

Bernoulli Graph Bounds for General Random Graphs*

Carter T. Butts[†]

07/14/10

Abstract

General random graphs (i.e., stochastic models for networks incorporating heterogeneity and/or dependence among edges) are increasingly widely used in the study of social and other networks, but few techniques other than simulation have been available for studying their behavior. Random graphs with independent edges (i.e., the Bernoulli graphs), on the other hand, are well-studied, and a large literature exists regarding their properties. In this paper, we demonstrate a method for leveraging this knowledge by constructing families of Bernoulli graphs that bound the behavior of an arbitrary random graph in a well-defined sense. By studying the behavior of these Bernoulli graph bounds, one can thus constrain the properties of a given random graph. We illustrate the utility of this approach via application to several problems from the social network literature, including identifying degeneracy in Markov graph models, studying the potential impact of tie formation mechanisms on epidemic potential in sexual contact networks, and robustness testing of inhomogeneous Bernoulli models based on geographical covariates.

Keywords: random graphs, discrete exponential families, Bernoulli graphs, graph theory, social networks

1 Introduction

Explaining the structure of social networks has long been recognized as a core sociological problem (see, e.g., the early intuitions of Simmel (1897) and later refinements by Coleman (1964) and Mayhew (1980)), but the mathematical, computational, and empirical advances required for substantial progress in this area have coalesced only recently (Freeman, 2004). Central to these developments has been the study of general random graphs (i.e., stochastic models for networks incorporating heterogeneity and/or dependence among edges), without which it is difficult or impossible to account for the structural properties observed within individual, organizational, and other networks (see Holland and Leinhardt (1981); Frank and Strauss (1986); Wasserman and Pattison (1996) for some classic motivating arguments). Although great strides have been made in recent decades regarding the modeling of general random graphs, analytical tools for studying the behavior of these models have proved elusive. The special case of random graph models with independent edges – the Bernoulli graphs – is much better understood, particularly under conditions of homogeneity.¹ Even where analytical results are unavailable, the comparative simplicity of simulating Bernoulli graphs makes them much easier to study than their general counterparts, particularly for graphs with large vertex sets. In this paper, we provide a mechanism for leveraging

*This research was supported in part by ONR award N00014-08-1-1015 and NSF award BCS-0827027.

[†]Department of Sociology and Institute for Mathematical Behavioral Sciences; University of California, Irvine; SSPA 2145; Irvine, CA 92697-5100; buttsc@uci.edu

¹See, e.g., Wasserman and Robins (2005) for a recent overview of general random graph models, and Bollobás (2001) for a canonical treatment of classical random (primarily Bernoulli) graph theory.

these advantages by constructing families of Bernoulli graphs that bound the behavior of general random graphs in a well-specified manner. Given a general random graph distribution specified in exponential family form, we specifically construct “upper” and “lower” Bernoulli graph distributions whose realizations are (when appropriately coupled to the general random graph) supergraphs and subgraphs of corresponding realizations from the target family. Using this result, we are able to bound the distributional properties of a wide range of graph statistics for the target distribution in terms of the corresponding properties on the upper and lower graphs, with the latter being easily constrained using known results from the theory of (Bernoulli) random graphs. To demonstrate the utility of our approach, we apply these results to three illustrative problems from the social network literature: identifying asymptotically pathological behavior in Markov graph models, studying the impact of social mechanisms on connectivity and epidemic potential in heterosexual partnership networks, and robustness testing of covariate-based, inhomogeneous Bernoulli models for large-scale social networks.

1.1 Initial Concepts and Notation

A *graph*, $G = (V, E)$ is a structure consisting of a set of distinct elements (V , the *vertex set*), and a set of *edges* (E). The elements of E consist of ordered or unordered pairs of elements in V (*vertices*); graphs with ordered edges are said to be *directed*, and those with unordered edges are said to be *undirected*, the order of elements within an edge generally corresponding to distinct roles of “sender” and “receiver” in empirical settings. An edge consisting of pair (v, v) (or multiset pair $\{v, v\}$) for vertex v is called a *loop*, and may or may not be permissible – we do not disallow loops in the development which follows, but may do so (via support constraints) in specific applications. In general, we denote the *order* of a given graph (i.e., the cardinality of the vertex set) by $N = |V|$. We will here take N to be finite for any given G , although graph sequences of arbitrarily increasing order will be encountered in Section 3.

In practice, it is often useful to represent a graph via its *adjacency matrix*. The adjacency matrix for graph G is the $N \times N$ binary matrix Y such that $Y_{ij} = 1$ if (i, j) (directed case) or $\{i, j\}$ (undirected case) belong to E , and 0 otherwise. Y_{ij} can thus be thought of as an indicator variable for the presence of an i, j edge. If the elements of E are random, then Y is also a random matrix; the graph corresponding to such an entity is referred to as a *random graph*. It is often convenient to refer to the support of a random graph via the support of its adjacency matrix, and indeed we shall use that device in our exposition. Likewise, note that we can represent the restriction that G be undirected by restricting Y to the set of symmetric matrices, we can represent the restriction that G be loopless by requiring that the diagonal elements of Y be equal to zero, etc.

Another concept that will be employed in this paper is the *subgraph* relation. Specifically, $G = (V, E)$ is said to be a subgraph of $G' = (V', E')$ (denoted $G \subseteq G'$) iff $V \subseteq V'$ and $E \subseteq E'$. Intuitively, G is a subgraph of G' (equivalently, G' is a supergraph of G) if G can be constructed from G' by discarding some collection of edges and/or vertices. For graphs of identical order $G \subseteq G'$ clearly implies and is implied by $Y_{ij} \leq Y'_{ij}$ for all i, j , where Y, Y' are the respective adjacency matrices of G and G' . We will make use of this fact in deriving bounds for the behavior of random graphs. Also useful is the observation that some functions of graphs (or, equivalently, adjacency matrices) are monotonic in the subgraph operation (i.e., $G \subseteq G'$ implies that $f(G) \leq f(G')$ in the increasing case, or $f(G) \geq f(G')$ in the decreasing case). We will call such functions *subgraph monotone* in the development which follows.

1.2 Exponential Family Models for General Random Graphs

Let Y be a random adjacency matrix of fixed order N on support \mathcal{Y}_N . Without loss of generality, we may write the pmf of Y in exponential family form as

$$\Pr(Y = y|\theta, t, \mathcal{Y}_N, X) = \frac{\exp(\theta^T t(y, X))}{\sum_{y' \in \mathcal{Y}_N} \exp(\theta^T t(y', X))} \mathbb{I}_{\mathcal{Y}_N}(y), \quad (1)$$

with $t : \mathcal{Y}_N, X \mapsto \mathbb{R}^p$ being a vector of sufficient statistics, $\theta \in \mathbb{R}^p$ a real-valued vector of parameters, $\mathbb{I}_{\mathcal{Y}_N}$ an indicator for membership in \mathcal{Y}_N , and X an arbitrary set of covariates. Exploiting the one-to-one mapping between graphs and adjacency matrices, we refer to Equation 1 as specifying an exponential family random graph model (ERGM) for Y , with the random variable Y specified in this manner referred to as an exponential family random graph (ERG). Since any random graph distribution on \mathcal{Y}_N can be written in this fashion, Equation 1 provides a natural starting point for the treatment of general random graphs.

Although Equation 1 provides a “holistic” expression for the pmf of Y , we may also describe the pmf via its component parts (i.e., the individual elements of Y , each of which is a random variable on $\{0, 1\}$). Specifically, we may factor the joint pmf of Y as follows:

$$\Pr(Y = y|\theta, t, \mathcal{Y}_N, X) = \Pr(Y_{11} = y_{11}|\theta, t, \mathcal{Y}_N, X) \Pr(Y_{12} = y_{12}|\theta, t, \mathcal{Y}_N, X, Y_{11} = y_{11}) \dots \quad (2)$$

$$\times \Pr(Y_{NN} = y_{NN}|\theta, t, \mathcal{Y}_N, X, Y_{11} = y_{11}, \dots, Y_{N(N-1)} = y_{N(N-1)}) \\ = \prod_{i=1}^N \prod_{j=1}^N \Pr(Y_{ij} = y_{ij}|\theta, t, \mathcal{Y}_N, X, Y_{<ij} = y_{<ij}), \quad (3)$$

where we employ the shorthand notation $x_{<ij}$ to denote the sequence of lexically ordered matrix elements $x_{11}, x_{12}, \dots, x_{i(j-1)}$ (for $j > 1$) or $x_{11}, x_{12}, \dots, x_{(i-1)N}$ (otherwise). Thus, $\Pr(Y_{ij} = y_{ij}|\theta, t, \mathcal{Y}_N, X, Y_{<ij} = y_{<ij})$ is the pmf of the i, j th cell of Y , conditional on the preceding entries ($y_{<ij}$) and marginalized across the succeeding entries. We will refer to these distributions generically as the “partial conditionals” of Y , as opposed to the full conditional distributions discussed below. Note here that, depending on the support of Y (particularly whether the graph Y represents is directed or undirected, loopless or otherwise), some entries of Y_{ij} may be fixed. As this poses no particular threat to our analysis, our notation includes all $N \times N$ potential elements of Y .

Historically, the factorization of Equation 3 has not been regarded as being especially useful, since the partial conditional pmf of an arbitrary Y_{ij} is known for only a handful of ERG families (generally, those with full edgewise independence). By contrast, the *full* conditional pmf for an arbitrary Y_{ij} (where not trivially fixed by support constraints) can be easily derived from Equation 1. For compactness of notation, let x_{ij}^c denote the elements of the matrix x other than x_{ij} , and x_{ij}^+, x_{ij}^- denote the matrices with non- i, j entries set to their values in x , and i, j entries set to 1 or 0 (respectively). Let us further denote the Bernoulli pmf with parameter ϕ by $B(X = x|\phi)$. Then the pmf of non-fixed Y_{ij} given Y_{ij}^c is given by

$$\Pr(Y_{ij} = y_{ij}|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) = B\left(Y_{ij} = y_{ij} \left[1 + \exp\left[\theta^T \left(t(y_{ij}^-, X) - t(y_{ij}^+, X)\right)\right]\right]^{-1}\right). \quad (4)$$

In general, the conditional pmf of Y_{ij} given Y_{ij}^c is not equal to the marginal pmf of Y_{ij} (whether marginalized across a subset of matrix elements, or across the entire matrix), the major exception to this being the case in which the elements of Y_{ij} are independent. This special case is referred to as the Bernoulli graph family, and is of considerable importance; nevertheless, many ERGMs

of interest do not belong to the Bernoulli graphs, and their individual edge variables (matrix elements) cannot be trivially marginalized.² As a result, most current work on general ERGMs focuses on the use of the full conditional pmfs as tools for Markov chain Monte Carlo sampling (Snijders, 2002) and sampling-based inferential strategies (Geyer and Thompson, 1992; Hunter and Handcock, 2006). As we will show, however, there is another potentially useful role for the full conditionals in characterizing the behavior of ERGs: they can be used to place bounds on the partial conditional edge probabilities of Equation 3.

2 Bernoulli Graph Bounds for ERGs

Consider two discrete random variables, W and Z , with Z taking values on support \mathcal{Z} . By definition, $\Pr(W = w) = \sum_{z \in \mathcal{Z}} \Pr(W = w|Z = z) \Pr(Z = z)$. Since this is a convex combination, we may bound the marginal pmf of W by the extreme values of its corresponding conditional distributions, e.g. $\min_{z \in \mathcal{Z}} \Pr(W = w|Z = z) \leq \Pr(W = w) \leq \max_{z \in \mathcal{Z}} \Pr(W = w|Z = z)$. This is of particular utility in the case of binary variables, where it is natural to characterize the distribution of a given random variable in terms of the probability that it will take a value of 1 (versus 0). Applying this logic to the factored ERG representation of Equation 3, we observe that

$$\min_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y'_{ij}) \leq \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{<ij} = y_{<ij}) \quad (5)$$

$$\leq \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y'_{ij}), \quad (6)$$

with the full conditionals defined as per Equation 4. Note that, since they constitute extrema over all possible values of Y , the upper and lower bounds for each Y_{ij} do not depend upon $Y_{<ij}$ (or, indeed, any particular value of Y_{ij}^c). We may therefore treat the two collections of extrema as matrices in their own right, with $\Lambda, \Psi \in [0, 1]^{\{N \times N\}}$ such that $\Lambda_{ij} \leq \min_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y'_{ij})$ and $\Psi_{ij} \geq \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y'_{ij})$. For purposes that will become clear presently, it is likewise useful to assemble the i, j partial conditionals from Equation 3 themselves in matrix form; we refer to this matrix as Γ , with $\Gamma_{ij} \equiv \Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{<ij} = y_{<ij})$. Although it is important to remember that Γ is not a fixed matrix of real numbers (since each entry following Γ_{11} may depend on those before it), the i, j th element of Γ does not depend on any element that lexically succeeds it. We will exploit this property in the development that follows.

Considering for a moment the matrices Λ and Ψ , we observe that each provides a fixed set of lower and upper bounds (respectively) on the edge probabilities of the graph described by Y . Intuitively, we may imagine that if we were to construct Bernoulli graphs using Λ and Ψ as respective parameter matrices, these graphs would in some sense bound the behavior of Y . To make this intuition precise, let us define the subgraph operator, \subseteq , with $A \subseteq B$ for fixed graphs $A = (V, E), B = (V', E')$ iff $V \subseteq V'$ and $E \subseteq E'$. For graph adjacency matrices W, Z , with V fixed, this corresponds directly to the condition that $W_{ij} \leq Z_{ij}$ for all i, j ; we thus employ the notation $W \subseteq Z$ to mean that the graph represented by W is a subgraph of that represented by Z . Given three graphs $A \subseteq B \subseteq C$, A and C bound B in a very strong sense: every edge of A is also in B , and no edge is contained in B unless it is also in C . Many graph properties of interest (e.g., density, connectivity, degree) are monotone in the subgraph relation, in the sense that $A \subseteq B$ implies $f(A) \leq f(B)$ for property f (with the direction of inequality being reversed if the property

²The approximation of the factorization in Equation 3 by the product of full conditional densities is referred to as the (edgewise) *pseudolikelihood* approximation of the pmf of Y , and has historically played an important (albeit sometimes problematic) role in inference for ERGMs; see, e.g. Besag (1975); Strauss and Ikeda (1990); Wasserman and Robins (2005).

is monotone decreasing). In our present case, finding graphs A, B such that $A \subseteq Y \subseteq B$ would seem to give us considerable leverage in understanding the behavior of Y , in that Y 's subgraph monotone properties could be bounded by those of A and B . Because Y is not a fixed graph, such a trivial approach is not viable. However, we can obtain a closely analogous result by means of an appropriately defined stochastic process involving Λ , Γ , and Ψ .

Our construction proceeds as follows. Let R_{11}, \dots, R_{NN} be a sequence of iid uniform random variables on the $[0, 1]$ interval. Define the random matrix L as the $N \times N$ matrix such that $L_{ij} = 1$ if $R_{ij} < \Lambda_{ij}$ and 0 otherwise. Likewise, let U be the $N \times N$ matrix such that $U_{ij} = 1$ if $R_{ij} < \Psi_{ij}$ and 0 otherwise. Clearly, L and U are Bernoulli graphs with respective parameter matrices Λ and Ψ . Now, consider forming Y by setting $Y_{ij} = 1$ if $R_{ij} < \Gamma_{ij}$ and 0 otherwise. Since each element in Γ depends only on those before it (in the order of R), it follows that each Γ_{ij} depends only on known terms and on elements of R lexically prior to R_{ij} . Likewise, the factorization of Equation 3 implies that $Y_{ij}|Y_{<ij}$ is Bernoulli distributed with parameter Γ_{ij} . It therefore follows that the Y resulting from this process is distributed as per Equation 1 (i.e., that it has the desired ERG distribution). While these results hold for any R sequence as defined above, we may couple the three random graphs L , U , and Y directly by employing the same R for each. This construction leads immediately to the following simple theorem:

Theorem 1. *Let l , u , and y be realizations of L , U , and Y as defined above, for common realization vector r_{11}, \dots, r_{NN} of R . Then $l \subseteq y \subseteq u$.*

Proof. By construction, $\Lambda_{ij} \leq \Gamma_{ij} \leq \Psi_{ij}$, $l_{ij} = 1$ iff $r_{ij} < \Lambda_{ij}$, $y_{ij} = 1$ iff $r_{ij} < \Gamma_{ij}$, and $u_{ij} = 1$ iff $r_{ij} < \Psi_{ij}$ for all i, j (with 0 otherwise in each case). It follows then that $l_{ij} \leq y_{ij} \leq u_{ij}$ for all i, j , and hence $l \subseteq y \subseteq u$. \square

Theorem 1 characterizes a formal sense in which L and U bound the behavior of Y : when coupled by sharing the same part of randomness, the realizations of Y always lie between the corresponding realizations of L (the ‘‘lower’’ graph) and U (the ‘‘upper’’ graph) in the sense of the subgraph relation. An important corollary of this result is the following:

Corollary 1. *Let f be a graph statistic that is monotone in \subseteq . Then $\Pr(f(L) \leq x) \geq \Pr(f(Y) \leq x) \geq \Pr(f(U) \leq x)$ for all x if f is monotone increasing in \subseteq , with $\Pr(f(L) \leq x) \leq \Pr(f(Y) \leq x) \leq \Pr(f(U) \leq x)$ otherwise.*

Proof. By Theorem 1, $l \subseteq y \subseteq u$ for l, y, u arising from any realization of R . Let R_x^L , R_x^Y , and R_x^U be the sets of potential realizations of r such that (respectively) $f(L|r) \leq x$, $f(Y|r) \leq x$, and $f(U|r) \leq x$. For f monotone increasing in \subseteq , it follows that $R_x^U \subseteq R_x^Y \subseteq R_x^L$ for any x , and hence that $\Pr(f(L) \leq x) \geq \Pr(f(Y) \leq x) \geq \Pr(f(U) \leq x)$. By the same logic, if f is monotone decreasing in \subseteq then $R_x^L \subseteq R_x^Y \subseteq R_x^U$, and thus $\Pr(f(L) \leq x) \leq \Pr(f(Y) \leq x) \leq \Pr(f(U) \leq x)$. \square

It is a trivial consequence of the above that $\mathbf{E}f(L) \leq \mathbf{E}f(Y) \leq \mathbf{E}f(U)$ for monotone increasing, with the reverse order holding for monotone decreasing f . Allowing f to be an indicator variable for some graph property (e.g., strong connectivity) leads to corresponding bounds on the probability that Y will possess the property in question.

The power of Theorem 1 (and Corollary 1) lies in the fact that the properties of Bernoulli graphs are much better understood than the properties of general random graphs, the former being the subject of a large and growing portion of the graph theoretic literature (see, e.g. Bollobás, 2001). In some cases, analytical results are available for the properties of Bernoulli graphs, with good examples being bounds on the probability of connectivity, and distributional results on the incidence of subgraphs (e.g. the numbers of cycles or cliques). Where analytical results are not

available, exploration of model properties via simulation is far easier for Bernoulli graphs than for other random graph families. Thus, by “sandwiching” the behavior of an arbitrary graph model between two Bernoulli models, we may hope to obtain results regarding the behavior of the former by leveraging the ease of studying the latter. Of course, it does not follow that the bounds that result will always be sufficiently constraining to be helpful – particularly in the case of homogeneous graph models with strong edgewise dependence, the gap between Λ and Ψ may be too large to be effectively constraining. For models with weak dependence, however, and/or where dependence is heavily constrained by strong covariate effects, the bounding processes may be quite informative. In the next section, we demonstrate the utility of the bounding scheme by applying it to three basic problems regarding the behavior of general random graph models. As we shall see, the Bernoulli graph bounds allow us to draw several useful deductions regarding processes that would be difficult to study using other means.

3 Sample Applications

To illustrate the use of the Bernoulli graph bounds, we here consider three sample applications of the method to typical problems from the social network literature. In the first, we consider a well-known (and generally defective) model for networks with bias towards local clustering – the edge/triangle model – and show how the bounding approach can be used to quickly identify the model’s pathological asymptotic behavior. We then show how the bounds can be used to identify an alternative, well-behaved parameterization, and to establish that this latter model does not possess the problems of the original. In our second application, we consider a simplified model for human sexual contact networks, and demonstrate the use of Bernoulli graph bounds in identifying the nature and extent of tie formation biases (specifically, taboos on concurrent relationships and partner exchange) that are necessary to inhibit connectedness within large populations (a key factor in the diffusion of sexually transmitted diseases). This application demonstrates the power of the Bernoulli graph bounds to allow exploration of potentially complex social mechanisms, without recourse to simulation or numerical procedures. Our third, and final, application illustrates the use of Bernoulli graph bounds for inhomogeneous networks as a mechanism for evaluating the robustness of tie formation models to omitted sources of edgewise dependence. Specifically, we consider the impact of adding local triadic clustering to a spatial Bernoulli graph model of personal associations among residents of a U.S. county, in the context of rumor diffusion. While the system in question is too large to allow direct simulation of the extended model, the Bernoulli graph bounds allow us to place an upper bound on the effect of clustering bias on diffusion potential, thereby establishing a range of parameter values over which the omission of the dependence effect can be shown to have little impact on the outcome of interest.

3.1 Triadic Clustering and the “Density Explosion”

A long-standing interest in the development of general random graph models has been the ability to capture the excess of local clustering (relative to homogeneous Bernoulli graphs) found in many types of empirical networks (see, e.g. Davis, 1970; Holland and Leinhardt, 1970). Although described in general terms by Holland and Leinhardt (1981), the classic early development in this area is that of Frank and Strauss (1986), which established a principled basis for the modeling of triadic clustering via the inclusion of triangle statistics in ERG distributions. Paradoxically, the intuitive and well-motivated *Markov graph* family developed by Frank and Strauss was immediately recognized to give rise to asymptotic difficulties (as demonstrated by Strauss (1986)), and much subsequent work has been devoted to finding alternative parameterizations that are better

behaved (see, e.g. Pattison and Robins, 2002; Snijders et al., 2006; Hunter and Handcock, 2006; Hunter, 2007). The primary problem encountered in work with the Markov graphs has been that of *degeneracy*: given a model (specified by a fixed choice of θ and t) and taking a series of draws from the associated distributions on the graphs of increasing order (i.e., graphs of k vertices, $k + 1$ vertices, etc.), then almost all probability mass becomes concentrated in the large-order limit on a set of graphs having measure zero. Although degeneracy has many subtle dimensions (see the investigation of Handcock, 2003, as well as the above-cited Strauss paper), as a practical matter it frequently manifests in a phenomenon of “density explosion”: with very high probability, models incorporating elements such as triangle terms produce graphs that are nearly complete, even when the parameters associated with said terms are quite small. While detailed analysis of Markov graph behavior is complex, the use of Bernoulli bounds provide a very simple method of detecting the potential for runaway density, and for identifying parameterizations that avoid it.

For illustrative purposes, we will restrict ourselves here to the simple edge/triangle model family, which is parameterized in terms of the counts of each of these features within the graph. In ERG form, we may specify this family by defining \mathcal{Y}_N to be the set of all symmetric, zero-diagonal adjacency matrices of order N and setting $t(y, X) = (t_e(y), t_t(y))$ with $t_e(y) = \sum_{i=1}^N \sum_{j=i+1}^N y_{ij}$ and $t_t(y) = \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N y_{ij}y_{jk}y_{ik}$. Intuitively, our interest is in the case in which $\theta_2 > 0$, this being the region of the parameter space in which triangle formation is enhanced. To obtain Bernoulli graph bounds for this family, we must determine the values of Λ and Ψ ; as this model is fully homogeneous, we may consider an arbitrary off-diagonal i, j pair to analyze. For the lower-bound case, we proceed by substituting the above definition of t (and the θ constraint) into that of Λ_{ij} and carrying out the appropriate minimization:

$$\Lambda_{ij} = \min_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) \quad (7)$$

$$= \min_{y' \in \mathcal{Y}_N} \left[1 + \exp \left[\theta_1 \left(t_e \left(y_{ij}^- \right) - t_e \left(y_{ij}^+ \right) \right) + \theta_2 \left(t_t \left(y_{ij}^- \right) - t_t \left(y_{ij}^+ \right) \right) \right] \right]^{-1} \quad (8)$$

$$= [1 + \exp [-\theta_1]]^{-1}. \quad (9)$$

Observe that $t_e \left(y_{ij}^- \right) - t_e \left(y_{ij}^+ \right)$ is identically -1 , so θ_1 carries forward deterministically. For t_t , we note that the minimum edge probability case occurs for θ_2 when $t_t \left(y_{ij}^- \right) - t_t \left(y_{ij}^+ \right) = 0$; such a state does exist within the support (it is trivially satisfied by the empty graph), and hence the θ_2 term vanishes. We are left with a homogeneous Bernoulli graph whose expected density is the inverse logit of θ_1 as the lower bound on Y .

For the upper bound, we follow the analogous procedure:

$$\Psi_{ij} = \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) \quad (10)$$

$$= \max_{y' \in \mathcal{Y}_N} \left[1 + \exp \left[\theta_1 \left(t_e \left(y_{ij}^- \right) - t_e \left(y_{ij}^+ \right) \right) + \theta_2 \left(t_t \left(y_{ij}^- \right) - t_t \left(y_{ij}^+ \right) \right) \right] \right]^{-1} \quad (11)$$

$$= [1 + \exp [-\theta_1 - \theta_2 (N - 2)]]^{-1}. \quad (12)$$

In this case, note that the maximum edge probability occurs when the change in the number of triangles is maximized; since adding a single edge can create at most $N - 2$ triangles (this occurring when y_{ij}^c is complete), this adds an effect of $\theta_2 (N - 2)$ to the θ_1 baseline. While this again gives us a homogeneous Bernoulli graph, it is important to observe that the expected density in this case is not independent of N – on the contrary, density here increases logistically in N . Thus, we can immediately see evidence of the “density explosion” alluded to earlier: for any positive value of θ_2

(and any finite θ_1), there exists a finite N^* such that the expected density, δ of the upper bound is greater than any value $\delta^* < 1$ for all $N > N^*$. Although this does not in and of itself establish that the same is true of Y , it does mean that no large fluctuations in density can be ruled out. In fact, precisely this behavior is observed in practice, for the very reason suggested by the derivation of Ψ – triangle formation in this model is conditionally self-reinforcing, and thus graphs with many triangles quickly become orders of magnitude more probable than those without. While negative values of θ_1 can stave off this effect for small N , the number of potential triangles grows much faster than the number of potential edges ($O(N^3)$ versus $O(N^2)$), and the latter will eventually overwhelm the former. When this occurs, probability mass concentrates on high-density graphs.

Examination of the form of Equation 12 suggests a possible resolution to this problem. Since the maximum conditional probability is driven by the maximum number of triangles that can be generated by adding a single edge – a number that scales with N – substituting a related statistic whose difference (or *changescore*) grows more slowly can potentially avoid the density explosion. For instance, consider $t_{\nu}(y) = 3/N \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N y_{ij}y_{jk}y_{ik}$ (the mean number of triangle memberships per vertex) as an alternative clustering statistic to t_{ν} . The lower bound process for the modified edge/triangle model with this statistic is identical to the above (since triangle addition still increases conditional edge probability) but Ψ_{ij} is modified by the $3/N$ multiplier of t_i . The corresponding bounding model parameter now becomes

$$\Psi_{ij} = [1 + \exp[-\theta_1 - \theta_2 3(N-2)/N]]^{-1}. \quad (13)$$

As $N \rightarrow \infty$, we now have $\Psi_{ij} \rightarrow [1 + \exp[-\theta_1 - 3\theta_2]]$, and the density explosion is thus curbed. Moreover, since Ψ_{ij} is the parameter for the upper bounding process (and since density is a subgraph monotone property), it follows that the expected density of Y cannot exceed this bound – indeed, from Corollary 1, we know that no density quantile of Y can exceed the corresponding quantiles of a Bernoulli graph with parameter Ψ_{ij} . By using the properties of the bounding process, we can easily establish the safety of this alternative model family (at least, with respect to density explosion).

As a side note, density explosion is not the only pathology to which this model family is subject. For instance, in many practical applications, the mean degree (i.e., $\delta(N-1)$) is bounded (if not fixed), and cannot increase indefinitely with N . That is clearly a problem for the above, as even our lower bound leads to an expected degree of $(N-1)\Lambda_{ij} = (N-1)[1 + \exp[-\theta_1]]^{-1} \xrightarrow{N \rightarrow \infty} \infty$. While we might think to solve this problem by dividing t_e by N (as we did with t_t), this does not lead to the intended result. (Indeed, by effectively shrinking the impact of θ_1 towards 0, one approaches the uniform random graph (and thus $\delta = 0.5$) in the limit of N !) In this case, a more natural solution lies in retaining the edge statistic, and seeking a curved parameterization for θ_1 in terms of N : solving for constant mean degree leads to the parameterization $\theta_1 = \log\left(\frac{\beta}{N-1-\beta}\right)$, where β is the expected degree of the lower bound process. Combining this modification with the development of Equation 13 gives us the limit $\Psi_{ij} \xrightarrow{N \rightarrow \infty} \left[1 + \exp\left[-\log\left(\frac{\beta}{N-1-\beta}\right) - 3\theta_2\right]\right]$, which (multiplying by $N-1$) leads to the asymptotic expected mean degree bound $\beta \exp(3\theta_2)$. Thus, while the exact properties of this modified edge/triangle model may be complex, the Bernoulli bounds assure us that its mean degree will vary within predictable limits (limits that are easily “tuned” by β and θ_2).

3.2 Connectedness Inhibition and Epidemic Potential in Bipartite Graphs

In cases like the edge/triangle model, above, our interest was in either identifying possible pathologies in model behavior, or in identifying parameterizations that could be guaranteed to have reason-

able behavior. Another, more substantively theoretical, use of bounding graphs is to determine the nature and magnitude of effects that are needed to give rise to complex phenomena. Consider, for instance, the case of disease propagation over a heterosexual contact network. Setting aside (perhaps at peril, per Butts, 2009) the detailed dynamics of relational turnover, a first cut at estimating the maximum epidemic potential of such a network may be given by considering the connectedness of the network over some appropriately selected time interval. We may naturally represent the network itself as a bipartite graph, with ϵN members of one sex and $(1 - \epsilon)N$ of the other (taking $\epsilon \geq 0.5$, and being at present disinterested in which group corresponds to males versus females). Morris et al. (2007) have emphasized the role of relational concurrency (here, degree greater than 1) as a key enabler of disease transmission, and it goes without saying that, in many cultures, norms limiting concurrency are a potentially important factor in tie formation. Another, more subtle effect on tie formation is suggested by Bearman et al. (2004), who present evidence for suppression of 4-cycles in adolescent romantic relationships (presumably stemming from the fact that 4-cycles represent “partner swapping,” which may be taboo in some groups). We may reasonably wonder whether factors such as concurrency or partner swapping taboos are strong enough to potentially inhibit the connectedness of such networks, and if so what degree of effect strength is adequate for such an effect. While our methods cannot tell us when inhibition *will* occur, they nevertheless allow us to identify *minimal conditions* for inhibition – put another way, they give us a very simple tool to quickly determine when complex effects such as concurrency or 4-cycle suppression will be insufficient to disconnect a bipartite graph with high probability. At the opposite extreme from full connectivity (which represents the “worst case” epidemic potential), we can also consider bounds on the *basic reproductive number*, R_0 , associated with a sexual contact network. R_0 is defined as the expected number of new cases arising from the infection of a randomly chosen individual in an uninfected population, and is an important indicator of the capacity of a network to support an epidemic: where $R_0 < 1$, new infections will tend to die out, and thus $R_0 > 1$ is typically necessary for an epidemic to arise. By identifying minimal conditions for R_0 to be less than unity, we can therefore draw some basic conclusions regarding the circumstances under which epidemics may be fully suppressed (at least potentially).

As a first step in this analysis, we require a characterization of the conditions under which a random bipartite graph is connected. As with conventional random graphs, connectedness occurs as a threshold behavior as $N \rightarrow \infty$, although the detailed results are necessarily somewhat different. Most relevant in our case is Theorem 1 of Klee et al. (1981), which provides one such result for bipartite graphs in which the population is not dominated almost entirely by a single group (i.e., males or females).³ Translated into our notation, let G_k, G_{k+1}, \dots be an independent sequence of Bernoulli bipartite graphs defined as above, with G_N being of order N and having density δ_N . We define the quantity

$$\gamma_N = (1 - \epsilon) N \left[1 - (1 - \delta_N) (\epsilon N)^{\frac{1}{(1-\epsilon)N}} \right], \quad (14)$$

a function that grows slightly faster than the number of edges. (It is, in fact, asymptotically equal to the number of edges of G_N divided by ϵ .) Klee et al.’s theorem then states that if $\gamma_N \xrightarrow{N \rightarrow \infty} c$, then $\Pr(G_N \text{ connected}) \xrightarrow{N \rightarrow \infty} \exp(-\exp(-c))$. Since this is true for any fixed c , it thus follows that if for all $c > 0$ there exists an N^* such that $\gamma_N > c$ for all $N > N^*$, $\Pr(G_N \text{ connected}) \xrightarrow{N \rightarrow \infty} 1$.

³Specifically, the required condition in our notation is $(2 - 1/\epsilon) \log N \xrightarrow{N \rightarrow \infty} \infty$, recalling that $\epsilon \geq 0.5$. The case of $\epsilon = 0.5$ technically belongs to Theorem 2 of Klee et al. (1981), but leads to the same result as Theorem 1. We thus employ the latter here.

When will this occur? Taking $\partial\gamma_N/\partial N = 0$ and solving for δ provides one line of attack:

$$\delta_N^* = 1 - \frac{(1 - \epsilon)N}{(1 + (1 - \epsilon)N - \log(\epsilon N)) (\epsilon N)^{\frac{1}{(1-\epsilon)N}}}, \quad (15)$$

with δ_N^* being the maximum density at which γ_N remains unchanged. δ_N^* provides a useful bound for assessing asymptotic connectedness, in the sense that any series of density values increasing strictly faster than δ_N^* will eventually lead to bipartite graphs that are connected with probability 1. A minimal condition for the possibility of disconnection in our general bipartite graph process is thus that the expected density sequence of its corresponding lower bounding process fall at least as quickly as δ_N^* for sufficiently large N . In assessing this, it is useful to work with a simpler, limiting form of the above. We begin by consolidating and rearranging terms:

$$\delta_N^* = \frac{(1 + (1 - \epsilon)N - \log(\epsilon N)) (\epsilon N)^{\frac{1}{(1-\epsilon)N}} - (1 - \epsilon)N}{(1 + (1 - \epsilon)N - \log(\epsilon N)) (\epsilon N)^{\frac{1}{(1-\epsilon)N}}} \quad (16)$$

$$= \frac{1 + (1 - \epsilon)N \left(1 - (\epsilon N)^{\frac{-1}{(1-\epsilon)N}}\right) - \log(\epsilon N)}{1 + (1 - \epsilon)N - \log(\epsilon N)} \quad (17)$$

$$= \frac{1 + (\epsilon - 1)N \left((\epsilon N)^{\frac{1}{(\epsilon-1)N}} - 1\right) - \log(\epsilon N)}{1 + (1 - \epsilon)N - \log(\epsilon N)}. \quad (18)$$

Observe that, as $N \rightarrow \infty$, $(\epsilon - 1)N \left((\epsilon N)^{\frac{1}{(\epsilon-1)N}} - 1\right) \rightarrow \log(\epsilon N)$. The numerator thus reduces to 1. For large N , the denominator will be dominated by the order N term; combining these results then gives us the limiting form

$$\delta_N^* \xrightarrow{N \rightarrow \infty} \frac{1}{(1 - \epsilon)N}. \quad (19)$$

Now, let us turn to our network model. By assumption, our support is the set of bipartite adjacency matrices having ϵN and $(1 - \epsilon)N$ vertices, and our model terms (in ERGM form) are defined by $t(y, x) = (t_e(y), t_c(y), t_{C_4}(y))$, with t_e being the edge count, t_c being the number of vertices having degree greater than 1 (i.e., concurrent partnerships), and t_{C_4} being the number of 4-cycles. We are interested in the region of the parameter space for which $\theta_2, \theta_3 \leq 0$, i.e. for which concurrency and partner-swapping are both discouraged. Furthermore, it seems sensible to assume here that the expected mean degree in the absence of secondary effects is asymptotically constant in N , since there is no reason to assume that the average number of short-term sex partners will grow substantially with population. Thus, we employ the mean degree stabilizing parameterization $\theta_1 = \log[\beta / (2\epsilon(1 - \epsilon)N - \beta)]$ (the bipartite analog of that used in the previous example), where β is the expected mean degree.

Our interest now is on the lower bounding graph, which is in this case a homogeneous undirected bipartite Bernoulli graph with common parameter Λ_{ij} (for any i and j belonging to different vertex sets). The maximum changescores here are 2 in the case of t_c (since adding an edge can increase the maximum number of persons with concurrent partnerships by at most 2), and $(\epsilon N - 1)((1 - \epsilon)N - 1)$ in the case of t_{C_4} (this being equal to the number of male-female pairs other than the i, j th). It is noteworthy that these two bounds cannot both be realized simultaneously for large N , since at most one 4-cycle can be created by an edge addition without the endpoints previously having degree greater than 1. The two extreme cases are thus the increase of t_c by 2 and t_{C_4} by 1, and the increase of t_{C_4} by $(\epsilon N - 1)((1 - \epsilon)N - 1)$ (with t_c constant). As

always, the changescore for t_e is 1. Given this, we may derive Λ_{ij} in the usual fashion:

$$\Lambda_{ij} = \min_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^{c'}) \quad (20)$$

$$= \min_{y' \in \mathcal{Y}_N} \left[1 + \exp \left[\theta_1 \left(t_e \left(y_{ij}^- \right) - t_e \left(y_{ij}^+ \right) \right) + \theta_2 \left(t_c \left(y_{ij}^- \right) - t_c \left(y_{ij}^+ \right) \right) + \theta_3 \left(t_{C_4} \left(y_{ij}^- \right) - t_{C_4} \left(y_{ij}^+ \right) \right) \right] \right]^{-1} \quad (21)$$

$$= \begin{cases} \left[1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - 2\theta_2 - \theta_3 \right] \right]^{-1} & \text{if } 2\theta_2 < \theta_3 [(\epsilon N - 1)((1 - \epsilon)N - 1)] \\ \left[1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - \theta_3 (\epsilon N - 1)((1 - \epsilon)N - 1) \right] \right]^{-1} & \text{otherwise} \end{cases} \quad (22)$$

$$\geq \left[1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - 2\theta_2 - \theta_3 (\epsilon N - 1)((1 - \epsilon)N - 1) \right] \right]^{-1}. \quad (23)$$

As noted, there are two potential lower bounds for this model, depending on the relative maximum impact of concurrency versus partner swapping events. For any fixed $\theta_3 < 0$, however, the latter will dominate the former for sufficiently large N , and thus the second case of Equation 22 is the appropriate large-population bound. On the other hand, the first case continues to be useful where $\theta_3 = 0$ (i.e., only concurrency effects are present). Incorporating both effects leads to the even looser bound of Equation 23, which may be adequate in some circumstances (and which does not require one to specify which of the two conditions of Equation 22 applies). Here, we will focus on the two cases in which $\theta_3 = 0$, and the large- N case of $\theta_3 < 0$ (in which the partner swapping effect dominates concurrency).

Considering the first of these bounds, we have a Bernoulli bipartite graph with expected density $\left[1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - 2\theta_2 - \theta_3 \right] \right]^{-1}$, versus threshold density δ_N^* . Our necessary condition for connectedness inhibition is

$$\lim_{N \rightarrow \infty} \frac{\delta_N}{\delta_N^*} < 1 \quad (24)$$

$$\frac{(1 - \epsilon)N}{1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - 2\theta_2 - \theta_3 \right]} < 1 \quad (25)$$

$$(1 - \epsilon)N < 1 + \left(\frac{2\epsilon(1 - \epsilon)N}{\beta} - 1 \right) \exp [-2\theta_2 - \theta_3] \quad (26)$$

$$1 - \epsilon < 1/N + \left(\frac{2\epsilon(1 - \epsilon)}{\beta} - 1/N \right) \exp [-2\theta_2 - \theta_3], \quad (27)$$

which further reduces in the large- N limit to

$$1 - \epsilon < \frac{2\epsilon(1 - \epsilon)}{\beta} \exp [-2\theta_2 - \theta_3], \quad (28)$$

and finally

$$\beta \exp [2\theta_2 + \theta_3] < 2\epsilon. \quad (29)$$

Recalling that β is the “baseline” mean degree (in the absence of other effects), this tells us that connectedness can be avoided only if the effective mean degree can be reduced to less than 2ϵ (which

will be in the interval $[1, 2)$, depending on the sex ratio) through the action of the concurrency and partner swapping parameters. As a practical matter, we observed earlier that this limit is only relevant when $\theta_3 = 0$, and thus this condition corresponds to the case in which concurrency is the only factor of consequence; solving for θ_2 in this case gives us $\theta_2 < \log(2\epsilon/\beta)/2$ as the threshold parameter value. When the concurrency penalty is greater (i.e., less negative) than this value, the underlying network will be almost surely connected for sufficiently large N .

Having explored the case in which the concurrency penalty dominates, we now turn to the case in which partner swapping taboos have a greater effect. Substituting the appropriate bound into our connectedness inhibition condition gives us

$$\lim_{N \rightarrow \infty} \frac{\delta_N}{\delta_N^*} < 1 \quad (30)$$

$$\frac{(1 - \epsilon)N}{1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - \theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1) \right]} < 1 \quad (31)$$

$$(1 - \epsilon)N < 1 + \left(\frac{2\epsilon(1 - \epsilon)N}{\beta} - 1 \right) \exp [-\theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1)] \quad (32)$$

reducing to

$$1 < \frac{2\epsilon}{\beta} \exp [-\theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1)] \quad (33)$$

and thus

$$\beta \exp [\theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1)] < 2\epsilon. \quad (34)$$

The form here is as before, save that an order N^2 term remains within the weighting factor on the left hand side. Assuming that $\theta_3 < 0$, this term must eventually overwhelm β . Thus, we can see that – unlike concurrency – a non-zero partner swapping penalty will *always* meet the necessary conditions for connectedness inhibition for N sufficiently large. This does not, of course, establish that such a penalty is also a *sufficient* condition for inhibiting connectedness, since the impact of suppressing 4-cycles will in practice accumulate more slowly than the extremal construction suggests. Nevertheless, this result gives us considerable intuition regarding the way in which this mechanism functions, and highlights an important (and a priori non-obvious) difference between the behavior of concurrency and partner swapping taboos (at least for the mechanisms implemented here).

How does all of this compare to the corresponding behavior of the basic reproductive number, R_0 ? As the expected number of new infections arising from single infective, clearly R_0 must be equal to the mean degree multiplied by the marginal probability of infection per contact. We here treat this rate as an arbitrary disease-specific constant, ζ . Given ζ , we can bound R_0 easily via Λ_{ij} :

$$R_0 \geq \zeta 2\epsilon(1 - \epsilon)N\Lambda_{ij} \quad (35)$$

$$= \begin{cases} \frac{2\zeta\epsilon(1-\epsilon)N}{1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - 2\theta_2 - \theta_3 \right]} & \text{if } 2\theta_2 < \theta_3 [(\epsilon N - 1) ((1 - \epsilon) N - 1)] \\ \frac{2\zeta\epsilon(1-\epsilon)N}{1 + \exp \left[-\log \left[\frac{\beta}{2\epsilon(1-\epsilon)N-\beta} \right] - \theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1) \right]} & \text{otherwise} \end{cases} \quad (36)$$

$$\xrightarrow{N \rightarrow \infty} \begin{cases} \zeta\beta \exp [2\theta_2 + \theta_3] & \text{if } 2\theta_2 < \theta_3 [(\epsilon N - 1) ((1 - \epsilon) N - 1)] \\ \zeta\beta \exp [\theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1)] & \text{otherwise} \end{cases} \quad (37)$$

Recalling that β is the baseline mean degree, the lower bound on R_0 approaches a “baseline” expected number of new infections ($\zeta\beta$) scaled by a “penalty” due to the action of concurrency and/or partner swapping taboos. In a loose sense, we may thus think of the additive inverse of the penalty factor as the fraction of new infections that are prevented by the proposed social mechanisms. To assess epidemic potential, we ask where the R_0 bounds cross 1:

$$R_0 < 1 \tag{38}$$

$$\zeta\beta \exp [2\theta_2 + \theta_3] < 1 \tag{39}$$

$$\log \left[\frac{1}{\zeta\beta} \right] > 2\theta_2 + \theta_3, \tag{40}$$

or, in the case of the second bound,

$$\zeta\beta \exp [\theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1)] < 1 \tag{41}$$

$$\log \left[\frac{1}{\zeta\beta} \right] > \theta_3 (\epsilon N - 1) ((1 - \epsilon) N - 1) \tag{42}$$

$$\frac{\log \left[\frac{1}{\zeta\beta} \right]}{(\epsilon N - 1) ((1 - \epsilon) N - 1)} > \theta_3. \tag{43}$$

As these are based on the lower bound, they imply that the sexual contact network will support an epidemic so long as θ_2, θ_3 are above a fixed level that scales as the negative log of the baseline number of new infections (where the concurrency taboo dominates), or so long as θ_3 is greater than the above threshold reduced by an $\mathcal{O}(N^2)$ term (partner swap case). Interventions aimed at eliminating epidemic potential must thus achieve at least this level of relational inhibition in order to have a chance at effectiveness.

Before leaving this example, it is worth commenting briefly on the upper bound for this model family. From the above, we can obtain Ψ_{ij} by setting $\theta_2 = \theta_3 = 0$, in which case we can see that $\lim_{N \rightarrow \infty} \frac{\delta_N}{\delta_N^*} < 1$ where $\beta < 2\epsilon$. In the event that the baseline expected degree is sufficiently low (the threshold being between 1 and 2, depending on ϵ), connectedness is unlikely regardless of other factors. Considering the upper bound serves to remind us that the impact of secondary mechanisms is only consequential here (from the point of view of connectedness) when the “natural” number of sex partners is reasonably high.

3.3 Adding Dependence to an Inhomogeneous Bernoulli Model

While we have focused thus far on the use of Bernoulli graph bounds to study asymptotic properties of homogeneous models with edgewise dependence, it should be noted that these bounds are also applicable in the inhomogeneous case. Of particular interest here are networks for which strongly predictive covariates are available, as one may hope in such cases to capture the behavior of such networks through Bernoulli models whose rate parameters are determined by the available covariates. (This is, indeed, the basis for *network regression* procedures (Krackhardt, 1988; Dekker et al., 2007), which reduce the problem of modeling edge structure to a more conventional generalized linear modeling problem.) One important application of covariate-based models is to the extrapolative simulation of large-scale networks, since Bernoulli models based on geographical or other similar information can be used to predict properties of social structure on large populations for which more complex models would prove computationally infeasible. This raises an important question, however: how sensitive are such models to the assumption of conditional independence among edges? If the covariates employed are sufficiently powerful (and the model correct), dependence assumptions are unnecessary to account for the bulk of the information content in the

modeled network (see, e.g. Butts, 2003), but this does not guarantee that specific features of importance will be sufficiently well-described for any given application. As a practical matter, one would like to have some means of verifying that the Bernoulli assumption is inconsequential, at least in the limited sense that adding one or more sources of dependence will not change model behavior in such a way as to make it unsuited for its intended application. While model elaboration followed by simulation analysis would seem to be a natural approach to this problem, the poor scalability of simulation methods for random graphs with edgewise dependence makes this impractical for large-scale structures. In this case, Bernoulli graph bounds can often be used to provide an effective “check” on the robustness of model behavior to the addition of edgewise dependence. These bounds can also be employed more affirmatively as a scalable way of investigating the impact of weak dependence mechanisms in large-scale systems, by first approximating the network in inhomogeneous Bernoulli form, and then exploring the change in bounding graph behavior as dependence is added. Whether intended as an adequacy checking procedure or as an exploratory device, the relative ease of simulating draws from Bernoulli graphs makes them useful tools even when analytic results are unavailable.

As our final application, we demonstrate the use of Bernoulli graph bounds to examine the robustness of a large-scale geographical network model to the introduction of endogenous triadic clustering. The network model we employ is an extrapolative application of the “social friendship” model of (Butts, 2002), an inhomogeneous Bernoulli graph model based on friendship data from a housing study by Festinger et al. (1950). Written in conditional ERG form, this model is a curved exponential family (Hunter and Handcock, 2006) specified by

$$\Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) = \left[1 + \exp \left[-\log \left(\frac{(1 + \alpha D_{ij})^\gamma}{p_b} - 1 \right) \right] \right]^{-1} \quad (44)$$

where D_{ij} is the geographical distance between i and j , $p_b \in (0, 1]$ is the baseline probability of an i, j edge (at distance 0), $\alpha \in (0, \infty)$ is a scaling parameter, and $\gamma \in (0, \infty)$ is a shape parameter governing the form of the decay of tie probabilities at long distances. Here, we use the median parameter estimates from Butts (2002), specifically $p_b = 0.529$, $\alpha = 0.031423$, and $\gamma = 2.768$. Vertex positions for this simulation are based on block level data from the year 2000 US Census, for the Choctaw County, MS, metropolitan statistical area. Specifically, each household within each census block is assigned a graphical location uniformly at random within the block area (with individuals not belonging to households being treated as households of size 1). After assigning households to positions, each household member is assigned a position from a circular jittering distribution centered on the household location, with a maximum radius of 5 meters. This placement process thus preserves within-block population densities, while also respecting local clustering associated with variable household size. The resulting population distribution is depicted in Figure 1; note that the total number of vertices is 9,758. Although this is far too large a network to permit simulation of general exponential family models, it is well within the realm of feasible Bernoulli graphs.

To illustrate the use of Bernoulli graph bounds to study the effects of complex processes, we focus our attention on a dynamic phenomenon: the diffusion of a hypothetical rumor through the target community. Specifically, we consider a process in which a rumor is introduced to a randomly chosen member of the population, who then conveys the rumor to his or her alters. The alters, in turn, convey the message to their peers, and so on. For our purposes, we will restrict attention the maximum number of persons reachable through such a process (assuming that the message diffuses until it can no longer be transmitted); this corresponds to the distribution of the component sizes of the simulated network, weighted by probability of selection (here, proportionate to component

Choctaw, MS, Population Distribution

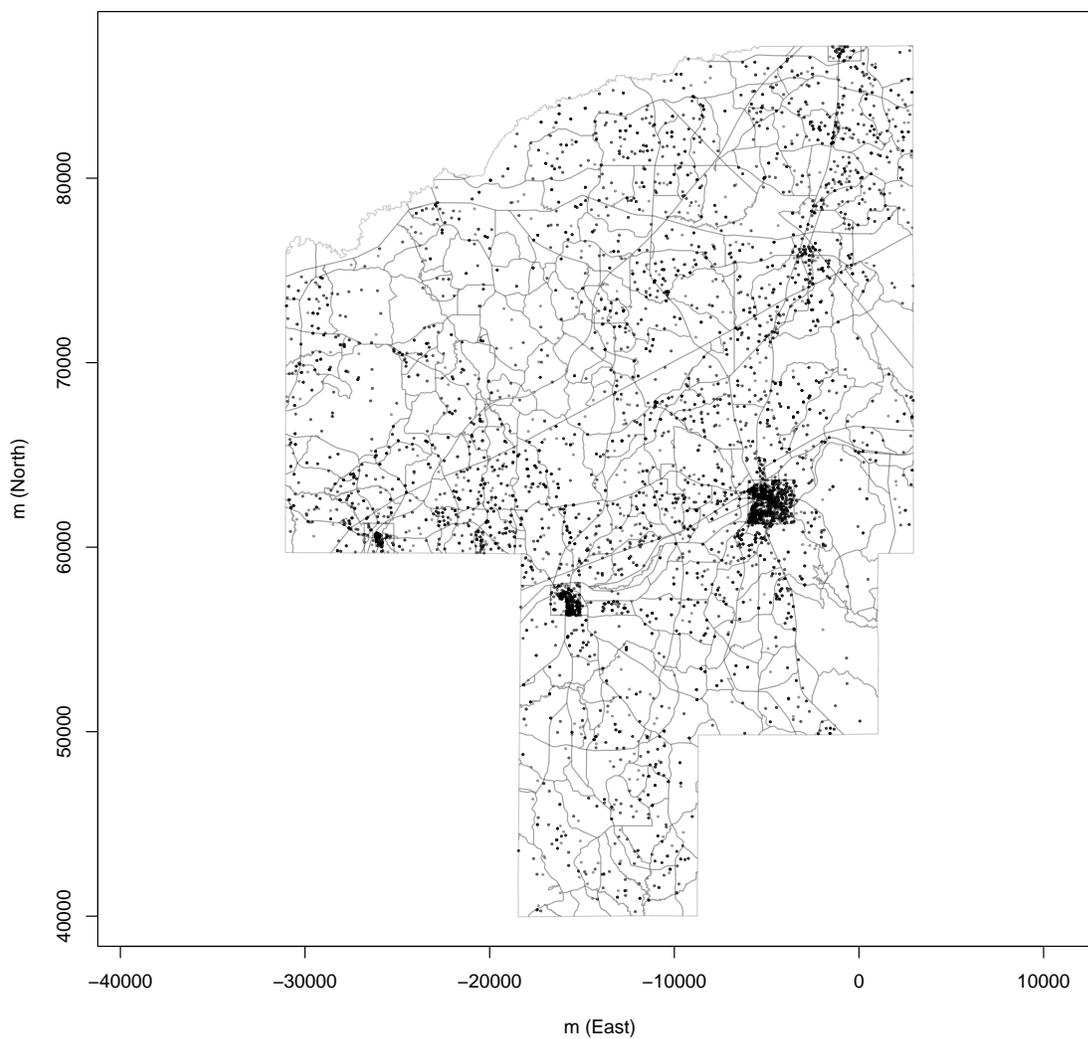


Figure 1: Population distribution for Choctaw County, MS. Points are placed uniformly by census block ($N = 9,758$).

size). It should be noted that both the mean and the order statistics for the number of people to be reached by such a process are subgraph monotone, as adding edges can only increase the number of persons eventually reached. Thus, we can bound the number of persons reached by the diffusing rumor within an arbitrary ERG by its associated Bernoulli graph bounds.

Given this phenomenon of interest, our central question is the extent to which behavior of this process as realized on the inhomogeneous Bernoulli model applied to the Choctaw, MS case, is robust to the assumption of edgewise independence. Specifically, we consider the possibility that we may have omitted a process of geographically local clustering, in which the presence of shared partners proximate to a pair of actors increases the propensity of that pair to form an edge (e.g., due to participation in shared social events, opportunities to encounter or learn about one another through the shared partners, etc.). The form we use for the corresponding statistic reflects a *neighborhood dependence* (specifically, a *setting structure*, in the sense of Pattison and Robins (2002)): edges are only dependent within this model if all three parties (ego, alter, and the third party) are within some specified distance τ of one another. This leads naturally to a *local triangle statistic*, of the form $t_{t_\ell}(y, D, \tau) = \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N I(D_{ij} < \tau) I(D_{jk} < \tau) I(D_{ik} < \tau) y_{ij} y_{jk} y_{ik}$. Combined with the initial random graph model, this generates the expanded model family

$$\Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) = \left[1 + \exp \left[\log \left(\frac{(1 + \alpha D_{ij})^\gamma}{p_b} - 1 \right) - \theta I(D_{ij} < \tau) \sum_{k \neq i, j} I(D_{jk} < \tau) I(D_{ik} < \tau) y_{jk} y_{ik} \right] \right]^{-1}. \quad (45)$$

As remarked earlier, triangle terms reflect a very strong form of triadic coupling, and imposing the neighborhood condition weakens this only slightly: while the number of potential triangles is reduced by the condition that all members be spatially proximate, no restriction is placed on coupling within that set. Thus, the expanded family represents a relatively strong perturbation to the initial Bernoulli family. We can obtain Bernoulli bounds for the expanded family in the usual manner; our interest here is on the impact of positive clustering, and hence the upper bound

$$\Psi_{ij} = \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}^c) \quad (46)$$

$$= \left[1 + \exp \left[\log \left(\frac{(1 + \alpha D_{ij})^\gamma}{p_b} - 1 \right) - \theta I(D_{ij} < \tau) \sum_{k \neq i, j} I(D_{jk} < \tau) I(D_{ik} < \tau) \right] \right]^{-1}. \quad (47)$$

Unlike our previous examples, this bounding graph is inhomogeneous, with both the base tie probability and the clustering effect varying due to the underlying geography. Rather than analytical results, we thus employ simulation to examine the impact of clustering on the diffusion process, studying the distribution of persons reached by a randomly chosen initial contact as a function of θ (the clustering parameter) and τ (the radius over which the clustering process operates). For sufficiently large values of θ and τ , the triangle term is expected to dominate the base tie probability, and the graph's behavior will be much like the edge/triangle model. For small values of θ and τ , however, we may imagine that the graph distribution will not change sufficiently to produce a large change in this behavior of the diffusion process. The operative question is thus how large these parameters must be to have a substantial impact on the number of persons reached. If the latter quantity changes little under the Ψ -based graph for small to moderate θ and τ values, that suggests that the inhomogeneous graph model is relatively robust to local triangulation. If, on the other hand, we find that the introduction of even low levels of dependence produces dramatic changes in

the number of persons reached, then we may conclude that the model is non-robust (at least for this case), possibly suggesting the need for a more elaborate treatment.

Before examining the simulation results, it is useful to consider the question of where we would *expect* to see a large effect of local triadic coupling on the Bernoulli upper bound. Let the *triadic exposure* of a vertex pair be the maximum number of shared partners for that pair, under the constraint that all three pairs (the original, and each of the members of the original with the partner) are within distance τ of one another. For a pair within τ of each other, this reduces to the number of vertices within τ of both members. To get some idea of how this might scale, let us assume that our focal pair resides at some distance $\ell\tau$ from one another, with $0 \leq \ell < 1$, and let us further assume that the population density within the vicinity of the pair is uniformly ρ (with τ being in units of distance, and ρ in units of persons per distance squared). The expected triadic exposure of the pair is then equal to ρ times the area of the intersection of two τ -radius circles with centers $\ell\tau$ apart, or $\rho\tau^2 (2 \cos^{-1}(\ell/2) - \sin(2 \cos^{-1}(\ell/2)))$. To simplify this expression, it is helpful to note (from its Taylor series expansion) that $2 \cos^{-1}(\ell/2) - \sin(2 \cos^{-1}(\ell/2)) = \pi - 2\ell + \ell^3/12 + \dots$, with higher order terms falling rapidly to 0. Since $0 < \ell < 1 \ll \tau$, it is thus reasonable to approximate the expected triadic exposure of a given edge by $\rho\tau^2(\pi - 2\ell)$, with upper bound $\pi\rho\tau^2$. It follows, then, that the triadic exposure will be dominated primarily by τ , and that it will grow rapidly as τ becomes large. The baseline odds of an i, j edge are thus multiplied by a factor that grows at maximum with $\exp[\theta\pi\rho\tau^2]$, and we can expect the impact of triadic clustering to be poorly constrained whenever $\tau > 1/\sqrt{\theta\pi\rho}$. For the Choctaw, MS case, $\rho \approx 9\text{persons/km}^2$ (although this estimate is for the entire region, and underestimates the higher local population densities experienced by most persons). Using this as a starting point, we might expect to see substantial triadic exposure effects for this area where $\tau > 188/\sqrt{\theta}$ (or, equivalently, $\theta > (188/\tau)^2$). That would imply, for instance, that effects due to triangulation for $\tau \approx 1\text{km}$ should be modest for $\theta < 0.035$, but extending the critical radius by a factor of 2 would cut the permissible θ by a factor of 4. How well does this heuristic do at accounting for the real impact of local triangulation on the upper bound? For that, we turn to simulation.

To bound the effect of local triangulation on the number of persons potentially reached by a rumor started by a randomly chosen member of the population, we proceed by simulating the corresponding process on draws from the upper Bernoulli bound for various values of τ and θ . Noting that the number of persons reachable from a given individual is simply the size of said person's component, we speed computation by simply identifying the size of the component belonged to by each member of the network (each of whom has an equal chance to be the seed node), replicating this process with a new graph 500 times. The resulting mean and median numbers of persons reached are shown in Table 1. As can be seen, the distribution of components (and hence numbers reached) is quite skewed, with a random seed under the baseline process reaching about 100 persons on average versus a median of 3. Our illustrative relation is thus heavily fragmented.⁴ Adding triangulation effects, we find little impact on the expected number of persons reached until around 80m (for large values of θ), with substantial effects for low θ appearing at around 730m. While large in relative terms (e.g., a doubling of the expectation), these effects are still modest in reference to the larger population until scales of approximately 2km are reached, at which point moderate to large θ lead to expected diffusion potential that is on par with the entire population. The median number of persons reached is noticeably more robust, with no substantial change until approximately 730m (and then only for very large θ), and qualitatively large changes emerging only for moderate to

⁴Although we are not here concerned with the plausibility of this network as a realistic model for rumor transmission, we do note that the highly skewed distribution of diffusion sizes is consistent with empirical findings (e.g. Back et al., 1950; Schall et al., 1950; Schachter and Burdick, 1955) suggesting that most randomly inserted rumors do *not* diffuse under normal conditions, with a small number becoming widely disseminated.

		Mean Number Reached									
θ	0.0000	0.0020	0.0039	0.0078	0.0156	0.0312	0.0625	0.1250	0.2500	0.5000	
$\tau = 0\text{m}$	103.02	103.31	103.04	104.04	103.67	103.47	103.70	103.76	103.29	103.93	
3m	103.29	103.89	103.06	103.48	102.96	103.66	104.31	104.45	104.62	107.07	
9m	103.45	104.63	104.18	104.15	104.42	106.11	108.36	112.27	118.59	125.94	
27m	104.47	103.62	103.98	104.48	105.75	107.35	112.34	118.70	127.97	138.63	
81m	104.25	105.29	107.07	111.53	119.09	129.35	149.81	171.89	197.90	224.19	
243m	103.27	122.46	139.58	165.49	197.61	241.84	292.11	339.01	364.75	380.68	
729m	102.68	225.80	284.04	331.03	363.88	394.30	441.22	474.45	554.09	800.55	
2187m	103.10	409.02	460.34	513.53	590.84	995.69	4779.88	7270.19	8947.38	9651.42	
		Median Number Reached									
θ	0.0000	0.0020	0.0039	0.0078	0.0156	0.0312	0.0625	0.1250	0.2500	0.5000	
$\tau = 0\text{m}$	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	
3m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	
9m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	
27m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	
81m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00	
243m	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00	4.00	5.00	
729m	3.00	3.00	3.00	3.00	4.00	4.00	5.00	6.00	26.00	139.00	
2187m	3.00	4.00	4.00	5.00	7.00	553.00	6825.00	8383.00	9345.00	9706.00	

Table 1: Upper bounds for the mean and median numbers of persons reachable from a randomly chosen seed as a function of θ and τ , based on 500 simulated networks.

large θ at $\tau \approx 2\text{km}$. How do these effects compare with the scaling heuristic? Examination of Table 1 suggests that the heuristic does a reasonable job of indicating where substantial deviations from the baseline begin to accumulate, although there is considerable quantitative detail that appears only in the simulations. Moreover, the simulations allow us to distinguish between changes that are noteworthy but qualitatively limited (e.g., a doubling or tripling of the initially meager diffusion potential) and those that are qualitatively substantial (e.g., diffusions that reliably reach most members of the population). In practice, both are useful tools, and both are made possible by our ability to place simple bounds on a highly complex random process.

To summarize our findings for this illustrative case, we can clearly say that the inhomogeneous Bernoulli model is robust to local triangulation (in terms of diffusion potential) where the radius of the local clustering process is 250m-730m for moderate to large values of θ (e.g., each two-path increasing the odds of an edge by 13% or more), with robustness to 2km or more when θ is small (e.g., each two-path increasing the odds of an edge by less than about 0.4%). If one is concerned with qualitative shifts (i.e., movement from a world of small outbreaks to diffusion on par with population size), then the model is robust to triangulation out to at least the 2km scale if θ is not large (in the above sense). We expect for the θ sensitivity to grow roughly in the square of τ , so substantially larger radii of dependence will require vanishingly weak two-path effects for robustness to be guaranteed. These being upper bounds on the triangulation effect – and the form of local triangulation here being the strongest possible (i.e., all edges are potentially affected by all shared partners within the local radius) – we cannot guarantee that the original model would be a poor approximation in plausible scenarios with longer-range dependence. We can, however, establish that positive triadic dependence associated with processes unfolding on the scales of neighborhoods

or small communities would not result in radical changes in diffusion potential for this particular case, so long as the impact of any given two-path on edge formation is reasonably modest.

4 Conclusion

Although the ability to specify, simulate, and estimate parameters for general random graph models has represented a tremendous advance in our capacity to study relational systems, this ability has not brought with it a similarly rich set of methods for understanding the models themselves. With the exception of simulation – a powerful, but limited, tool – there have been few if any viable techniques for studying the behavior of network models incorporating general forms of edgewise dependence. The Bernoulli graph bounds developed here represent one attempt at such a development. While not all problems can be formulated in terms that are amenable to analysis using the Bernoulli bounds, the sample applications shown above demonstrate the value of the approach for solving problems that would be difficult or impossible to solve using previously existing methods. It is hoped that this development will encourage further research along these lines, thereby expanding our ability to employ general random graph models across a range of scientific contexts.

5 References

- Back, K., Festinger, L., Hymovitch, B., Kelley, H., Schachter, S., and Thibaut, J. (1950). The methodology of studying rumor transmission. *Human Relations*, 3:307–312.
- Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110:44–91.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.
- Bollobás, B. (2001). *Random Graphs*. Cambridge University Press, Cambridge, second edition.
- Butts, C. T. (2002). *Spatial Models of Large-scale Interpersonal Networks*. Doctoral Dissertation, Carnegie Mellon University.
- Butts, C. T. (2003). Predictability of large-scale spatially embedded networks. In Breiger, R., Carley, K. M., and Pattison, P., editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. National Academies Press, Washington, D.C.
- Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, 325:414–416.
- Coleman, J. S. (1964). *Introduction to Mathematical Sociology*. Free Press, New York.
- Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two theoretical models on 742 sociograms. *American Sociological Review*, 35:843–852.
- Dekker, D., Krackhardt, D., and Snijders, T. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, 72(4):563–581.
- Festinger, L., Schachter, S., and Back, K. (1950). *Social Pressures in Informal Groups*. Stanford University Press, Stanford, California.
- Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81:832–842.

- Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, Vancouver.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood calculations (with discussion). *Journal of the Royal Statistical Society, Series C*, 54:657–699.
- Handcock, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In Breiger, R., Carley, K. M., and Pattison, P., editors, *Dynamic Social Network Modeling and Analysis*, pages 229–240. National Academies Press, Washington, DC.
- Holland, P. W. and Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 70:492–513.
- Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373):33–50.
- Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, 29:216–230.
- Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583.
- Klee, V. L., Larman, D. G., and Wright, E. M. (1981). The proportion of labelled bipartite graphs which are connected. *Journal of the London Mathematical Society*, 24:397–404.
- Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analyses of dyadic data. *Social Networks*, 10:359–382.
- Mayhew, B. H. (1980). Structuralism versus individualism: Part i, shadowboxing in the dark. *Social Forces*, 59:335–375.
- Morris, M., Goodreau, S., and Moody, J. (2007). Sexual networks, concurrency, and STD/HIV. In Holmes, K. K., Sparling, P. F., Stamm, W. E., , P. P., Wasserheit, J. N., and Corey, L., . e. a., editors, *Sexually Transmitted Diseases*, chapter 7. McGraw-Hill, New York, 4th edition.
- Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32:301–337.
- Schachter, S. and Burdick, H. (1955). Field experiment on rumor transmission and distortion. *Journal of Abnormal and Social Psychology*, 50:363–371.
- Schall, H. M., Levy, B., and Tresselt, M. E. (1950). A sociometric approach to rumor. *Journal of Social Psychology; Political, Racial, and Differential Psychology*, 31(1):121–129.
- Simmel, G. (1897). The persistence of social groups: I. *American Journal of Sociology*, 3:662–698.
- 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., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36.
- Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*, 28(4):513–527.

- Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p^* . *Psychometrika*, 60:401–426.
- Wasserman, S. and Robins, G. (2005). An introduction to random graphs, dependence graphs, and p^* . In Carrington, P. J., Scott, J., and Wasserman, S., editors, *Models and Methods in Social Network Analysis*, chapter 10, pages 192–214. Cambridge University Press, Cambridge.