8

*An Introduction to Random Graphs, Dependence Graphs, and p**

Stanley Wasserman
University of Illinois

Garry Robins
University of Melbourne

We begin with a graph (or a directed graph), a single set of nodes $\mathcal{N}$, and a set of lines or arcs $\mathcal{L}$. It is common to use this mathematical concept to represent a *social network*. We use the notation of Wasserman and Faust (1994), especially Chapters 13 and 15. There are extensions of these ideas to a wide range of social networks, including multiple relations, affiliation relations, valued relations, and social influence and selection situations (in which information on attributes of the nodes is available). Later chapters in this volume discuss such generalizations.

The model p^* was first discussed by Frank and Strauss (1986), who termed it a distribution for a *Markov random graph*. Further developments, especially commentary on estimation of distribution parameters, were given by Strauss and Ikeda (1990). Wasserman and Pattison (1996) further elaborated this family of models, showing how a Markov parametric assumption provides just one of many possible sets of parameters. This family, with its variety and extensions, was named p^*, a label by which it has come to be known. The parameters reflect structural concerns, which are assumed to govern the probabilistic nature of the underlying social and/or behavioral process.

The development of p^* presented here is different from that found in Wasserman and Pattison (1996) and Anderson, Wasserman, and Crouch (1999), but similar to the presentation in Pattison and Wasserman (1999). Rather than looking at p^* as an approximate autologistic regression model, we begin with *dependence graphs* and show how this family of models follows naturally from such graphs. Dependence graphs are quite useful when distinguishing among a variety of different random graph types, with unique dependence assumptions – which is one of the goals of this chapter. This chapter provides an introduction to dependence graphs, with more detail discussed in later chapters.

8.1 Some Notation

A *social network* is a set of n actors and a collection of r social relations that specify how these actors are related to one another. As defined by Wasserman and Faust (1994,

Chapter 3), a social network can also contain a collection of attribute characteristics, measured on the actors. Such actor attribute variables can be important "explanatory" variables for relational "response variables"; due to the introductory nature of this chapter, such variables, as well as the *social influence* (Robins, Pattison, and Elliot 2001) and *social selection* (Robins, Elliot, and Pattison 2001) models, are discussed in later chapters.

Here, we let $r = 1$, focusing just on networks with single relations and assuming that relational ties take on just two values (see also Frank and Strauss 1986; Strauss and Ikeda 1990; Strauss 1992; Rennolls 1995). Extensions to multiple and valued relations (more recently presented by Pattison and Wasserman 1999, and Robins, Pattison, and Wasserman 1999) are also possible (see also Frank 1991, 1997, and Frank and Nowicki 1993).

We let $\mathcal{N} = \{1, 2, \ldots, g\}$ denote the set of actors, and $\mathcal{X}$ denote a particular relation defined on the actors. Specifically, $\mathcal{X}$ is a set of ordered pairs recording the presence or absence of relational ties between pairs of actors. This social relation can be represented by a $g \times g$ matrix $\mathbf{X}$, with elements

$$X_{ij} = \begin{cases} 1 \text{ if } (i, j) \in \mathcal{X}, \\ 0 \text{ otherwise.} \end{cases}$$

We use a variety of graph characteristics and statistics throughout this chapter; such quantities are defined in the early chapters of Wasserman and Faust (1994). We assume throughout that $\mathbf{X}$ and its elements are random variables. Typically, these variables are assumed to be interdependent, given the interactive nature of the social processes that generate and sustain a social network. In fact, one of the new ideas for social network analysis used by the p^* family of models is a *dependence graph*, a device that allows one to consider which elements of $\mathbf{X}$ are independent. We define this graph in the next section.

One of the "tricks" that allows the basic p^* model to be extended to multivariate and valued relations is the creation of new relations that are the converses, compositions, or intersections of the measured relations. Here, we define three new arrays for each relational tie, $\mathbf{X}_{ij}^{+}$, $\mathbf{X}_{ij}^{-}$, and $\mathbf{X}_{ij}^{c}$, which will be useful for the estimation of model parameters.

Let $\mathbf{X}_{ij}^{+}$ be the array formed from $\mathbf{X}$ where the tie from i to j is forced to be present:

$$\begin{aligned} (X_{ij}^{+})_{mn} &= X_{ij}, \text{ if } (m, n) \neq (i, j) \\ &= 1, \text{ if } (m, n) = (i, j). \end{aligned}$$

Thus, $\mathbf{X}_{ij}^{+}$ differs at most from $\mathbf{X}$ by the (i, j)th entry, which is forced to be 1. Define $\mathbf{X}_{ij}^{-}$ as the array formed from $\mathbf{X}$ where the tie from i to j is forced to be absent:

$$\begin{aligned} (X_{ij}^{-})_{mn} &= X_{ij}, \text{ if } (m, n) \neq (i, j) \\ &= 0, \text{ if } (m, n) = (i, j). \end{aligned}$$

Last, define $\mathbf{X}_{ij}^{c}$ as the matrix for the *complement* relation for $\mathbf{X}$ of the tie from i to j:

$$\begin{aligned} (X_{ij}^{c})_{mn} &= X_{ij}, \text{ if } (m, n) \neq (i, j) \\ &= \text{ undefined, if } (m, n) = (i, j). \end{aligned}$$

The complement relation has no relational tie coded from i to j – one can view this single variable as missing. These arrays are used when estimating the parameters of p^* via an approximate, maximum pseudolikelihood estimation.

8.2 Dependence Graphs

Recall that we have defined a set of random variables based on the relational ties in the network. The first step for any probabilistic model of a network is to consider the statistical dependencies among the elements of this set. To do this, we construct a dependence graph. Such a device allows us to distinguish among the many possible graph probability distributions, which can often be characterized by considering which relational ties are assumed to be statistically independent.

We define a dependence graph (as it applies to network relational variables) and then show how it can distinguish among basic graph distributions (such as those described in Wasserman and Faust 1994, Chapter 13). This dependence graph is also the starting point for the Hammersley-Clifford theorem (Besag 1974), which posits a very general probability distribution for these network random variables using the postulated dependence graph. The exact form of the dependence graph depends on the nature of the substantive hypotheses about the social network under study; we briefly discuss several such hypotheses.

(A) *Theory*

Any observed single relational network may be regarded as a realization $\mathbf{x} = [x_{ij}]$ of a random two-way binary array $\mathbf{X} = [X_{ij}]$. In general, the entries of the array $\mathbf{X}$ cannot be assumed to be independent; consequently, it is helpful to specify a dependence structure for the random variables $\{X_{ij}\}$, as originally suggested by Frank and Strauss (1986).

The dependence structure for these random variables is determined by the *dependence graph* $\mathcal{D}$ of the random array $\mathbf{X}$. $\mathcal{D}$ is itself a graph whose nodes are elements of the index set $\{(i, j); i, j \in \mathcal{N}, i \neq j\}$ for the random variables in $\mathbf{X}$, and whose edges signify pairs of the random variables that are assumed to be conditionally dependent (given the values of all other random variables).

More formally, a dependence graph for a univariate social network has node set

$$\mathcal{N}_D = \{(i, j); i, j \in \mathcal{N}, i \neq j\}.$$

The edges of $\mathcal{D}$ are given by $\mathcal{E}_D = \{((i, j), (k, l)),$ where X_{ij} and X_{kl} are not conditionally independent$\}$. This specific dependence graph is a version of an *independence* graph, as it is termed in the graphical modeling literature (for example, Lauritzen 1996; Robins 1997); see Robins (1998) for an extended discussion of the application of graphical modeling techniques to social network models.

(B) *Applications*

As Frank and Strauss (1986) observed for univariate graphs and associated two-way binary arrays, several well-known classes of distributions for random graphs may be

specified in terms of the structure of the dependence graph. Pattison and Wasserman (2001) and Wasserman and Pattison (2000) noted that there are three major classes – Bernoulli graphs and conditional uniform graph distributions, dyadic dependence distributions, and p^*. Other probabilistic graph models are described by Bollobas (1985), although a primary issue in the mathematics literature is on asymptotic behavior of various graph statistics as the size of the node set increases (whereas typically in social network analysis we want to analyze social networks on a fixed-node set). We briefly focus on the first two classes mentioned previously, describing the third at length later in this chapter.

The assumption of conditional independence for all pairs of random variables representing distinct relational ties (that is, X_{ij} and X_{kl} are independent whenever $i \neq k$ and $j \neq l$) leads to the class of Bernoulli graphs (see Frank and Nowicki 1993). The dependence graph for such a distribution has no edges; it is empty. A Bernoulli graph assumes complete independence of relational ties; the probability that the tie $i \rightarrow j$ is present is P_{ij}. If the $P_{ij} = 0.5$ for all ties, the distribution is often referred to as the uniform random (di)graph distribution, U. All (di)graphs are equally likely to occur; hence, the uniform probability aspect of the distribution. A more general Bernoulli graph distribution fixes the P_{ij} at P; each edge can be viewed as the outcome of a biased coin toss, with probability P of a "success."

The uniform distribution U conditions on no graph properties, whereas the uniform distribution $U|L$ statistically conditions on the number L of edges in the graph. All (di)graphs with $L = l$ lines (arcs) are equally likely; (di)graphs with $L \neq l$ lines (arcs) have probability 0. There are many other conditional uniform distributions, including the classic $U|MAN$ distribution, which fixes the counts of the dyad states and assumes that all digraphs with the specified dyad census are equally likely, and $U|\{X_{i+}\}, \{X_{+j}\}$, which fixes the out-degrees and in-degrees. Many such conditional uniform distributions are described in Chapter 13 of Wasserman and Faust (1994). Some of these distributions have simple dependence graphs; for example, the $U|\{X_{i+}\}$ distribution, which fixes only the out-degrees, has a dependence graph with edge set $\mathcal{E}_D = \{((i, j), (i, k))$, for all $j \neq k$ for every $i\}$.

The assumption of conditional dependence of X_{ij} and X_{kl}, if and only if $\{k, l\} = \{j, i\}$, leads to the class of dyad dependence models (see Wasserman 1987; Holland & Leinhardt 1981), the second family of graph distributions mentioned previously. These "multinomial dyad" distributions assume all dyads are statistically independent and postulate substantively interesting parameterizations for the probabilities of the various dyad states. The dependence graph for such distributions has an edge set with edges connecting only the two random variables within each dyad: $\mathcal{E}_D = \{((i, j), (j, i))$, for all $i \neq j\}$. This class of models was termed p_1 by Holland and Leinhardt (1977, 1981) and has a long history (see Chapters 15 and 16 of Wasserman and Faust 1994). Although for some parameterizations it is easy to fit, its assumption of independence across dyads is not terribly realistic.

Consider now a general dependence graph, with an arbitrary edge set. Such a dependence graph yields a very general probability distribution for a (di)graph, which we term p^* and focus on as follows. One dependence graph, for which this distribution was first developed, assumes conditional independence of X_{ij} and X_{kl}, if and only if $\{i, j\} \cap \{k, l\} = \emptyset$. This dependence graph links any two relational ties

involving the same actor(s); thus, any two relational ties are associated if they involve the same actor(s). This type of dependency resembles a Markov spatial process, so these dependencies were defined as a Markov graph by Frank and Strauss (1986). This p^* family of distributions has been extended in many ways, and estimates of its parameters scrutinized.

Of course, if the dependence graph is fully connected, then a general class of random graphs is obtained. We note, however, that any model deriving from a fully connected dependence graph is not identifiable. Later chapters introduce more complex dependence structures that permit models more general than Markov random graphs, but avoid fully connected dependence graphs.

8.3 *p**

For an observed network, which we consider to be a realization $\mathbf{x}$ of a random array $\mathbf{X}$, we assume the existence of a dependence graph $\mathcal{D}$ for the random array $\mathbf{X}$. The edges of $\mathcal{D}$ are crucial here; consider the set of edges, and determine if there are any *complete subgraphs*, or cliques, found in the dependence graph. [For a general dependence graph, a subset A of the set of relational ties $\mathcal{N}_D$ is *complete* if every pair of nodes in A (that is, every pair of relational ties) is linked by an edge of $\mathcal{D}$. A subset comprising a single node is also regarded as complete.] These cliques specify which subsets of relational ties are all pairwise, conditionally dependent on each other.

The Hammersley-Clifford theorem (Besag 1974) establishes that a probability model for $\mathbf{X}$ depends only on the cliques of the dependence graph $\mathcal{D}$. In particular, application of the Hammersley-Clifford theorem yields a characterization of $Pr(\mathbf{X} = \mathbf{x})$ in the form of an exponential family of distributions:

$$Pr(\mathbf{X} = \mathbf{x}) = \left(\frac{1}{\kappa}\right) \exp\left(\sum_{A \subseteq \mathcal{N}_D} \lambda_A \prod_{(i,j) \in A} x_{ij}\right), \tag{8.1}$$

where:

- $\kappa = \sum_{\mathbf{x}} \exp\{\sum_{A \subseteq \mathcal{D}} \lambda_A \prod_{(i,j) \in A} x_{ij}\}$ is a normalizing quantity.
- $\mathcal{D}$ is the dependence graph for $\mathbf{X}$; the summation is over all subsets A of nodes of $\mathcal{D}$.
- $\prod_{(i,j) \in A} x_{ij}$ is the sufficient statistic corresponding to the parameter λ_A.
- $\lambda_A = 0$ whenever the subgraph induced by the nodes in A is not a clique of $\mathcal{D}$.

The set of nonzero parameters in this probability distribution for $Pr(\mathbf{X} = \mathbf{x})$ depends on the *maximal* cliques of the dependence graph (a maximal clique is a complete subgraph that is not contained in any other complete subgraph). Any subgraph of a complete subgraph is also complete (but not maximal), so if A is a maximal clique of $\mathcal{D}$, then the probability distribution for the (di)graph will contain nonzero parameters for A and all its subgraphs.

Clearly, the number of parameters can be overwhelming. Thus, it is wise to limit these numbers by either postulating a simple dependence graph or making assumptions

about the parameters. Our usual assumption is *homogeneity*, in which parameters for isomorphic *configurations* of nodes are equated. As defined by Pattison and Wasserman (1999), this assumption equates parameters for the various isomorphic subgraphs that arise. For example, there can be two parameters for the isomorphic dyads (null, asymmetric, and mutual), up to fifteen parameters for the isomorphic triad states, and so forth (there is a loss of one parameter for each class of subgraph due to redundancies). See Figure 10.3.2 in Robins and Pattison (Chapter 10, this volume). Detailed definitions and consequences of homogeneity can be found in Pattison and Wasserman (1999). It is also possible to equate parameters by relying on the common practice of assuming *a priori* stochastic blockmodels (Anderson, Wasserman, and Faust 1992).

Even with homogeneity imposed, models may not be identifiable. Typically, parameters for higher-order configurations (for example, higher-order stars or triads) are set to zero (equivalent to setting higher-order interactions to zero in general linear models). An interpretation of the resulting model in terms of constrained social settings is given in later chapters.

Tables of parameters, and associated minimal sufficient statistics, can be found in the trilogy of p^* papers: Wasserman and Pattison (1996), Pattison and Wasserman (1999), and Robins et al. (1999).

The interpretation of parameters can be complicated given that higher-order configurations contain within them lower-order configurations (see Robins et al. 1999). At its simplest, a substantial positive parameter estimate for a triangle (for instance) suggests that, given the number of other configurations in the observed graph, there are more triangles present than would be expected by chance.

(A) *Estimation*

The probability distribution arising from the Hammersley-Clifford theorem, (8.1), can be written in the general form:

$$Pr(\mathbf{X} = \mathbf{x}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}}{\kappa(\boldsymbol{\theta})} \tag{8.2}$$

where $\boldsymbol{\theta}$ is a vector of model parameters and $\mathbf{z}(\mathbf{x})$ is a vector of network statistics. $\kappa(\boldsymbol{\theta})$ is a normalizing constant, which guarantees that the distribution is proper. Which network statistics appear in the model depends on the structure of the hypothesized dependence graph, and on whether any homogeneity constraints have been proposed. One could view this model as an autologistic regression, as described by Wasserman and Pattison (1996). Regardless of the motivation, the elements of $\boldsymbol{\theta}$ are unknown and must be estimated.

The likelihood function for the distribution is quite simple:

$$L(\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}}{\kappa(\boldsymbol{\theta})}.$$

Even though it has a simple expression, the function is not easy to work with, due to the dependence of $\kappa(\boldsymbol{\theta})$ on the unknown parameters. Direct, exact differentiation of the

log likelihood is difficult, if not impossible. Some exact results can be found in Walker (1995), but are specific to very small (di)graphs.

Two approaches have been used to date: (1) a maximum pseudolikelihood estimation technique, pioneered by Besag (1975, 1977a, 1977b) and refined and applied by Strauss (1986) and Strauss and Ikeda (1990); and (2) Markov chain Monte Carlo maximum likelihood estimation, being applied to p^* by Crouch and Wasserman (1998), Snijders (2002), and Handcock (2003).

(B) *Pseudolikelihood Estimation*

For probabilities to be computed, one must be able to calculate κ, which is just too difficult for most networks. Hence, alternative model formulations and approximate estimation techniques are important. One such alternative, which we now describe, uses log odds ratios of the conditional probabilities of each element of $\mathbf{X}$.

The Logit Model

We can turn model (8.2) into a autologistic regression model, not for the probability of the (di)graph, but for the conditional probabilities of the relational ties. This produces an approximate likelihood function that is much easier to deal with than the likelihood function described previously.

We first condition on the complement of X_{ij}, defined earlier in this chapter, and consider just the probability that the dichotomous random variable X_{ij} is unity. Specifically, consider

$$\begin{aligned} Pr(X_{ij} = 1|\mathbf{X}^c_{ij}) &= \frac{Pr(\mathbf{X} = \mathbf{x}^+_{ij})}{Pr(\mathbf{X} = \mathbf{x}^+_{ij}) + Pr(\mathbf{X} = \mathbf{x}^-_{ij})} \\ &= \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x}^+_{ij})\}}{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x}^+_{ij})\} + \exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x}^-_{ij})\}}. \end{aligned} \tag{8.3}$$

The odds ratio, which simplifies model (8.3), is

$$\begin{aligned} \frac{Pr(X_{ij} = 1|\mathbf{X}^c_{ij})}{Pr(X_{ij} = 0|\mathbf{X}^c_{ij})} &= \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x}^+_{ij})\}}{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x}^-_{ij})\}} \\ &= \exp\{\boldsymbol{\theta}'[\mathbf{z}(\mathbf{x}^+_{ij}) - \mathbf{z}(\mathbf{x}^-_{ij})]\} \end{aligned} \tag{8.4}$$

yielding the simple "linear" model

$$\varpi_{ijm} = \log\left\{\frac{Pr(X_{ij} = 1|\mathbf{X}^c_{ij})}{Pr(X_{ij} = 0|\mathbf{X}^c_{ij})}\right\} = \boldsymbol{\theta}'[\mathbf{z}(\mathbf{x}^+_{ij}) - \mathbf{z}(\mathbf{x}^-_{ij})]. \tag{8.5}$$

If we define $\boldsymbol{\delta}(x_{ij}) = [\mathbf{z}(\mathbf{x}^+_{ij}) - \mathbf{z}(\mathbf{x}^-_{ij})]$, then the logit model (8.5) simplifies succinctly to $\varpi_{ij} = \boldsymbol{\theta}'\boldsymbol{\delta}\,(x_{ij})$.

The important quantities here are the elements of $\boldsymbol{\delta}\,(x_{ij})$, that is, the vector of network statistics whose elements measure the changes in the statistics when the relational tie x_{ij} changes from 1 to 0. These odds ratios form the basis of an approximate likelihood function, which is constructed by assuming they are conditionally independent.

Estimation

An approximate estimation approach, proposed by Besag (1975, 1977b), and adopted by Strauss (1986), Strauss and Ikeda (1990), and the trilogy of p^* papers, uses a *pseudolikelihood function*

$$PL(\boldsymbol{\theta}) = \prod_{i \neq j} Pr(X_{ij} = 1 | \mathbf{X}_{ij}^c)^{x_{ij}} Pr(X_{ij} = 0 | \mathbf{X}_{ij}^c)^{1-x_{ij}}. \tag{8.6}$$

A *maximum pseudolikelihood estimator* (MPLE) is the value of $\boldsymbol{\theta}$ that maximizes (8.6). This approach assumes conditional independence of the random variables representing the relational ties. Details on the literature on approximate likelihood estimation can be found in Pattison and Wasserman (1999). Estimation of $\boldsymbol{\theta}$ for single, dichotomous relations can be accomplished via logistic regression. Maximizing the pseudolikelihood given in 8.6 is equivalent to maximizing the likelihood function for the fit of logistic regression to the model (8.5) (for independent observations $\{x_{ij}\}$).

In practice, pseudolikelihood estimation is not overly complex because standard logistic regression techniques can be used. To set up the data file for pseudolikelihood estimation, each possible binary tie (X_{ij}) becomes a "case," with the "independent variables" constituted by the parameters in the model (the isomorphic configurations – for instance, stars and triads of various types). For each case, the statistic associated with a variable is the difference in the number of the relevant configurations between the graph with $x_{ij} = 1$ and the graph with $x_{ij} = 0$. Standard logistic regression can then be applied to this data file, with the observations on the ties as the dependent variable. Nevertheless, this is not a standard logistic regression because of the dependencies within the data, and the usual tests of model fit do not strictly apply. For instance, the pseudolikelihood deviance (which a standard logistic regression package will normally compute) is not necessarily an asymptotic chi-square random variable. Accordingly, measures of fit are usually taken as heuristic guides. The pseudolikelihood deviance is often presented, along with simple goodness-of-fit statistics, such as mean absolute residual.

(C) *Maximum Likelihood Estimation: Simulating p* Models and Model Degeneracy*

Maximum pseudolikelihood estimation was suggested by Besag (1975) for dealing with data in spatial statistics. It has the virtue of being relatively easy to fit, even for complicated models. Moreover, the maximum pseudolikelihood estimator satisfies unbiased estimating equations, is consistent, and is asymptotically normal under suitable conditions (Baddeley and Turner 2000; see also Geys, Molenberghs, and Ryan 1997, 1999; Le Cessie and van Houwielingen 1994). Nevertheless, because the properties of the pseudolikelihood estimator are not well-understood, a more recent body of work has developed Monte Carlo techniques for maximum likelihood estimates for p^* models.

Various approaches to Monte Carlo estimation are available. For the particular case of p^* models, the central idea is to simulate a distribution of random graphs from a starting set of parameter values, and then to refine these estimated parameter values by comparing the distribution of graphs with the observed graph. The process is repeated until the parameter estimates stabilize. Simulation procedures establish a Markov chain

of graphs that, under suitable conditions, will converge to a stationary graph distribution. Two of the most popular algorithms that can produce such a Markov chain are the Gibbs sampler (Geman and Geman 1984) and the Metropolis-Hastings algorithm (which includes the Gibbs sampler as a special case; see Chib and Greenberg 1995). Geyer and Thompson (1992) presented a general Markov chain Monte Carlo maximum likelihood method (see also Besag 2000). Crouch and Wasserman (1998) described how this technique could be applied specifically to *p** models (see also Snijders 2002, and Corander, Dahmström, and Dahmström 2002).

Studies on Monte Carlo techniques have thrown new light on an important issue, that of model degeneracy. For certain parameter values, a *p** model may produce a distribution of graphs in which only a handful of graphs (sometimes only one) have nonzero probability; moreover, these graphs are often uninteresting, such as the full or empty graph. Such a model is termed *degenerate*. For degenerate models, estimation techniques will not perform well, irrespective of method.

Strauss (1986) was the first to use the Metropolis-Hastings algorithm to simulate Markov random graph distributions. Strauss observed a variation on the problem of degeneracy, by noting that asymptotically there may be no finite normalizing constant for a distribution with certain parameter values. For these regions of the parameter space, simulations are thus not adequate in producing a stationary distribution.

Based on the results in Besag (2000) and Handcock (2000), Hoff, Raftery, and Handcock (2002) suggested that commonly used *p** models may have model degeneracy and instability problems that are not resolved by alternative forms of estimation, but rather may represent defects in the models themselves. This is a strong judgment that is not necessarily borne out in subsequent work. However, a similar point is also implied by Snijders (2002), who presented a Monte Carlo estimation technique for *p** models using the Gibbs sampler and the Munro-Robbins algorithm to estimate the moments of the sufficient statistics. Snijders presented several examples of simple Markov graph models that involved degeneracy and instability. For certain parameter values, sufficient statistics had bimodal distributions. Snijders' suggestions to overcome these problems worked satisfactorily for some data sets, especially for small graphs, but the bimodality remained a limitation. Snijders and van Duijn (2002) suggested that to address bimodality, model estimation might be conditional on the number of observed edges in the graph.

The most extensive study on degeneracy for these models has been conducted by Handcock (2003). Handcock defined "near degeneracy" of a graph model as occurring when the model places disproportionate probability mass on only a few of the possible graphs (often empty or full graphs). He examined in detail the simplest Markov model for very small nondirected graphs – the two-star model with only edge and two-star parameters. He delineated the region of parameter space in which nondegeneracy occurs, where Monte Carlo estimation techniques will operate satisfactorily and the models are statistically well-behaved. This work also makes clear that the bimodality observed by Snijders (2002) is to be expected for certain parameter values that are on the edge of the nondegenerate area.

Simulation studies by Robins, Pattison, and Woolcock (in press) suggested that nondegenerate graphs may be more readily achieved with more complex models, in

particular with a parameterization that includes three and four stars, that is, at least the first three moments of the degree distribution. Triangle, transitivity parameters are also desirable in a "realistic" model. Robins et al. showed that by varying values for such a parameter set, it is possible to simulate Markov graph distributions with properties akin to "small worlds" (Watts 1999), or with other global features such as long paths or a high proportion of four cycles.

Following Grenander (1993), Robins et al. (in press) demonstrated that increasing all parameter values by the same factor results in movement toward degenerate regions. Contrary to Handcock (2003), who argued that degenerate regions imply uninteresting models, Robins et al. noted that some degenerate models result in graphs of theoretical import, such as graphs of disconnected complete components (the so-called caveman graphs of Watts 1999) or complete bipartite graphs. Robins et al. interpreted degenerate regions as areas where "stochasticism" breaks down and deterministic structures emerge (see also Pattison 2002). Their simulation technique permits examination of how "close" an observed graph is to this phase transition. Most human social structures are indeed stochastic, but because of tendencies toward transitivity and structural balance, it is possible that stochastic social systems may be not too "far" from determinism (Robins 2003). Such a conclusion would accord with the small world simulation results of Watts (1999), who showed that the addition of only a small random component to a highly structured graph resulted in small world properties.

In summary, Monte Carlo estimation techniques for these models are now well-established and their development has shed new light on model behavior, particularly on model degeneracy. It is important for any estimation procedure that the model be non-degenerate; otherwise, there will not be satisfactory convergence of parameter estimates. Programs for Monte Carlo estimation are now available or becoming available; one example being the estimation procedures in the StOCNET suite of network programs (Snijders 2002). This program estimates Markov random graph models for directed and nondirected graphs with a choice of parameterizations, including reciprocity, triadic, and higher-order star parameters. Output from the program includes an assessment of convergence and reliable standard errors of parameter estimates.

(D) *Comparing Pseudolikelihood and Monte Carlo Maximum Likelihood Estimation*

Maximum likelihood estimation is undoubtedly optimal in the sense of having a principled statistical basis and producing reliable standard errors from which statistical inferences can be made. These are not qualities of pseudolikelihood estimation. However, Monte Carlo approaches to maximum likelihood estimation can be computer intensive, so estimation for networks with a large number of nodes, or for a complex model, may not be possible or may take an unacceptably long time. One very important question is "in what, if any, circumstances might pseudolikelihood be an acceptably approximate technique to obtain parameter estimates?"

There have been some interesting studies comparing the methods of estimation. Corander, Dahmström, and Dahmström (1998) used a Metropolis-Hastings algorithm, together with formulae for the first three cumulants (moments) of the sufficient statistics,

to obtain maximum likelihood estimates for simple Markov random graph models (see also Corander et al. 2002). Comparing these estimates with pseudolikelihood estimates, they concluded that for graphs of up to approximately 40 nodes, the maximum likelihood estimator performed better in the nondegenerate regions of the parameter space. The pseudolikelihood estimator was more biased. In larger graphs, of 40 to 100 nodes, they concluded that the two estimators were nearly equivalent, although they showed that pseudolikelihood estimates could vary for different graphs with the same values of the sufficient statistics. Their simulation approach, however, fixed the number of edges in the graph, so it is not clear how to interpret their comparison of the two estimators because the pseudolikelihood estimation presumably included an edge parameter.

In other comparative studies, using models from Wasserman and Pattison (1996), Besag (2000) showed that a Monte Carlo goodness-of-fit test can lead to different conclusions from those that would apply if fits based on maximum pseudolikelihood estimates were treated as distributed as chi-square random variables. These results emphasize that standard hypothesis tests are simply not appropriate for pseudolikelihood estimates. Snijders (2002) presented examples where pseudolikelihood estimates were close to maximum likelihood estimates, as well as other examples where they were not.

One of the problems with assessing the performance of the maximum pseudolikelihood estimator is that of model specification. The simulation results of Robins, Pattison, and Woolcock (in press), noted previously, suggest that more complex models may have better properties in terms of avoiding degeneracy and instability. It might be expected that the maximum pseudolikelihood estimator performs less well for models that are close to degeneracy so the complexity of the models may affect the performance of the estimator. These results stress the potential importance of specification.

In summary, pseudolikelihood estimates have convenience, but are at best approximate. Whenever possible, Monte Carlo maximum likelihood approaches should be preferred. If pseudolikelihoods are to be used, estimates should be treated as exploratory, giving some possible evidence for effects that may be substantial in the network, with a clear recognition that point estimates are probably not precise. (For many research purposes, this may be sufficient.) Formal statistical inferences should not be made with pseudolikelihood estimation. Standard errors and other statistical results from the logistic regression procedure used for pseudolikelihood estimation should be treated with caution. Standard errors are likely to be too small, and at best should be viewed as very rough indicators of "scale" for the parameters. Wald statistics should not be regarded as reliable. The pseudolikelihood deviance remains a valid measure of model fit, in that models that better predict the data will have lower deviance, but it will not be distributed, even asymptotically, as chi-square; thus, statistical inference is not available. In terms of model specification, our current suggestion is that Markov random graph models should carry rather complicated parameterization, including at least a three-star parameter (three-in-star and three-out-star for directed graphs). If the highest-order star parameter has a positive estimate, beware of the possibility of degeneracy. In that case, even higher-order star parameters may be necessary.

In conclusion, an important focus for future work should be the specification of regions of degeneracy relative to model specification. Such a research approach may

permit a better understanding of when pseudolikelihood procedures can be considered adequate. The outcome may be a clearer idea of the type of parameters required for sensible *p** model formulation. The notion that a simple Markov parameterization sufficiently models social networks is also being taken up through more substantive discussion of social settings and of construals of social space in general (see Robins and Pattison, Chapter 10, this volume; also Hoff et al. 2002, and Pattison and Robins 2002). This may be an instance where substantive considerations on the one hand and technical estimation requirements on the other hand could jointly lead to better model formulation.

Acknowledgments

This research was supported by grants from the National Science Foundation and the U.S. Office of Naval Research, and the Australian Research Council.

References

Anderson, C. J., Wasserman, S., and Crouch, B. (1999). A *p** primer: Logit models for social networks. *Social Networks*, *21*, 37–66.

Anderson, C. J., Wasserman, S., and Faust, K. (1992). Building stochastic blockmodels. *Social Networks*, *14*, 137–161.

Baddeley, A., and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point processes (with discussion). *Australian and New Zealand Journal of Statistics*, *42*, 283–322.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, *36*, 192–236.

Besag, J. E. (1975). Statistical analysis of non-lattice data. *The Statistician*, *24*, 179–195.

Besag, J. E. (1977a). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Association*, *47*, 77–92.

Besag, J. E. (1977b). Efficiency of pseudo-likelihood estimation for simple Gaussian random fields. *Biometrika*, *64*, 616–618.

Besag, J. E. (2000). *Markov Chain Monte Carlo for Statistical Inference*. Working Paper, University of Washington, Center for Statistics and the Social Sciences, Seattle.

Bollobas, B. (1985). *Random Graphs*. London: Academic Press.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, *49*, 327–335.

Corander, J., Dahmström, K., and Dahmström, P. (1998). *Maximum Likelihood Estimation for Markov Graphs*. Research Report 1998:8, University of Stockholm, Department of Statistics, Stockholm, Sweden.

Corander, J., Dahmström, K., and Dahmström, P. (2002). Maximum likelihood estimation for exponential random graph models. In J. Hagberg (ed.), *Contributions to Social Network Analysis, Information Theory and Other Topics in Statistics: A Festschrift in Honour of Ove Frank on the Occasion of His 65th Birthday* (pp. 1–17). Stockholm, Sweden: Universitet Stockholms.

Crouch, B., and Wasserman, S. (1998). *Fitting p*: Monte Carlo Maximum Likelihood Estimation*. Paper presented at International Conference on Social Networks, Sitges, Spain, May 28–31.

Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica*, *45*, 283–293.

Frank, O. (1997). Composition and structure of social networks. *Mathematiques, Informatique, et Science Humaines*, *137*, 11–23.

Frank, O., and Nowicki, K. (1993). Exploratory statistical analysis of networks. In J. Gimbel, J. W. Kennedy, and L. V. Quintas (eds.), *Quo Vadis Graph Theory? Annals of Discrete Mathematics, 55*, 349–366.

Frank, O., and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*, 832–842.

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

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

Geys, H., Molenberghs, G., and Ryan, L. M. (1997). Pseudolikelihood inference for clustered binary data. *Communications in Statistics, Part A – Theory and Methods*, *26*, 2743–2767.

Geys, H., Molenberghs, G., and Ryan, L. M. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, *94*, 734–745.

Grenander, U. (1993). *General Pattern Theory*, Oxford: Oxford University Press.

Handcock, M. S. (2000). *Progress in Statistical Modeling of Drug User and Sexual Networks*. Manuscript, University of Washington, Center for Statistics and the Social Sciences, Seattle.

Handcock, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, and P. Pattison (eds.), *Dynamic Social Network Modeling and Analysis* (pp. 229–240). Washington, DC: National Academies Press.

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

Holland, P. W., and Leinhardt, S. (1977). *Notes on the Statistical Analysis of Social Network Data*. Unpublished manuscript.

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

Lauritzen, S. (1996). *Graphical Models*. Oxford: Oxford University Press.

Le Cessie, S., and van Houwielingen, J. C. (1994). Logistic regression for correlated binary data. *Applied Statistics*, *43*, 95–108.

Pattison, P. E. (2002). *Social networks, social space, social structure*. Keynote address, Sunbelt XXII International Social Network Conference, New Orleans, February.

Pattison, P. E., and Robins, G. L. (2002). Neighbourhood-based models for social networks. *Sociological Methodology*, *32*, 301–337.

Pattison, P., and Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, *52*, 169–193.

Pattison, P., and Wasserman, S. (2001). Social network models, statistical. In Smelser, N. J., and Baltes, P. B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences*. London: Elsevier Science.

Rennolls, K. (1995). $p_{\frac{1}{2}}$. In Everett, M. G., and Rennolls, K. (eds.), *Proceedings of the 1995 International Conference on Social Networks. 1* (pp. 151–160). London: Greenwich University Press.

Robins, G. L. (1997). Graphical modelling. *Chance*, *10*, 37–40.

Robins, G. L. (1998). *Personal Attributes in Inter-personal Contexts: Statistical Models for Individual Characteristics and Social Relationships*. Unpublished doctoral dissertation, University of Melbourne, Department of Psychology, Melbourne, Australia.

Robins, G. L. (2003). *The Small Worlds of Small Social Networks*. Paper presented at the American Association for the Advancement of Science annual meeting, Denver, Colorado, February 13–18.

Robins, G. L., Elliot, P., and Pattison, P. (2001). Network models for social selection processes. *Social Networks*, *23*, 1–30.

Robins, G. L., Pattison, P., and Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, *66*, 161–190.

Robins, G. L., Pattison, P., and Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika*, *64*, 371–394.

Robins, G. L., Pattison, P., and Woolcock, J. (in press). Small and other worlds: Global network structures from local processes. *American Journal of Sociology*.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*, 2.

Snijders, T. A. B., and van Duijn, M. A. J. (2002). Conditional maximum likelihood estimation under various specifications of exponential random graph models. In J. Hagberg (ed.), *Contributions to Social Network Analysis, Information Theory and Other Topics in Statistics: A Festschrift in Honour of Ove Frank on the Occasion of His 65th Birthday* (pp. 117–134). Stockholm, Sweden: Universitet Stockholms.

Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*, *28*, 513–527.

Strauss, D. (1992). The many faces of logistic regression. *American Statistician*, *46*, 321–327.

Strauss, D., and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*, 204–212.

Walker, M. E. (1995). *Statistical Models for Social Support Networks: Application of Exponential Models to Undirected Graphs with Dyadic Dependencies*. Unpublished doctoral dissertation, University of Illinois, Department of Psychology, Chicago.

Wasserman, S. (1987). Conformity of two sociometric relations. *Psychometrika*, *52*, 3–18.

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Wasserman, S., and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov random graphs and *p**. *Psychometrika*, *60*, 401–426.

Wasserman, S., and Pattison, P. (2000). Statistical models for social networks. In Kiers, H., Rasson, J., Groenen, P., and Schader, M. (eds.), *Studies in Classification, Data Analysis, and Knowledge Organization*. Heidelberg: Springer-Verlag.

Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press.

9

Random Graph Models for Social Networks: Multiple Relations or Multiple Raters[1]

Laura M. Koehly
Texas A&M University

Philippa Pattison
University of Melbourne

9.1 Introduction

Several chapters in this book outline some of the significant advances that have been made in modeling networks and network-based processes (see, for example, Chapters 6, 7, 10, and 11). These models generally presuppose a single network of interest, such as a network of acquaintance ties or a network of advice-seeking ties, and they represent the interdependence of such ties with actor characteristics and other ties in some local network *neighborhood* (see, for example, Chapter 10). Yet, there are compelling theoretical and methodological reasons to extend these models to the case of multiple networks, and in this chapter we discuss the rationale and nature of these extensions, as well as a number of issues to which they give rise.

From a theoretical perspective, it is more than likely that network processes involve different kinds of relational ties; indeed, some well-known hypotheses about the nature of local network processes involve multiple types of tie. Cartwright and Harary's (1956) adaptation of Heider's (1946) balance model, for example, proposes a strong form of interdependence among positive and negative ties within triadic network structures, and Granovetter's (1973) "strength of weak ties" thesis involves an interdependence between strong, weak, and null ties. In addition, there is an impressive body of empirical work that points to the importance of multiplex ties, that is, those ties in which several types of relationships come together – such as friend and coworker, or advisor and supervisor – and also to the consequences of such ties for interpersonal processes. Indeed, theoretical arguments about such forms of interdependence are supported by a number of empirical studies. For example, Lazega and Pattison (1999) identified a number of separable forms of interdependence among three different types of ties – coworker, advisor, and friend – linking members of a law firm. Their results suggested the simultaneous presence not only of strong multiplexity and generalized reciprocity effects, but also of more complex triadic forms of interdependence involving several types of ties.

From a methodological perspective, there are also compelling reasons to develop models for multiple relational observations among network members. In insightful