382{1387, 2008.

[78] E. Cohen-Cole and J. M. Fletcher. Detecting implausible social network e ects in acne, height, and headaches: Longitudinal analysis. *British Medical Journal*, 337: a2533, 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*, 9(1), 2009.

[80] A. Davis, B. B. Gardner, M. R. Gardner, and J. J. Wallach. *Deep South: A Social*

[94] P. Erdős and A. Renyi. The evolution of random graphs. *Magyar Tud. Akad. Mat.*

[107] A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks II. *Internet Mathematics*, 4(1):87{112, 2007.

[108] J. Fowler and N. Christakis. Estimating peer e ects on health in social networks. *Journal of Health Economics*, 27(5):1400{1405, 2008.

[109] O. Frank. Network sampling and model tting. In P. J. Carrington, J. Scott, and S. S. Wasserman, editors, *Models and Methods in Social Network Analysis*, pages 31{56. Cambridge University Press, 2005.

[110] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832{842, 1986.

[111]

M. Tyers. Systematic identi cation of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415:180{183, 2002.

[144] P. D. Ho . Random e ects models for network data. In R. Breiger, K. M. Carley, and P. E. Pattison, editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 303{312. The National Academies Press, Washington, D.C., 2003.

[145] P. D. Ho . Modeling homophily and stochastic equivalence in symmetric relational data. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 657{664. MIT Press, 2008.

[146] P. D. Ho , A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090{1098, 2002.

[147] P. W. Holland and S. Leinhardt. Local structure in social networks. *Sociological Methodology*, 7:1{45, 1976.

[148] P. W. Holland and S. Leinhardt. A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5{20, 1977.

[149] P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373):33{65, 1981.

[150] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109{137, 1983.

[151] P. Holme, J. Karlin, and S. Forrest. An integrated model of tra c, geography and economy in the internet. *ACM SIGCOMM Computer Communication Review*, 38(3): 7{15, 2008.

[152] B. A. Huberman and L. A. Adamic. Growth dynamics of the world-wide web. *Nature*, 401:131, 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*, 30(4):297{308, 2008.

[155] D. R. Hunter and M. S. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565{583, 2006.

[170] D. Kempe, J. Kleinberg, and E. Tardos. In uential nodes in a di usion model for social networks. In *Automata, Languages and Programming*, volume 3580 of *Lecture Notes in Computer Science*, pages 1127{1138. Springer Berlin / Heidelberg, 2005.

[171] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604{632, 1999.

[172] J. M. Kleinberg. Navigation in a small world | it is easier to nd short chains between points in some networks than others. *Nature*, 406:845, 2000.

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

[174] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems (NIPS)*, volume 14. MIT Press, Cambridge, MA, 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*, volume 1627 of *Lecture Notes in Computer Science*, pages 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*, 38(1):79{88, 1994.

[177] E. D. Kolacyzk. *Statistical Anaysis of Network Models*. Springer, New York, 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*, 137(12):3930{3938, 2007.

[180] G. Kossinets. E ects of missing data in social networks. *Social Networks*, 28(3):247{268, 2006.

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

[182] V. E. Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43{52, 2002.

[183] P. N. Krivitsky, M. S. Handcock, A. E. Raftery, and P. D. Ho . Representing degree distributions, clustering, and homophily in social networks with latent cluster random e ects models. *Social Networks*, 31(3):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. Peregr n-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*, 440(7084):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*, pages 57{65, 2000.

[186] S. L. Lauritzen. Rasch models with exchangeable rows and columns. In J. M. Bernardo et al., *Bayesian Statistics 7*, pages 215{232. Oxford University Press, 2003.

[187] S. L. Lauritzen. Exchangeable Rasch matrices. *Rendiconti di Matematica, Serie VII*, 28(1):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*, 104(31):12931{12935, 2007.

[189] R. T. A. J. Leenders. Models for network dynamics: A Markovian framework. *Journal of Mathematical Sociology*, 20:1{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*, 59(1):75{83, 2007.

[191] J. Leskovec and C. Faloutsos. Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 631{636. ACM Press, New York, 2006.

[192] 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*, volume 3721 of *Lecture Notes in Computer Science*, pages 133{145. Springer Berlin / Heidelberg, 2005.

[193] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densi cation laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 177{187. ACM Press, New York, 2005.

[194] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densi cation and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1): 2, 2007.

[195] 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.

[196] L. Li, D. Alderson, J. C. Doyle, and W. Willinger. Towards a theory of scale-free graphs: De nition, proper ties, and implications. *Internet Mathematics*, 2(4):431{523, 2005.

[197]

[207] A. McCallum, A. Corradda-Emmanuel, and X. Wang. Topic and role discovery in social networks. In *Proceedings of the International Joint Conference on Arti cial Intelligence*, pages 786{791, 2005.

[208] A. McCallum, X. Wang, and N. Mohanty. Joint group and topic discovery from relations and text. In E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. Zheng, editors, *Statistical Network Analysis: Models, Issues and New Directions*, volume 4503 of *Lecture Notes in Computer Science*, pages 28{44. Springer Berlin / Heidelberg, 2007.

[209] K. McComb, C. Moss, S. M. Durant, L. Baker, and S. Sayialel. Matriarchs as repositories of social knowledge in African elephants. *Science*, 292(5516):491{494, 2001.

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

[211] J. W. McDonald, P. W. F. Smith, and J. J. Forster. Markov chain Monte Carlo exact inference for social networks. *Social Networks*, 29(1):127{136, 2007.

[212] 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*, pages 524{532. ACM Press, New York, 2008.

[213] M. M. Meyer. Transforming contingency tables. *Annals of Statistics*, 10(4):1172{1181, 1982.

[214] M. Middendorf, E. Ziv, and C. H. Wiggins. Inferring network mechanisms: The Drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences*, 102(9):3192{3197, 2005.

[215] S. Milgram. The small world problem. *Psychology Today*, 1(1):60{67, 1967.

[216] D. Mimno and A. McCallum. Mining a digital library for in uential authors. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 105{106. ACM Press, New York, 2007.

[217] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan. Finding strongly-knit clusters in social networks. *Internet Mathematics*, 5(1-2):155{174, 2008.

[218] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226{251, 2004.

[219] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2{3):161{180, 1995.

[220] M. Molloy and B. Reed. The size of the largest component of a random graph on a xed degree sequence. *Combinatorics, Probability and Computing*, 7:295{306, 1998.

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

[222] M. Morris and M. Kretzschmar. Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17(3{4):299{318, 1995.

[223] 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*, 96(6):1091{1097, 2006.

[224] M. Morris, M. S. Handcock, and D. R. Hunter. Speci cation of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software*, 24(4), 2008. http://www.jstatsoft.org/v24/i04.

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

[226] 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*, 9(Suppl. 1):S4, 2008.

[227] 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*, 21(Suppl. 1):i302{i310, 2005.

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

[229] J. Neville and D. Jensen. Collective classi cation 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.

[230] J. Neville, D. Jensen, L. Friedland, and M. Hay. Learning relational probability trees. In

[233] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577{8582, 2006.

[234] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 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*, 99(Suppl. 1):2566{2572, 2002.

[236] K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077{1087, 2001.

[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*, 5(4):11, 2006.

[249] E. Reid and H. Chen. Mapping the contemporary terrorism research domain: Researchers, publications, and institutions analysis. In *Intelligence and Security Informatics*, volume 3495 of *Lecture Notes in Computer Science*, pages 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*, volume 3073 of *Lecture Notes in Computer Science*, pages 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*, 3:446{484, 2009.

[252] J. M. Roberts, Jr. Simple methods for simulating sociomatrices with given marginal totals. *Social Networks*, 22(3):273{283, 2000.

[253] G. L. Robins and P. E. Pattison. Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25:5{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*, 64(3):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*, 26(3): 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*, 29(2):192{215, 2007.

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

[258] M. J. Salganik and D. D. Heckathorn. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, 34:193{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 G•teborg University, G•teborg, Sweden, 2005.

[261] O. Sandberg. Neighbor selection and hitting probability in small-world graphs. *Annals of Applied Probability*, 18(5):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. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18, pages 1145{1152. MIT Press, Cambridge, MA, 2005.

[264] P. Sarkar and A. W. Moore. Dynamic social network analysis using latent space models. *SIGKDD Explorations: Special Edition on Link Mining*, 7(2):31{40, 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 Arti cial 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 E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. Zheng, editors, *Statistical Network Analysis: Models, Issues and New Directions*, volume 4503 of *Lecture Notes in Computer Science*, pages 140{157. Springer Berlin / Heidelberg, 2007.

[267] B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):733{740, 2006.

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

[269] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3{4):425{440, 1955.

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

[271] B. Singer and S. Spilerman. The representation of social processes by Markov models. *The American Journal of Sociology*, 82(1):1{54, 1976.

[272] T. A. B. Snijders. The transition probabilities of the reciprocity model. *Journal of Mathematical Sociology*, 23(4):241{253, 1999.

[273] T. A. B. Snijders. The statistical evaluation of social network dynamics. *Sociological Methodology*, 31:361{395, 2001.

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

[275] T. A. B. Snijders. Models for longitudinal network data. In P. J. Carrington, J. Scott, and S. S. Wasserman, editors, *Models and Methods in Social Network Analysis*, chapter 11. Cambridge University Press, New York, 2005.

[276] T. A. B. Snijders. Statistical methods for network dynamics. In S. R. Luchini et al., editors, *Proceedings of the XLIII Scienti c Meeting, Italian Statistical Society*, pages 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 Classi cation*, 14(1):75{100, 1997.

[278] T. A. B. Snijders and M. A. J. van Duijin. Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, and P. Terna, editors, *Simulating Social Phenomena*, pages 493{512. Springer, Berlin, 1997.

[279] T. A. B. Snijders and M. A. J. van Duijn. Conditional maximum likelihood estimation under various speci cations of exponential random graph models. In J. Hagberg, editor, *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank*, pages 117{134. Department of Statistics, University of Stockholm, Stockholm, Sweden, 2002.

[280] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New speci cations for exponential random graph models. *Sociological Methodology*, 36:99{153, 2006.

[281] R. Solomono and A. Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biology*, 13(2):107{117, 1951.

[282] S. Spilerman. Structural analysis and the generation of sociograms. *Behavioral Science*, 11:312{318, 1966.

[283] M. Stephens. Bayesian analysis of mixtures with an unknown number of components | an alternative to reversible jump methods. *Annals of Statistics*, 28(1):40{74, 2000.

[284] D. Stork and W. Richards. Nonrespondents in communication network studies. *Group & Organization Management*, 17(2):193{209, 1992.

[285] D. B. Stou er, R. D. Malmgren, and L. A. N. Amaral. Comment on Barabasi, Nature 435, 207 (2005). http://arXiv.org/abs/physics/0510216, 2005.

[286] D. B. Stou er, 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*, 85(409):204{212, 1990.

[288] M. P. H. Stumpf and T. Thorne. Multi-model inference of network properties from incomplete data. *Journal of Integrative Bioinformatics*, 3(2):32, 2006. http://journal.imbio.de/index.php?paper_id=32.

[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*, 102(12):4221{4224, 2005.

[290] S. Swasey. Net ix awards $1 million Net ix 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*, 320(5882):1465{1470, 2008.

works[292] A. Hadi. Car(I.11.9552 Tf 108.918 0 T.7219)-326(102)50(a27(t38r)50c)50(e)50(ducfor)-3239(to7(t38r,)-

J. M. Rothberg. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403(6770):623{627, 2000.

[301] M. A. J. van Duijn, T. A. B. Snijders, and B. J. H. Zijlstra. p_2: A random e ects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2):234{254, 2004.

[302] 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*, 31(1):52{62, 2009.

[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. Vazquez, J. G. Oliveira, Z. Dezsö, K. Goh, I. Kondor, and A.-L. Barabasi. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73:036127, 2006.

[305] E. Volz and D. D. Heckathorn. Probability based estimation theory for respondent driven sampling. *Journal of O cial Statistics*, 24(1):79{97, 2008.

[306] E. Volz and L. A. Meyers. Epidemic thresholds in dynamic contact networks. *Journal of the Royal Society Interface*, 6(32):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*, 417(6887):399{403, 2002.

[308] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1{2):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*, 106(16):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)*, pages 25{34, 2003.

[311] Y. Y. Wang and G. Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):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*, 75(370):280{294, 1980.

[314] S. S. Wasserman and C. Anderson. Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks*, 9(1):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*, 61(3):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 E. M. Airoldi, D. M. Blei, S. E. Fienberg,

[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*, 101(44):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. Barabasi, J. Tavernier, D. E. Hill, and M. Vidal. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898): 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*, 213:21{87, 1925.

[330]