NEW SPECIFICATIONS FOR EXPONENTIAL RANDOM GRAPH MODELS

*Tom A. B. Snijders**
Philippa E. Pattison[†]
Garry L. Robins[†]
Mark S. Handcock[‡]

The most promising class of statistical models for expressing structural properties of social networks observed at one moment in time is the class of exponential random graph models (ERGMs), also known as p^* models. The strong point of these models is that they can represent a variety of structural tendencies, such as transitivity, that define complicated dependence patterns not easily modeled by more basic probability models. Recently, Markov chain Monte Carlo (MCMC) algorithms have been developed that produce approximate maximum likelihood estimators. Applying these models in their traditional specification to observed network data often has led to problems, however, which can be traced back to the fact that important parts of the parameter space correspond to nearly degenerate distributions, which may lead to convergence problems of estimation algorithms, and a poor fit to empirical data.
This paper proposes new specifications of exponential random graph models. These specifications represent structural properties

We thank Emmanuel Lazega for permission to use data collected by him. A portion of this paper was written in part while the first author was an honorary senior fellow at the University of Melbourne.

*University of Groningen
[†]University of Melbourne
[‡]University of Washington

such as transitivity and heterogeneity of degrees by more complicated graph statistics than the traditional star and triangle counts. Three kinds of statistics are proposed: geometrically weighted degree distributions, alternating k-triangles, and alternating independent two-paths. Examples are presented both of modeling graphs and digraphs, in which the new specifications lead to much better results than the earlier existing specifications of the ERGM. It is concluded that the new specifications increase the range and applicability of the ERGM as a tool for the statistical analysis of social networks.

1. INTRODUCTION

Transitivity of relations—expressed for friendship by the adage "friends of my friends are my friends"—has resisted attempts to be expressed in network models in such a way as to be amenable for statistical inference. Davis (1970) found in an extensive empirical study on relations of positive interpersonal affect that transitivity is the outstanding feature that differentiates observed data from a pattern of random ties. Transitivity is expressed by triad closure: if i and j are tied, and so are j and h, then closure of the triad i, j, h would mean that i and h are also tied. The preceding description is for nondirected relations, and it applies in modified form to directed relations. Davis found that triads in data on positive interpersonal affect tend to be transitively closed much more often than could be accounted for by chance, and that this occurs consistently over a large collection of data sets. Of course, in empirically observed social networks transitivity is usually far from perfect, so the tendency towards transitivity is stochastic rather than deterministic.

Davis's finding was based on comparing data with a nontransitive null model. More sophisticated methods along these lines were developed by Holland and Leinhardt (1976), but they remained restricted to the testing of structural characteristics such as transitivity against null models expressing randomness or, in the case of directed graphs, expressing only the tendency toward reciprocation of ties. A next step in modeling is to formulate a stochastic model for networks that expresses transitivity and could be used for statistical analysis of data. Such models have to include one or more parameters indicating the strength of transitivity, and these parameters have to be estimated and tested, controlling for other effects—such as covariate and node-level effects. Then, of course, it would be interesting to model other network effects in addition to transitivity.

The importance of controlling for node-level effects, such as actor attributes, arises because there are several distinct localized social processes that may give rise to transitivity. In the first, social ties may "self-organize" to produce triangular structures, as indicated by the process noted above, that the friends of my friends are likely to become my friends (i.e., a structural balance effect). In other words, the presence of certain ties may induce other ties to form, in this case with the triangulation occurring explicitly as the result of a social process involving three people. Alternatively, certain actors may be very popular, and hence attract ties, including from other popular actors. This process may result in a core-periphery network structure with popular actors in the core. Many triangles are likely to occur in the core as an outcome of tie formation based on popularity. Both of these triangulation effects are structural in outcome, but one represents an explicit social transitivity process whereas the other is the outcome of a popularity process. In the second case, the number of triangles could be accounted for on the basis of the distribution of the actors' degrees without referring to transitivity. In a separate third possibility, however, ties may arise because actors select partners based on attribute homophily, as reviewed in McPherson, Smith-Lovin, and Cook (2001), or some other process of social selection, in which case triangles of similar actors may be a by-product of homophilous dyadic selection processes. An often important question is whether, once accounting for homophily, there are still structural processes present. This would indicate the presence of organizing principles within the network that go beyond dyadic selection. In that case, can we determine whether this self-organization is based within triads, or whether triangulation is the outcome of some other organizing principle? Given the diversity of processes that may lead to transitivity, the complexity of statistical models for transitivity is not surprising.

It can be concluded that transitivity is widely observed in networks. For a full understanding of the processes that give rise to and sustain the network, it is crucial to model transitivity adequately, particularly in the presence of—and controlling for—attributes. In a wide-ranging review, Newman (2003) deplores the inadequacy of existing general network models in this regard. When the requirement is made that the model is tractable for the statistical analysis of empirical data, exponential random graph (or p^*) models offer the most promising framework within which such models can be developed. These models are described in the next section; it will be explained, however, that current specifications of these models often do not provide adequate accounts of empirical data. It is the aim of this paper to present some new specifications for exponential random graph models that considerably extend our capacity to model observed social networks.

1.1. *Exponential Random Graph Models*

The following terms and notation will be used. A *graph* is the mathematical representation of a relation, or a binary network. The number of nodes in the graph is denoted by n. The random variable Y_{ij} indicates whether there exists a tie between nodes i and j ($Y_{ij} = 1$) or not ($Y_{ij} = 0$). We use the convention that there are no self-ties—i.e., $Y_{ii} = 0$ for all i. A random graph is represented by its adjacency matrix Y with elements Y_{ij}. Graphs are by default nondirected (i.e., $Y_{ij} = Y_{ji}$ holds for all i, j), but much attention is given also to directed relations, represented by directed graphs, for which Y_{ij} indicates the existence of a tie *from i to j*, and where Y_{ij} is allowed to differ from Y_{ji}. Denote the set of all adjacency matrices by $\mathcal{Y}$. The notational convention is followed where random variables are denoted by capitals and their outcomes by small letters. We do not consider nonbinary ties here, although they may be considered within this framework (e.g., Snijders and Kenny 1999; Hoff 2003).

A stochastic model expressing transitivity was proposed by Frank and Strauss (1986). According to their definition, a probability distribution for a graph is a Markov graph if the number of nodes is fixed at n and possible edges between disjoint pairs of nodes are independent conditional on the rest of the graph. This can be formulated less compactly, for the case of a nondirected graph: if i, j, u, v are four distinct nodes, the Markov property requires that Y_{ij} and Y_{uv} are independent, conditional on all other variables Y_{ts}. This is an appealing but quite restrictive definition, generalizing the idea of Markovian dependence for random processes with a linearly ordered time parameter and for spatial processes on a lattice (Besag 1974). The basic idea is that two possible social ties are dependent only if a common actor is involved in both. In Section 3.2 we shall discuss the limitations of this dependence assumption in modeling observed social structures.

Frank and Strauss (1986) obtained an important characterization of Markov graphs. They used the assumption of permutation invariance, stating that the distribution remains the same when the nodes are relabeled. Making this assumption and using the Hammersley-Clifford theorem (Besag 1974), they proved that a random graph is a Markov graph if and only if the probability distribution can be written as

$$\mathrm{P}_\theta\{Y=y\} = \exp\left(\sum_{k=1}^{n-1} \theta_k\, S_k(y) + \tau\, T(y) - \psi(\theta,\tau)\right) \quad y \in \mathcal{Y} \tag{1}$$

where the statistics S_k and T are defined by

$$\begin{aligned}
S_1(y) &= \textstyle\sum_{1\le i<j\le n} y_{ij} && \text{number of edges} \\
S_k(y) &= \textstyle\sum_{1\le i\le n} \binom{y_{i+}}{k} && \text{number of } k\text{-stars } (k \ge 2) \\
T(y) &= \textstyle\sum_{1\le i<j<h\le n} y_{ij}\, y_{ih}\, y_{jh} && \text{number of triangles,}
\end{aligned} \tag{2}$$

the Greek letters θ_k and τ indicate parameters of the distribution, and $\psi(\theta, \tau)$ is a normalizing constant ensuring that the probabilities sum to 1. Replacing an index by the + sign denotes summation over the index, so y_{i+} is the degree of node i. A configuration $(i, j_1, \ldots, j_k)$ is called a k-star if i is tied to each of j_1, j_2, up to j_k. For all k, the number of k-stars in which node i is involved, is given by $\binom{y_{i+}}{k}$. An edge is a one-star, so $S_1(y)$ is also equal to the number of one-stars. Some of these configurations are illustrated in Figure 1.

It may be noted that this family of distributions contains for $\theta_2 = \ldots = \theta_{n-1} = \tau = 0$ the trivial case of the Bernoulli graph—i.e., the purely random graph in which all edges occur independently and have the same probability $e^{\theta_1}/(1 + e^{\theta_1})$.

Frank and Strauss (1986) elaborated mainly the three-parameter model where $\theta_3 = \ldots = \theta_{n-1} = 0$, for which the probability distribution depends on the number of edges, the number of two-stars, and the number of transitive triads. They observed that parameter estimation for this model is difficult, and they presented a simulation-based method for the maximum likelihood estimation of any one of the three parameters in this model, given that the other two are fixed at 0, which is only of theoretical value. They also proposed the so-called pseudo-likelihood

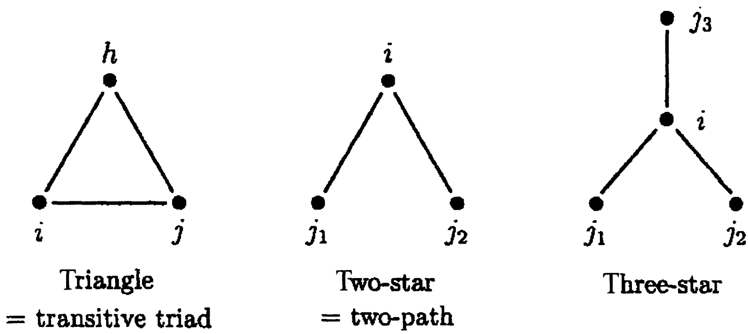

FIGURE 1. Some configurations for nondirected graphs.

estimation method for estimating the complete vector of parameters. This is based on maximizing the pseudo-loglikelihood defined by

$$\ell(\theta) = \sum_{i<j} \ln\Big(\mathrm{P}_\theta\left\{ Y_{ij} = y_{ij} \mid Y_{uv} = y_{uv} \text{ for all } u < v, (u, v) \neq (i, j)\right\}\Big). \tag{3}$$

This method can be carried out relatively easily, as the algorithm is formally equivalent to a logistic regression. However, the properties of pseudo-loglikelihood estimators have not been adequately established for social networks. In analogous situations in spatial statistics, the maximum pseudo-loglikelihood estimator has been observed to overestimate the dependence in situations where the dependence is strong and to perform adequately when the dependence is weak (Geyer and Thompson 1992). For most social networks the dependence is strong and the maximum pseudo-loglikelihood is suspect.

The paper by Frank and Strauss (1986) was seminal and led to many papers published in the 1990s. In the first place, Frank (1991) and Wasserman and Pattison (1996) proposed to use a model of this form, both for nondirected and for directed graphs, with arbitrary statistics $u(y)$ in the exponent. This yields the probability functions

$$\mathrm{P}_\theta\{Y = y\} = \exp\left(\theta' u(y) - \psi(\theta)\right) \quad y \in \mathcal{Y} \tag{4}$$

where y is the adjacency matrix of a graph or digraph and $u(y)$ is any vector of statistics of the graph. Wasserman and Pattison called this family of distributions the p^* model. As this is an example of what statisticians call an exponential family of distributions (e.g., Lehmann 1983) with $u(Y)$ as the sufficient statistic, the family also is called an exponential random graph model (ERGM).

Various extensions of this model to valued and multivariate relations were published (among others, Pattison and Wasserman 1999; Robins, Pattison, and Wasserman 1999), focusing mainly on subgraph counts as the statistics included in $u(y)$, motivated by the Hammersley-Clifford theorem (Besag 1974). To estimate the parameters, the pseudo-likelihood method continued to be used, although it was acknowledged that the usual chi-squared likelihood ratio tests were not warranted here, and there remained uncertainty about the qualities and meaning of the pseudo-likelihood estimator. The concept of Markovian dependence as defined by Frank and Strauss was extended by Pattison and Robins (2002) to partial conditional independence, meaning that whether edges Y_{ij} and Y_{uv} are independent conditionally on the rest of the graph depends not only on whether they share nodes but also on the pattern of ties in the rest of the graph. This concept will be used later in this paper.

Recent developments in general statistical theory suggested Markov chain Monte Carlo (MCMC) procedures both for obtaining simulated draws from ERGMs, and for parameter estimation. MCMC algorithms for maximum likelihood (ML) estimation of the parameters in ERGMs were proposed by Snijders (2002) and Handcock (2003). This method uses a general property of maximum likelihood estimates in exponential families of distributions such as (4). That is to say, the ML estimate is the value $\hat{\theta}$ for which the expected value of the statistics $u(Y)$ is precisely equal to the observed value $u(y)$:

$$E_{\hat{\theta}}u(Y) = u(y). \tag{5}$$

In other words, the parameter estimates require the model to reproduce exactly the observed values of the sufficient statistics $u(y)$.

The MCMC simulation procedure, however, brought to light serious problems in the definition of the model given by (1) and (2). These were discussed by Snijders (2002), Handcock (2002a, 2002b, 2003), and Robins, Pattison, and Woolcock (2005), and they go back to a type of model degeneracy discussed in a more general sense by Strauss (1986). A probability distribution can be termed degenerate if it is concentrated on a small subset of the sample space, and for exponential families this term is used more generally for distributions defined by parameters on the boundary of the parameter space; near degeneracy here is defined by the distribution placing disproportionate probability on a small set of outcomes (Handcock 2003).

A simple instance of the basic problem with these models occurs as follows. If model (1) is specified with only an edge parameter θ_1 and a transitivity parameter τ, while θ_1 has a moderate and τ a sufficiently positive value, then the exponent in (1) is extremely large when y is the complete graph (where all edges are present—i.e., $y_{ij} = 1$ for all i, j) and much smaller for all other graphs that are not almost complete. This difference is so extreme that for positive values of τ—except for quite small positive values—and moderate values of θ_1, the probability is almost 1 that the density of the random graph Y is very close to 1. On the other hand, if τ is fixed at a positive value and the edge parameter θ_1 is decreased to a sufficient extent, a point will be reached where the probability mass moves dramatically from nearly complete graphs to predominantly low density graphs. This model has been studied asymptotically by Jonasson (1999) and Handcock (2002a). If τ is nonnegative, Jonasson shows that asymptotically the model produces only three types of distributions: (1) complete graphs, (2) Bernoulli graphs, and (3) mixture distributions with a probability p of complete graphs and a probability $1 - p$ of Bernoulli graphs. These distributions are not interesting in terms of transitivity. This near-degeneracy is related to the phase transitions known for the Ising and some other models (e.g., Besag 1974; Newman and Barkema 1999). The phase transition was studied for the triangle model by Häggström and Jonasson (1999) and Burda, Jurkiewicz, and Krzywicki (2004), and for the two-star model by Park and Newman (2004).

Some examples of more complex models are given in Sections 4 and 5 below. The phase transition occurs in such models as a near discontinuity of the expected value $\mathrm{E}_\theta u(Y)$ as a function of θ—i.e., as the existence of a value of θ where a plot of coordinates $\mathrm{E}_\theta u_k(Y)$ graphed as a function of the coordinate θ_k (or of other coordinates $\theta_{k'}$) shows a sudden and big increase, or jump (e.g., see, the Figure 16 a). Mathematically, the function still is continuous, but the derivative is extremely large. In many network data sets this increase of $\mathrm{E}_{\theta} u_k(Y)$ jumps right over the observed value $u_k(y)$; and for the parameter value where the jump occurs—which has to be the parameter estimate satisfying the likelihood equation (5)—the probability distribution of $u_k(y)$ has a bimodal shape, reflecting that here the random graph distribution is a mixture of the low-density graphs produced to the left of the jump, and the almost complete graphs produced to its right. Hence, although the parameter estimate does reproduce the observation $u(y)$ as the fitted *expected* value, this expected value is far from the two modes of the fitted distribution. This fitted model does not give a satisfactory representation of the data. Illustrations are given in later sections.

One potential way out of these problems might be to condition on the total number of ties—i.e., to consider only graphs having the observed number of edges. However, Snijders (2002) showed that although conditioning on the total number of ties does sometimes lead to improved parameter estimation, the mentioned problems still occur in more subtle forms, and there still are many data sets for which satisfactory parameter estimates cannot be obtained.

A question, then, must be answered: To what extent does model (1) when applied to empirical data produce parameter estimates that are in, or too close to, the nearly degenerate area, resulting in the impossibility of obtaining satisfactory parameter estimates. A next question is whether a model such as (1) will provide a good fit. Our overall experience is that, although sometimes it is possible to attain parameter estimates that work well, even though they are close to the nearly degenerate area, there are many empirically observed graphs having a moderate or large degree of transitivity and a low to moderate density, which cannot be well represented by a model such as (1), either because no satisfactory parameter estimates can be obtained or because the fitted model does not give a satisfactory representation of the observed network. This model offers little medium ground between a very slight tendency toward transitivity and a distribution that is for all practical purposes concentrated on the complete graph or on more complex "crystalline" structures as demonstrated in Robins, Pattison, and Woolcock (2005).

The present paper aims to extend the scope of modeling social networks using ERGMs by representing transitivity not only by the number of transitive triads but in other ways that are in accordance with the concept of partial conditional independence of Pattison and Robins (2002). We have couched this introduction in terms of the important issue of transitivity, but the modeling of transitivity also requires attention to star parameters, or equivalently, aspects of the degree distribution. New representations for transitivity and the degree distribution in the case of nondirected graphs are presented in Section 3, preceded by a further explanation of simulation methods for the ERGM in Section 2.

After the technical details in Section 3, we present in Section 4 some new modeling possibilities made possible by these specifications, based on simulations, showing that these new specifications push back some of the problems of degeneracy discussed above. In Section 5 the new models are applied to data sets that hitherto have not been amenable to convergent parameter estimation for the ERGM. A similar development for directed relations is given in Section 6.

2. GIBBS SAMPLING AND CHANGE STATISTICS

Exponential random graph distributions can be simulated, and the parameters can be estimated, by MCMC methods as discussed by Snijders (2002) and Handcock (2003). This is implemented in the computer programs SIENA (Snijders et al. 2005) and statnet (Handcock et al. 2005). A straightforward way to generate random samples from such distributions is to use the Gibbs sampler (Geman and Geman 1983): cycle through the set of all random variables Y_{ij} $(i \neq j)$ and simulate each in turn according to the conditional distribution

$$P_\theta\{Y_{ij} = y_{ij} \mid Y_{uv} = y_{uv} \quad \text{for all } (u, v) \neq (i, j)\}. \tag{6}$$

Continuing this procedure a large number of times defines a Markov chain on the space of all adjacency matrices that converges to the desired distribution. Instead of cycling systematically through all elements of the adjacency matrix, another possibility is to select one pair (i, j) randomly under the condition $i \neq j$, and then generate a random value of Y_{ij} according to the conditional distribution (6); this procedure is called mixing (Tierney 1994). Instead of Gibbs steps for stochastically updating the values Y_{ij}, another possibility is to use Metropolis-Hastings steps. These and some other procedures are discussed in Snijders (2002).

For the exponential model (4), the conditional distributions (6) can be obtained as follows, as discussed by Frank (1991) and Wasserman and Pattison (1996). For a given adjacency matrix y, define by $\tilde{y}^{(1)}(i, j)$ and $\tilde{y}^{(0)}(i, j)$, respectively, the adjacency matrices obtained by defining the (i, j) element as $\tilde{y}^{(1)}_{ij}(i, j) = 1$ and $\tilde{y}^{(0)}_{ij}(i, j) = 0$ and leaving all other elements as they are in y, and define the *change statistic* with (i, j) element by

$$z_{ij} = u(\tilde{y}^{(1)}(i, j)) - u(\tilde{y}^{(0)}(i, j)). \tag{7}$$

The conditional distribution (6) is formally given by the logistic regression with the change statistics in the role of independent variables,

$$\mathrm{logit}\big(\mathrm{P}_\theta\left\{Y_{ij} = 1 \mid Y_{uv} = y_{uv} \text{ for all } (u, v) \neq (i, j)\right\}\big) = \theta' z_{ij}. \tag{8}$$

This is also the form used in the pseudo-likelihood estimation procedure, shown in (3).

The change statistic for a particular parameter has an interpretation that is helpful in understanding the implications of the model. When multiplied by the parameter value, it represents the change in log-odds for the presence of the tie due to the effect in question. For instance, in model (1), if an edge being present on (i, j) would thereby form three new triangles, then according to the model the log-odds of that tie being observed would increase by 3τ due to the transitivity effect.

The problems with the exponential random graph distribution discussed in the preceding section reside in the fact that for specifications of the statistic $u(y)$ containing the number of k-stars for $k \geq 2$ or the number of transitive triads, if these statistics have positive parameters, changing some value y_{ij} can lead to large increases in the change statistic for other variables y_{uv}. The change in y_{uv} suggested by these change statistics will even further increase values of other change statistics, and so on, leading to an avalanche of changes which ultimately leads to a complete graph from which the probability of escape is negligible—hence the near degeneracy. Note that this is not intrinsically an algorithmic issue—the algorithm merely reflects the full-conditional probability distributions of the model. The cause is that the underlying model places significant mass on complete (or near complete) graphs. A theoretical analysis of these issues is given by Handcock (2003).

This can be illustrated more specifically by the special case of the Markov model defined by (1) and (2) for nondirected graphs where only edge, two-star, and triangle parameters are present. The change statistic is

$$\begin{pmatrix} z_{1ij} \\ z_{2ij} \\ z_{3ij} \end{pmatrix} = \begin{pmatrix} 1 \\ \tilde{y}^{(0)}_{i+}(i, j) + \tilde{y}^{(0)}_{j+}(i, j) \\ L_{2ij} \end{pmatrix} = \begin{pmatrix} 1 \\ y_{i+} + y_{j+} - 2\, y_{ij} \\ L_{2ij} \end{pmatrix} \tag{9}$$

where $\tilde{y}^{(0)}(i, j)$ denotes, as above, the adjacency matrix obtained from y by letting $\tilde{y}_{ij}^{(0)}(i, j) = 0$ and leaving all other y_{uv} unaffected, and $\tilde{y}_{i+}^{(0)}(i, j)$ and $\tilde{y}_{j+}^{(0)}(i, j)$ are for this reduced graph the degrees of nodes i and j; while L_{2ij} is the number of two-paths connecting i and j,

$$L_{2ij} = \sum_{h \neq i,j} y_{ih}\, y_{hj}. \tag{10}$$

The corresponding parameters are θ_1, θ_2, and τ. The avalanche effect, occurring for positive values of the two-star parameter θ_2 and the transitivity parameter τ, can be understood as follows.

All the change statistics are elementwise nondecreasing functions of the adjacency matrix y. Therefore, given that θ_2 and τ are positive, increasing some element y_{ij} from 0 to 1 will increase many of the change statistics and thereby the logits (8). In successive simulation steps of the Gibbs sampling algorithm, an accidental increase of one element y_{ij} will therefore increase the odds that a next variable y_{uv} will also obtain the value 1, which in the next simulation steps will further increase many of the change statistics, etc., leading to the avalanche effect. Note that the maximum value of z_2 is $2(n-2)$ and the maximum of z_3 is $(n-2)$, both of which increase indefinitely as the number of nodes of the graph increases, and this large maximum value is one of the reasons for the problematic behavior of this model. It may be tempting to reduce this effect by choosing the edge parameter θ_1 strongly negative. However, this forces the model toward the empty graph. If the two forces are balanced, the combined effect is a mixture of (near) empty and (near) full graphs with a paucity of the intermediate graphs that are closer to realistic observations. If the Markov random graph model contains a balanced mixture of positive and negative star parameter values, this avalanche effect can be smaller or even absent. This property is exploited and elaborated in the following section.

3. PROPOSALS FOR NEW SPECIFICATIONS FOR STAR AND TRANSITIVITY EFFECTS

We begin this section by considering proposals that will model all k-star parameters as a function of a single parameter. Since the number of stars is a function of the degrees, this is equivalent to modeling the degree distribution. Suitable functions will ensure that the avalanche effect referred to in the previous section will not occur, or will at least be constrained. These steps can be taken within the framework of the Markov dependence assumption.

We then turn to transitivity, which is more important from a theoretical point of view but is treated after the models for k-stars and the degree distribution because of the greater complexity of the graph structures involved. The model for transitivity uses a new graph configuration that we term a k-triangle. We model k-triangles in similar ways to the stars, in that all k-triangle parameters are modeled as a function of a single parameter. But these new parameters are not encompassed by Markov dependence, and we need to relax the dependence assumption to partial conditional dependence. The discussion is principally for nondirected graphs; the case of directed graphs is presented more briefly in a later section.

3.1. *Geometrically Weighted Degrees and Related Functions*

Expression (1)–(2) shows that the exponent of a Markov graph model can contain an arbitrary linear function of the k-star counts S_k, $k = 1, \dots, n-1$. These counts S_k are given by binomial coefficients, which are independent polynomials of the node degrees y_{i+}, S_k being a polynomial of degree k. But it is known that every function of the numbers 1 through $n-1$ can be expressed as a linear combination of polynomials of degrees $1, \dots, n-1$. Therefore, any function of the degree distribution (i.e., any function of the degrees that does not depend on the node labels) can be represented as a linear combination of the k-star counts $S_1, \dots, S_{n-1}$. In other words, we have complete liberty to include any function of the degree distribution in the exponent of (4) and still remain within the family of Markov graphs.

Saturated models for the degree sequence were discussed by Snijders (1991) and by Snijders and van Duijn (2002). These models have the virtue of giving a perfect fit to the degree distribution and controlling perfectly for the degrees when estimating and testing other parameters, but at the expense of an exceedingly high number of parameters and the impossibility to do more with the degree distribution than describe it. Therefore we do not discuss these models here.

3.1.1. *Geometrically Weighted Degree Counts*

A specification that has been traditional since the original paper by Frank and Strauss (1986) is to use the k-star counts themselves. Such subgraph counts, however, if they have positive weights θ_k in the exponent in (4), are precisely among the villains responsible for the degeneracy that has been plaguing ERGMs, as noted above. One primary difficulty is that the model places high probability on graphs with large degrees. A natural solution is to use a statistic that places decreasing weights on the higher degrees.

An elegant way is to use degree counts with geometrically decreasing weights, as in the definition

$$u_\alpha^{(d)}(y) = \sum_{k=0}^{n-1} e^{-\alpha k} d_k(y) = \sum_{i=1}^{n} e^{-\alpha y_{i+}}, \tag{11}$$

where $d_k(y)$ is the number of nodes with degree k and $\alpha > 0$ is a parameter controlling the geometric rate of decrease in the weights. We refer to α as the *degree weighting parameter*. For large values of α, the contribution of the higher degree nodes is greatly decreased. As $\alpha \to 0$ the statistic places increasing weight on the high degree graphs. This model is clearly a subclass of the model (4) where the vector of statistics is $u(y) = d(y) \equiv (d_0(y), \ldots, d_{n-1}(y))$ but with a parametric constraint on the natural parameters,

$$\theta_k = e^{-\alpha k} \quad k = 1, \ldots, n-1, \tag{12}$$

which may be called the *geometrically decreasing degree distribution assumption*. This model is hence a curved exponential family (Efron 1975). The statistic (11) will be called the geometrically weighted degrees with parameter α.

As the degree distribution is a one-to-one function of the number of k-stars, some additional insight can be gained by considering the equivalent model in terms of k-stars. Define

$$\begin{aligned} u_\lambda^{(s)}(y) &= S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \ldots + (-1)^{n-2}\frac{S_{n-1}}{\lambda^{n-3}} \\ &= \sum_{k=2}^{n-1}(-1)^k \frac{S_k}{\lambda^{k-2}}. \end{aligned} \tag{13}$$

Here the weights have alternating signs, so that positive weights of some k-star counts are balanced by negative weights of other k-star counts. This implies that, when considering graphs with increasingly high degrees, the contribution from extra k-stars is kept in check by the contribution from extra $(k+1)$-stars. Using expression (2) for the number of k-stars and the binomial theorem, we obtain that

$$u_{\lambda}^{(s)}(y) = \lambda^2 u_{\alpha}^{(d)}(y) + 2\lambda S_1 - n\lambda^2 \tag{14}$$

for $\lambda = e^{\alpha}/(e^{\alpha} - 1) \geq 1$; the parameters α and λ are decreasing functions of one another. This shows that the two statistics form the same model in the presence of an edges or 1-star term. This model is also a curved exponential family based on (1), and the constraints on the star parameters can be expressed in terms of the parameter λ as

$$\theta_k = -\theta_{k-1}/\lambda. \tag{15}$$

This equation is equivalent to the geometrically decreasing degree distribution assumption and can, alternatively, be called the *geometric alternating k-star assumption*. Statistic (13) will be called an alternating k-star with parameter λ.

As $\alpha \to \infty$, it follows that $\lambda \to 1$, and (11) approaches

$$u_{\infty}^{(d)}(y) = d_0(y). \tag{16}$$

Thus the boundary case $\alpha = \infty(\lambda = 1)$ implies that the number of isolated nodes is modeled distinctly from other terms in the model. This can be meaningful for two reasons. First, social processes leading to the isolation of some actors in a group may be quite different from the social processes that determine which ties the nonisolated actors have. Second, it is not uncommon that isolated actors are perceived as not being part of the network and are therefore left out of the data analysis. This is usually unfortunate practice. From a dynamic perspective, isolated actors may become connected and other actors may become isolated. To exclude isolated actors in a single network study is to make the implausible presupposition that such effects are not present.

The change statistic associated to statistic (11) is

$$z_{ij} = -\left(1 - e^{-\alpha}\right)\left(e^{-\alpha \tilde{y}_{i+}} + e^{-\alpha \tilde{y}_{j+}}\right) \tag{17}$$

where $\tilde{y} = \tilde{y}^{(0)}(i, j)$ is the reduced graph as defined above. This change statistic is an elementwise nondecreasing function of the adjacency matrix, but the change becomes smaller as the degrees $\tilde{y}_{i+}$ become larger, and for $\alpha > 0$ the change statistic is negative and bounded below by $2(e^{-\alpha} - 1)$. Thus, according to the criterion in Handcock (2003), a full-conditional MCMC for this model will mix close to uniformly. This should help protect such models from the inferential degeneracy that has hindered unconstrained models.

As discussed above, the change statistic aids interpretation. If the parameter value is positive, then we see that the conditional log-odds of a tie on (i, j) is greater among high-degree actors. In a loose sense, this expresses a version of preferential attachment (Albert and Barabási 2002) with ties from low degree to high degree actors being more probable than ties among low degree actors. However, preference for high degree actors is not linear in degree: the marginal gain in log-odds for connections to increasingly higher degree partners is geometrically decreasing with degree.

For instance, if $\alpha = \ln(2)$ (i.e., $\lambda = 2$) in equation (17), for a fixed degree of i, a connection to a partner j_1 who has two other partners is more probable than a connection to j_2 with only one other partner, the difference in the change statistics being 0.25. But if j_1 and j_2 have degrees 5 and 6 respectively (from their ties to others than i), the difference in the change statistics is less than 0.02. So, nodes with degree 5 and higher are treated almost equivalently. Given these two effects – a preference for connection to high degree nodes, and little differentiation among high degree nodes beyond a certain point, we expect to see two differences in outcomes from models with this specification compared to Bernoulli graphs with the same value for θ_1: a tendency for somewhat higher degree nodes, and a tendency for a core-periphery structure.

3.1.2. *Other Functions of Degrees*

Other functions of the node degrees could also be considered. It has been argued recently (for an overview, see Albert and Barabási 2002) that for many phenomena degree frequencies tend to 0 more slowly than exponential functions—for example, as a negative power of the degrees. This suggests sums of reciprocals of degrees, or higher negative powers of degrees, instead of exponential functions such as (14). An alternative specification of a slowly decreasing function that exploits the fact that factorials are recurrent in the combinatorial properties of graphs and that is in line with recent applications of the Yule distribution to degree distributions (see Handcock and Jones 2004), is a sum of ascending factorials of degrees,

$$u(y) = \sum_{i=1}^{n} \frac{1}{(y_{i+} + c)_r} \tag{18}$$

where $(d)_r$ for integers d is Pochhammer's symbol denoting the rising factorial,

$$(d)_r = d(d+1)\dots(d+r-1),$$

and the parameters c and r are natural numbers $(1, 2, \dots)$. The associated change statistic is

$$z_{ij}(y) = \frac{-r}{(\tilde{y}_{i+} + c)_{r+1}} + \frac{-r}{(\tilde{y}_{+j} + c)_{r+1}}. \tag{19}$$

The choice between this statistic and (13), and the choice of the parameters α or λ, c, and r, will depend on considerations of fit to the observed network. Since these statistics are linearly independent for different parameter values, several of them could in principle be included in the model simultaneously (although this will sometimes lead to collinearity-type problems and change the interpretation of the parameters).

3.2. *Modeling Transitivity by Alternating k-Triangles*

The issues of degeneracy discussed above suggest that in many empirical circumstances the Markov random graph model of Frank and Strauss (1986) is too restrictive. Our experience in fitting data suggests that problems particularly occur with Markov models when the observed network includes not just triangles but larger "clique-like" structures that are not complete but do contain many triangles. Each of the three processes discussed in the introduction are likely to result in networks with such denser "clumps." These are indeed the subject of much attention in network analysis (cohesive subset techniques), and the transitivity parameter in Markov models (and perhaps the transitivity concept more generally) can be regarded as the simplest way to examine such clique-like sections of the network because the triangle is the simplest clique that is not just a tie. But the linearity of the triangle count within the exponential is a source of the near-degeneracy problem in Markov models, when observed incomplete cliques are somewhat large and hence contain many triangles. What is needed to capture these "clique-ish" structures is a transitivity-like concept that expresses triangulation also within subsets of nodes larger than three, and with a statistic that is not linear in the triangle count but gives smaller probabilities to large cliquelike structures. Such a concept is proposed in this section.

From the problems associated with degeneracy, given the equivalence between the Markov conditional independence assumption and model (1), we draw two conclusions: (1) edges that do not share a tie may still be conditionally dependent (i.e., the Markov dependence assumption may be too restrictive); (2) the representation of the social phenomenon of transitivity by the total number of triangles is often too simplistic, because the conditional log-odds of a tie between two social actors often will not be simply a linear function of the total number of transitive triangles to which this tie would contribute.

A more general type of dependence is the *partial conditional independence* introduced by Pattison and Robins (2002), a definition that takes into account not only *which nodes* are being potentially tied, but also *the other ties that exist* in the graph—i.e., the dependence model is realization-dependent. We propose a model that satisfies the more general independence concept denoted here as [CD] for "Conditional Dependence."

Assumption [CD]: Two edge indicators Y_{iv} and Y_{uj} are conditionally dependent, given the rest of the graph, only if one of the two following conditions is satisfied:

1. They share a vertex—i.e., $\{i, v\} \cap \{u, j\} \neq \emptyset$ (the usual Markov condition).
2. $y_{iu} = y_{vj} = 1$, i.e., if the edges existed they would be part of a four-cycle (see Figure 2).

This assumption can be phrased equivalently in terms of *in*dependence: If neither of the two conditions is satisfied, then Y_{iv} and Y_{uj} are conditionally independent, given the rest of the graph.

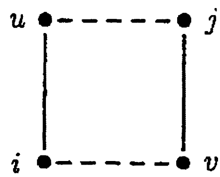

FIGURE 2. Partial conditional dependence when four-cycle is created.

One substantive interpretation of the partial conditional dependence assumption (2) is that the possibility of a four-cycle establishes the structural basis for a "social setting" among four individuals (Pattison and Robins 2002), and that the probability of a dyadic tie between two nodes (here, i and v) is affected not just by the other ties of these nodes but also by other ties within such a social setting, even if they do not directly involve i and v. A four-cycle assumption is a natural extension of modeling based on triangles (three-cycles) and was first used by Lazega and Pattison (1999) in an examination of whether such larger cycles could be observed in an empirical setting to a greater extent than could be accounted for by parameters for configurations involving at most three nodes.

We now seek subgraph counts that can be included among the sufficient statistics $u(y)$ in (4), expressing types of transitivity—therefore including triangles—and leading to graph distributions conforming to assumption [CD]. Under the Markov assumption (1), Y_{iv} is conditionally dependent on each of Y_{iu}, Y_{ij}, and Y_{jv}, because these edge indicators share a node. If $y_{iu} = y_{jv} = 1$, the precondition in the four-cycle partial conditional dependence (2), then Y_{iv} is conditionally dependent also on Y_{uj}, and hence (cf. Pattison and Robins 2002) the Hammersley-Clifford theorem implies that the exponential model (4) could contain the statistic defined as the count of such configurations. We term this configuration, given by

$$y_{iv} = y_{iu} = y_{ij} = y_{uj} = y_{jv} = 1,$$

a two-triangle (see Figure 3). It represents the edge $y_{ij} = 1$ as part of the triadic setting $y_{ij} = y_{iv} = y_{jv} = 1$ as well as the setting $y_{ij} = y_{iu} = y_{ju} = 1$.

Elaborating this approach, we propose a model that satisfies assumption [CD] and is based on a generalization of triadic structures in the form of graph configurations that we term *k-triangles*. It should be

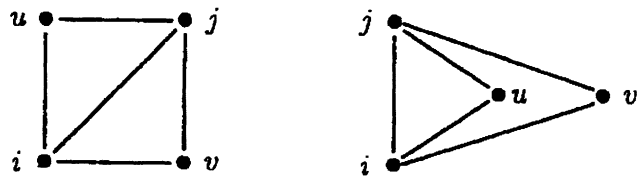

FIGURE 3. Two examples of a two-triangle.

noted that this model implies, but it is not implied by, assumption [CD]: It is a further specification.

For a nondirected graph, a k-triangle with base (i, j) is defined by the presence of a base edge $i - j$ together with the presence of at least k other nodes adjacent to both i and j. We denote a "side" of a k-triangle as any edge that is not the base. The integer k is called the *order of the k-triangle*. Thus a k-triangle is a combination of k individual triangles, each sharing the same edge $i - j$, as shown in Figure 4. The concept of a k-triangle can be seen as a triadic analogue of a k-star. If $k_{\max}$ denotes the highest value of k for which there is a k-triangle on a given base edge (i, j), then the larger $k_{\max}$, the greater the extent to which i and j are adjacent to the same nodes, or alternatively to which i and j share network partners. Because the notion of k-triangles incorporates that of an ordinary triangle ($k = 1$), k-triangle statistics have the potential for a more granulated description of transitivity in social networks. It should be noted that there are inclusion relations between the k-triangles for different k. A three-triangle configuration, for instance, necessarily comprises three two-triangles, so the number of three-triangles cannot be less than thrice the number of two-triangles.

A summary of how dependence structures relate to conditional independence models is given by Robins and Pattison (2005). Here we use the characterization, obtained by Pattison and Robins (2002),

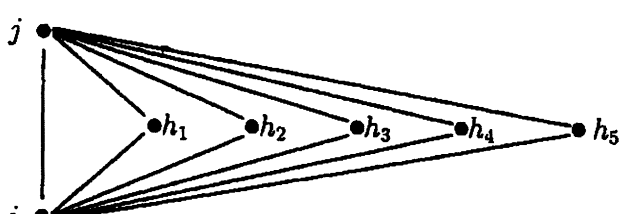

FIGURE 4. A k-triangle for $k = 5$, which is also called a five-triangle.

of the sufficient statistics $u(y)$ in (4) of partial conditionally independent graph models. In the model proposed below, the statistics $u(y)$ contain, in addition to those of the Markov model, parameters for all k-triangles. Such a model satisfies assumption [CD], which can be seen as follows. It was shown already above that this holds for two-triangles. Assuming appropriate graph realizations, [CD] implies that every possible edge in a three-triangle configuration can be conditionally dependent on every other possible edge through one or the other of the two-triangles, and hence as all possible edges are conditionally dependent, it follows from the characterization by Pattison and Robins (2002) that there is a parameter pertaining to the three-triangle in the model. Induction on k shows that the Markovian conditional dependence (1) with the four-cycle partial conditional dependence (2) implies that there can be a parameter in the model for each possible k-triangle configuration.

Our proposed model contains the k-triangle counts, but including these all as separate statistics in the exponent of (4) would lead to a large number of of statistical parameters. Therefore we propose a more parsimonious model specifying relations between their coefficients in this exponent, in much the same way as for alternating k-stars. The model expresses transitivity as the tendency toward a comparatively high number of triangles, without too many high-order k-triangles because this would lead to a (nearly) complete graph. Analogous to the alternating k-stars model, the k-triangle model described below implies a possibly substantial increase in probability for an edge to appear in the graph if it is involved in only one triangle, with further *but smaller* increases in probability as the number of triangles that would be created increases (i.e., as the edge would form k-triangles of higher and higher order). Thus, the increase in probability for creation of a k-triangle is a decreasing function of k. There is a substantively appealing interpretation: If a social tie is not present despite many shared social partners, then there is likely to be a serious impediment to that tie being formed at all (e.g., impediments such as limitations to degrees and to the number of nodes connected together in a very dense cluster, mutual antipathy, or geographic distance, depending on the empirical context). In that case, the addition of even more shared partners is not likely to increase the probability of the tie greatly.

This is expressed mathematically as follows. The number of k-triangles is given by the formula

$$\begin{aligned} T_k &= \sharp\{(\{i,j\},\{h_1,h_2,\dots,h_k\}) \mid \{i,j\} \subset V,\ \{h_1,h_2,\dots,h_k\} \subset V, \\ &\qquad y_{ij} = 1 \text{ and } y_{ih_\ell} = y_{h_\ell j} = 1 \text{ for } \ell = 1,\dots,k\} \\ &= \textstyle\sum_{i<j} y_{ij}\binom{L_{2ij}}{k} \quad (\text{for } k \ge 2), \end{aligned} \tag{20a}$$

where L_{2ij}, defined in (10), is the number of two-paths connecting i and j. (Note that for nondirected graphs, two-paths and two-stars refer to the same configuration $y_{ih} = y_{hj} = 1$; the name "star" points attention to the middle vertex h, the name "path" to the end vertices i and j.) If there exists a tie $i - j$, the value $k = L_{2ij}$ is the maximal order k for which a k-triangle exists on the base (i, j). The formula for the number of triangles, which can be called 1-triangles, is different, due to the symmetry of these configurations:

$$T_1 = \frac{1}{3}\sum_{i<j} y_{ij} L_{2ij}. \tag{20b}$$

We propose a model where these k-triangle counts occur as sufficient statistics in (4), but with weights for the k-triangle counts T_k that have alternating signs and are geometrically decreasing, like those in the alternating k-stars. We start with the 1-triangles—in contrast to (13)—these being the standard type of triangles on which the others are based, with a weight of 3 aimed at obtaining a simple expression in terms of the numbers of shared partners L_{2ij}. This leads to the following statistic. Analogous to the geometrically weighted degree count, an equivalent expression is given using (20) and the binomial formula,

$$\begin{aligned} u_\lambda^{(t)}(y) &= 3T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} - \dots + (-1)^{n-3}\frac{T_{n-2}}{\lambda^{n-3}} \\ &= \sum_{i<j} y_{ij} \sum_{k=1}^{n-2} \left(\frac{-1}{\lambda}\right)^{k-1} \binom{L_{2ij}}{k} \qquad (21a) \\ &= \lambda \sum_{i<j} y_{ij} \left\{1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}}\right\} \\ &= \lambda S_1 - \lambda \sum_{i<j} y_{ij} e^{-\alpha L_{2ij}}, \qquad (21b) \end{aligned}$$

where again $\lambda = e^\alpha/(e^\alpha - 1)$.

Expression (21a) shows that this is a linear function of the k-triangle counts, which is basic to the proof that this statistic satisfies assumption [CD]. As in the case of k-stars, the statistic imposes the constraint $\tau_k = -\tau_{k-1}/\lambda$ $(k \geq 3)$, where τ_k is the parameter pertaining to T_k. The alternating negative weights counteract the tendency to forming big cliquelike clusters that would be inherent in a model with only positive weights for k-triangle counts. Expression (21b) is (for $\alpha > 0$) an increasing function of the numbers L_{2ij} for which there is an edge $i - j$, but it increases very slowly as L_{2ij} gets large. This expresses that the tie $i - j$ has a higher probability accordingly as i and j have more shared partners, but this increase in probability is very small for higher numbers of shared partners.

We propose to use this statistic as a component in the exponential model (4) to express transitivity, with the purpose of providing a model that will be better able than the Markov graph model to represent empirically observed networks. In some cases, this statistic can be used alongside $T = T_1$ in the vector of sufficient statistics, in other cases only (21a) (or, perhaps, only T_1) will be used—depending on how the best fit to the empirical data is achieved and on the possibility of obtaining a nondegenerate model and satisfactory convergence of the estimation algorithm.

The associated change statistic is

$$z_{ij} = \lambda \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{\tilde{L}_{2ij}} \right\} + \sum_h \left\{ y_{ih} y_{jh} \left(1 - \frac{1}{\lambda}\right)^{\tilde{L}_{2ih}} + y_{hi} y_{hj} \left(1 - \frac{1}{\lambda}\right)^{\tilde{L}_{2hj}} \right\}, \tag{22}$$

where $\tilde{L}_{2uv}$ is the number of two-paths connecting nodes u and v in the reduced graph $\tilde{y}$ (where $\tilde{y}_{ij}$ is forced to be 0) for the various nodes u and v.

The change statistic gives a more specific insight into the alternating k-triangle model. Suppose $\lambda = 2$ and the edge $i - j$ is at the base of a k-triangle and consider the first term in the expression above. Then, similarly to the alternating k-stars, the conditional log-odds of the edge being observed does not increase strongly as a function of k for values of k above 4 or 5 (unless perhaps the parameter value is rather large compared to other effects in the model). The model expresses the notion that it is the first one to three shared partners that principally influence transitive closure, with additional partners not substantially increasing the chances of the tie being formed. The second and third terms of the change statistic relate to situations where the tie completes a k-triangle as a side rather than as the base. For example, for the second term, the edge $i - h$ is the base and h is a partner shared with j; the change statistic decreases as a function of the number of two-paths from i to h. This might be interpreted as actor i, already sharing many partners with h, feeling little impetus to establish a new shared partnership with j who is also a partner to h.

As was the case for the alternating k-stars, this statistic is considered for $\lambda \geq 1$, and the downweighting of higher-order k-triangles is greater accordingly as λ is larger. Again, the boundary case $\lambda = 1$ has a special interpretation. For $\lambda = 1$ the statistic is equal to

$$u_1^{(t)}(y) = \sum_{i<j} y_{ij} I\{L_{2ij} \geq 1\}, \tag{23}$$

the number of pairs (i, j) that are directly linked ($y_{ij} = 1$) but also indirectly linked ($y_{ih} = y_{hj} = 1$ for at least one other node h). In this case the change statistic is

$$z_{ij} = I\{\tilde{L}_{2ij} \geq 1\} + \sum_h \{y_{ih} y_{jh} I\{\tilde{L}_{2ih} = 0\} + y_{hi} y_{hj} I\{\tilde{L}_{2hj} = 0\}\}. \tag{24}$$

3.3. *Alternating Independent Two-Paths*

It is tempting to interpret the effect of the alternating k-triangles as an effect for a tie to form on a base, emergent from the various two-paths that constitute the sides. But the change statistic makes clear that formation of alternating k-triangles involves not only the formation of new bases of k-triangles but also new sides of k-triangles, which should be interpreted as contributing to prerequisites for transitive closure rather than as establishing transitive closure. In order to differentiate between these two interpretations, it is necessary to control for the prerequisites for transitive closure—i.e., the number of configurations that would be

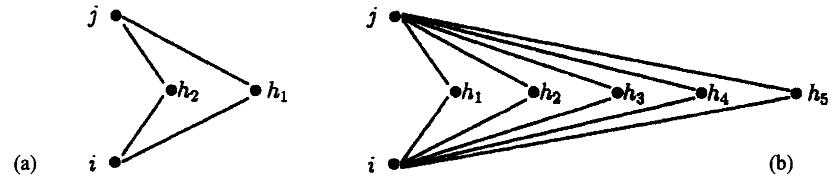

FIGURE 5. Two-independent two-paths (a) and five-independent two-paths (b).

the sides of k-triangles if there would exist a base edge. This means that we consider in addition the effect of connections by two-paths, irrespective of whether the base is present or not. This is precisely analogous in a Markov model to considering both preconditions for triangles—i.e., two-stars or two-paths—and actual triangles. For Markov models, the presence of the two-path effect permits the triangle parameter to be interpreted simply as transitivity rather than a combination of both transitivity and a chance agglomeration of many two-paths. Including the following configuration implies that the same interpretation is valid in our new model.

We introduced k-triangles as an outcome of a four-cycle dependence structure. A four-cycle is a combination of two two-paths. The sides of a k-triangle can be viewed as combinations of four-cycles. More simply, we construe them as independent (the graph-theoretical term for nonintersecting) two-paths connecting two nodes.

Thus, we define k-independent two-paths, illustrated in Figure 5, as configurations $(i, j, h_1, \ldots, h_k)$ where all nodes h_1 to h_k are adjacent to both i and j, irrespective of whether i and j are tied. Their number is expressed by the formula

$$U_k = \sharp\{(\{i,j\}, \{h_1, h_2, \ldots, h_k\}) \mid \{i,j\} \subset V,\ \{h_1, h_2, \ldots, h_k\} \subset V, \\ i \neq j,\ y_{ih_\ell} = y_{h_\ell j} = 1 \quad \text{for } \ell = 1, \ldots, k\}$$

$$= \sum_{i<j} \binom{L_{2ij}}{k} \qquad (\text{for } k \neq 2) \tag{25a}$$

$$U_2 = \frac{1}{2} \sum_{i<j} \binom{L_{2ij}}{2} \qquad (\text{number of four-cycles}); \tag{25b}$$

the specific expression for $k = 2$ is required because of the symmetries involved. The corresponding statistic, given as two equivalent expressions, of which the first one has alternating weights for the counts of independent two-paths while the second has geometrically decreasing weights for the counts of pairs with given numbers of shared partners, is

$$u_\lambda^{\mathrm{p}}(y) = U_1 - \frac{2}{\lambda} U_2 + \sum_{k=3}^{n-2} \left(\frac{-1}{\lambda} \right)^{k-1} U_k \tag{26a}$$

$$= \lambda \sum_{i<j} \left\{ 1 - \left(1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\},$$

$$= \lambda \binom{n}{2} - \sum_{i<j} e^{-\alpha L_{2ij}} \tag{26b}$$

where, in analogy to the statistic for the k-triangles, the extra factor 2 is used for U_2 in (26a) in order for the binomial formula to yield the expression (26b). As before, $\lambda = e^\alpha/(e^\alpha - 1)$.

This is called the *alternating independent two-paths statistic*. The change statistic is

$$z_{ij} = \sum_{h \neq i,j} \left\{ y_{jh} \left(1 - \frac{1}{\lambda} \right)^{\tilde{L}_{2ih}} + y_{hi} \left(1 - \frac{1}{\lambda} \right)^{\tilde{L}_{2hj}} \right\}. \tag{27}$$

As for the alternating k-star and k-triangle statistics, the alternating independent two-paths statistic can be generated by imposing the constraint $\nu_k = -\, \nu_{k-1}/\lambda$, where ν_k is the parameter corresponding to U_k.

For $\lambda = 1$ the statistic reduces to

$$u_1^{\mathrm{p}}(y) = \sum_{i<j} I\{L_{2ij} \geq 1\}, \tag{28}$$

the number of pairs (i, j) that are indirectly connected by at least one two-path. This statistic is counterpart to statistic (23), the number of pairs both directly and indirectly linked. Taken together they assess effects for transitivity in precise analogy with triangles and two-stars for Markov graphs. Since two nodes i and j are at a geodesic distance of two if they are indirectly but not directly linked, the number of nodes at a geodesic distance two is equal to (28) minus (23). The change statistic for $\lambda = 1$ is

$$z_{ij} = \sum_{h \neq i,j} \{y_{jh} I\{\tilde{L}_{2ih} = 0\} + y_{hi} I\{\tilde{L}_{2hj} = 0\}\}. \tag{29}$$

3.4. *Summarizing the Proposed Statistics*

Summarizing the preceding discussion, we propose to model transitivity in networks by exponential random graph models that could contain in the exponent $u(y)$ the following statistics:

1. The total number of edges $S_1(y)$, to reflect the density of the graph; this is superfluous if the analysis is conditional on the total number of edges—and this indeed is our advice.
2. The geometrically weighted degree distributions defined by (11), or equivalently the alternating k-stars (13), for a given suitable value of α or λ, to reflect the distribution of the degrees.
3. Next to, or instead of the alternating k-stars: the number of two-stars $S_2(y)$ or sums of reciprocals or ascending factorials (18); the choice between these degree-dependent statistics will be determined by the resulting fit to the data and the possibility of obtaining satisfactory parameter estimates.
4. The alternating k-triangles (21a) and the alternating independent two-paths (26a), again for a suitable value of λ (which should be the same for the k-triangles and the alternating independent two-paths but may differ from the value used for the alternating k-stars), to reflect transitivity and the preconditions for transitivity.
5. Next to, or instead of, the alternating k-triangles: the triad count $T(y) = T_1(y)$, if a satisfactory estimate can be obtained for the corresponding parameter, and if this yields a better fit as shown from the t-statistic for this parameter.

Of course, actor and dyadic covariate effects can also be added. The choice of suitable values of α and λ depends on the data set. Fitting this model to a few data sets, we had good experience with $\lambda = 2$ or 3 and the corresponding $\alpha = \ln(2)$ or $\ln(1.5)$. In some cases it may be useful to include the statistics for more than one value of λ—for example, $\lambda = 1$ (with the specific interpretations as discussed above) together with $\lambda = 3$. Instead of being determined by trial and error, the parameters λ (or α) can also be estimated from the data, as discussed in Hunter and Handcock (2005).

This specification of the ERGM satisfies the conditional dependence condition [CD]. This dependence extends the classical Markovian dependence in a meaningful way to a dependence within social settings. It should be noted, however, that this type of partial conditional dependence is satisfied by a much wider class of stochastic graph models than the transitivity-based models proposed here. Parsimony of modeling leads to restricting attention primarily to the statistics proposed here. Further modeling experience and theoretical elaboration will have to show to what extent it is desirable to continue modeling by including counts of other higher-order subgraphs, representing more complicated group structures.

4. NEW MODELING POSSIBILITIES WITH THESE SPECIFICATIONS

In this section, we present some results from simulation studies of these new model specifications. This section is far from a complete exploration of the parameter space. It only provides examples of the types of network structures that may emerge from the new specifications. More particularly, it illustrates how the new alternating k-triangle parameterization avoids certain problems with degeneracy that were noted above in regard to Markov random graph models.

We present results for distributions of nondirected graphs of 30 nodes. The simulation procedure is similar to that used in Robins et al. (2005). In summary, we simulate graph distributions using the Metropolis-Hastings algorithm from an arbitrary starting graph, choosing parameter values judiciously to illustrate certain points. Typically we have simulation runs of 50,000, with a burn-in of 10,000, although when MCMC diagnostics indicate that burn-in may not have been achieved we carry out a longer run, sometimes up to half a million iterations.

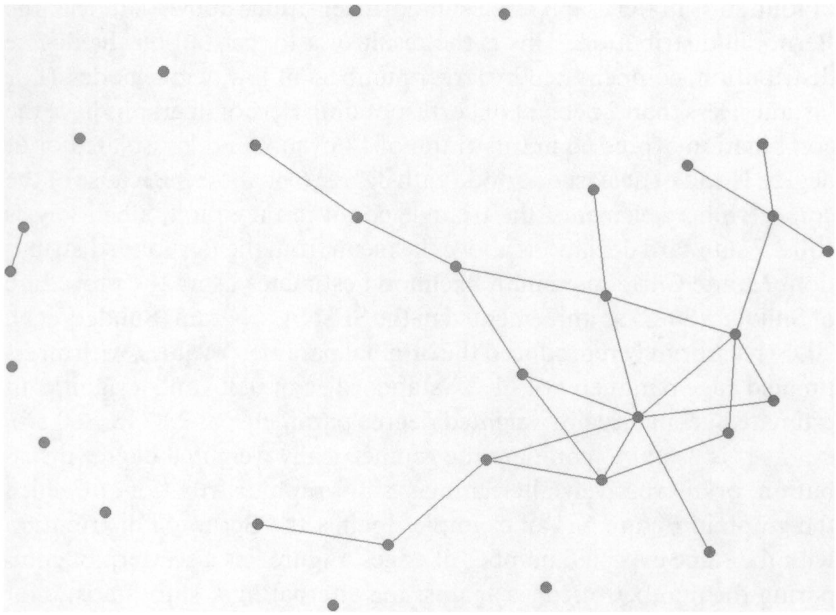

FIGURE 6. A graph from an alternating k-star distribution.

We sample every 100th graph from the simulation, examining graph statistics and geodesic and degree distributions.

4.1. *Geometrically Weighted Degree Distribution*

The graph in Figure 6 is from a distribution obtained by simulating with an edge parameter of -1.7 and a degree weighting parameter (for $\alpha = \ln(2) = 0.693$, corresponding to $\lambda = 2$) of 2.6. This is a low-density graph with 25 edges and a density of 0.06, and in terms of graph statistics is quite typical of graphs in the distribution. Even despite the low density, the graph shows elements of a core-periphery structure, with some relatively high degree nodes (one with degree 7), several isolated nodes, and some low degree nodes with connections into the higher degree "core." What particularly differentiates the graph from a comparable Bernoulli graph distribution with a mean of 25 edges is the number of stars, especially higher order stars. For instance, the number of four-stars in the graph is 3.5 standard deviations above that from the Bernoulli distribution. This is the result of a longer tail on the degree distribution, compensated by larger numbers of low degree nodes. (For instance, less than 2 percent of corresponding Bernoulli graphs have the combination expressed in this graph of 18 or more nodes isolated or of degree 1, and of at least one node with degree 6 or above.) Because of the core-periphery elements, the triangle count in the graph, albeit low, is still 3.7 standard deviations above the mean from the Bernoulli distribution. Monte Carlo maximum likelihood estimates using the procedure of Snijders (2002) as implemented in the SIENA program (Snijders et al. 2005) reassuringly reproduced the original parameter values, with an estimated edge parameter of –1.59 (standard error 0.35) and a significant estimated geometrically weighted degree parameter of 2.87 (S.E. 0.86).

It is useful to compare the geometrically weighted degree distribution, or alternatively alternating k-star graph distribution, of which the graph in Figure 6 is an example, against the Bernoulli distribution with the same expected number of edges. Figure 7 is a scatterplot comparing the number of edges against the alternating k-stars statistic for both distributions. The figure demonstrates a small but discernible difference between the two distributions in terms of the number of k-stars for a given number of edges. There is also a tendency here for greater dispersion of edges and alternating k-stars in the k-star distribution. As with our example graph, in the alternating k-star distributions there are more graphs with high degree nodes, as well as graphs with more low degree nodes.

Finally, in Figure 8, we illustrate the behavior of the model as the alternating k-star parameter increases. The figure plots the mean number of edges for models with an edge parameter of –4.3 and varying alternating k-star parameters, keeping $\lambda = 2$. Equation (13) implies that, as a graph becomes denser, the change statistic for alternating k-stars becomes closer to its constant maximum, so that high-density distributions are very similar to Bernoulli graphs. For an alternating k-star parameter of 1.0 or above, the properties of individual graphs generated within these distributions are difficult to differentiate from realizations of Bernoulli graphs. Even so, the distributions themselves (except those that are extremely dense) tend to exhibit much greater dispersion in graph statistics, including in the number of edges. An important point to note in Figure 8 is that there is a relatively smooth transition from low-density to high-density graphs as the parameter increases,

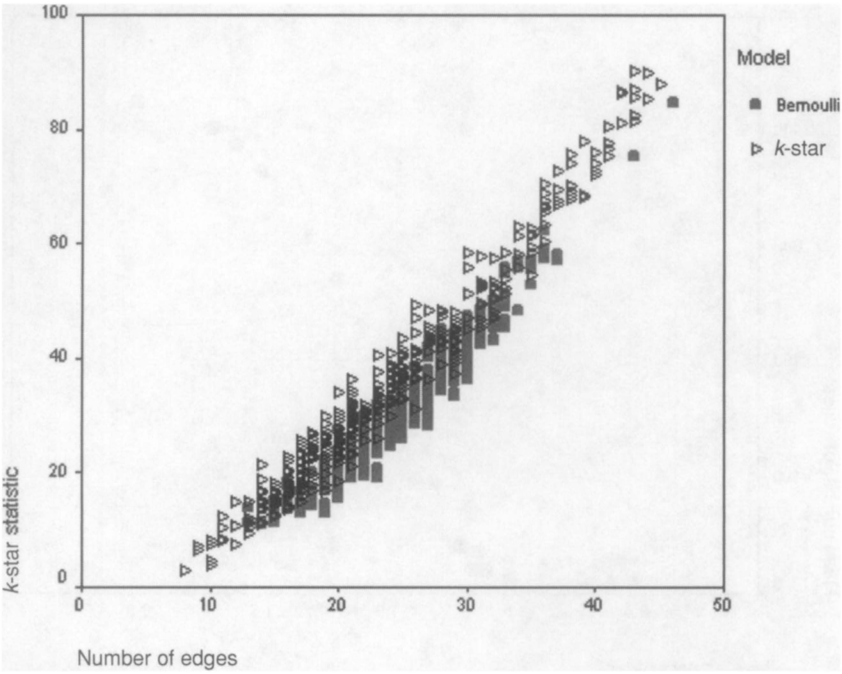

FIGURE 7. Scatterplot of edges against alternating k-stars for Bernoulli and alternating k-star graph distributions.

without the almost discontinuous jumps that betoken degeneracy and are often exhibited in Markov random graph models with positive star parameters.

4.2. *Alternating k-Triangles*

The degeneracy issue for transitivity models and the advance presented by the alternating k-triangle specification are illustrated in Figure 9. This figure depicts the mean number of edges for three transitivity models for various values of a transitivity-related parameter. Each of these models contains a fixed edge parameter, set at –3.0, plus certain other parameters.

The first model (labeled "triangle without star parameters" in the figure) is a Markov model with simply the edge parameter and a triangle parameter. For low values of the triangle parameter, only very

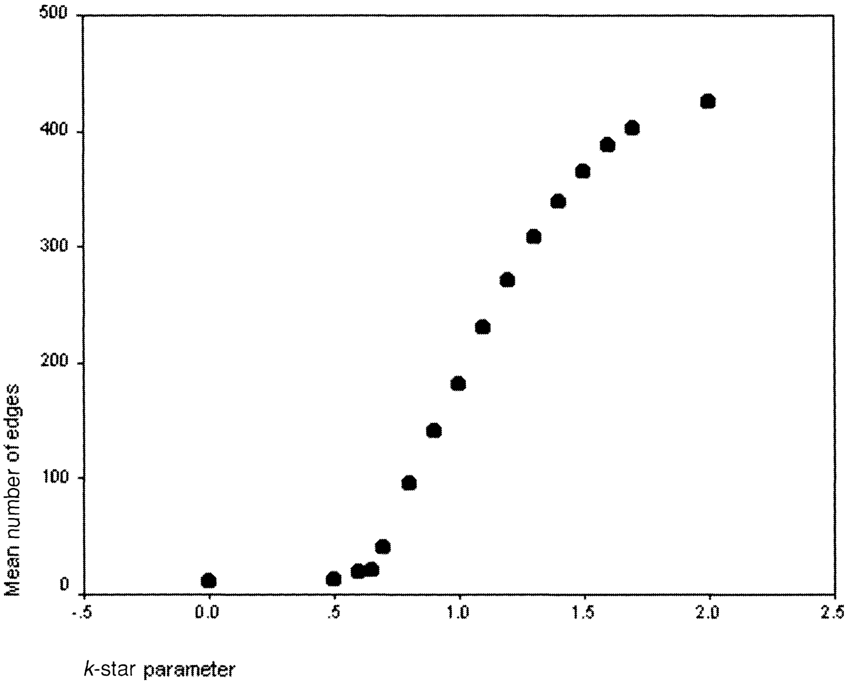

FIGURE 8. Mean number of edges in alternating k-star distributions with different values of the alternating k-star parameter.

low-density graphs are observed; for high values only complete graphs are observed. There is a small region, with a triangle parameter between 0.8 and 0.9, where either a low-density or a complete graph may be the outcome of a particular simulation. This bimodal graph distribution for certain triangle parameter values corresponds to the findings of Jonasson (1999) and Snijders (2002). Clearly, this simple two-parameter model is quite inadequate to model realistic social networks that exhibit transitivity effects.

The second model (labeled "triangle with negative star parameters" in Figure 9) is a Markov model with the inclusion of two- and three-star parameters as recommended by Robins et al. (2005), in particular a positive two-star parameter value (0.5) and a negative three-star parameter value (–0.2), and a triangle parameter with various values. The negative three-star parameter widens the nondegenerate region of the parameter space, by preventing the explosion of edges that leads

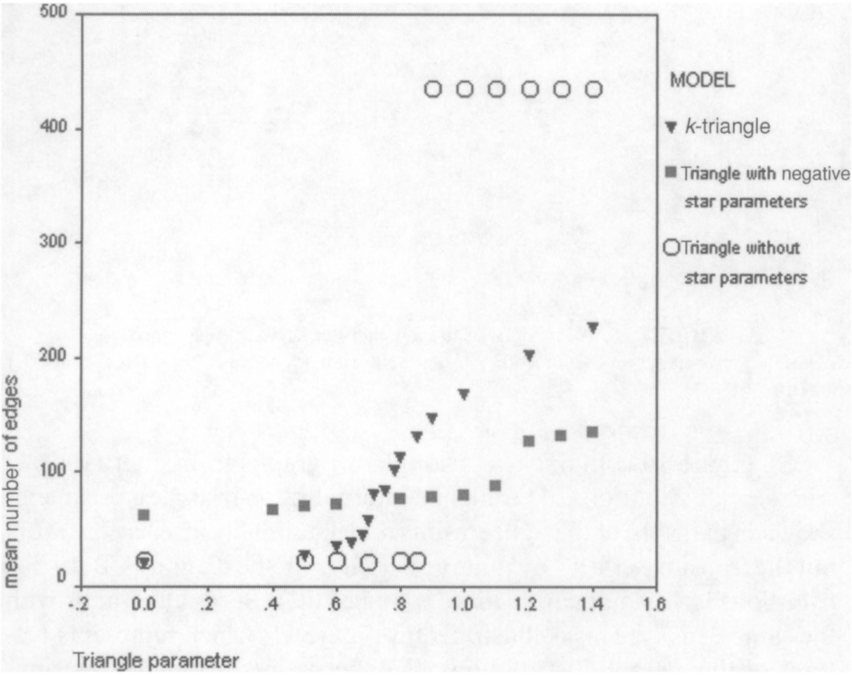

FIGURE 9. Mean number of edges in various graph distributions with different values of a triangle parameter.

to complete graphs. In this example, this works well until the triangle parameter reaches about 1.1. Below this value, the graph distributions are stochastic and of relatively low density, and they tend to have high clustering relative to the number of edges (in comparison to Bernoulli graph distributions). With a triangle parameter above 1.1, however, the graph distribution tends to be "frozen," not on the empty or full graph but on disconnected cliques akin to the caveman graphs of Watts (1999). This area of near degeneracy was observed by Robins et al. (2005).

The third model (labeled "ktriangle" in Figure 9), on the other hand, does not seem to suffer the discontinuous jump, nor the caveman area of near degeneracy, of the first and second models. It is a two-parameter model with an edge parameter and an alternating k-triangles parameter, and the expected density increases smoothly as a function of the latter parameter.

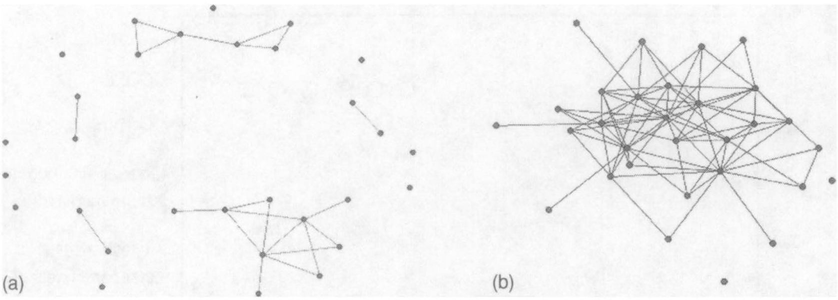

FIGURE 10. A low-density and a higher-density k-triangle graph.
Note: Edge parameter = –3.7 for both; alternating k-triangles parameter = 1.0 for (a) and 1.1 for (b).

Figure 10 contains two examples of graphs from alternating k-triangles distributions. The higher alternating k-triangles parameter shown in panel (b) of the figure results understandably in a denser graph, but the transitive effects are quite apparent from the diagrams. Both distributions have significantly more triangles than Bernoulli graphs with the same density. This is illustrated in Figure 11, which represents features of three graph distributions: the alternating k-triangles distribution of which Figure 10 (b) is a representative (edge parameter = –3.7; alternating k-triangle parameter = 1.1); the Bernoulli graph distribution with mean number of edges identical to this alternating k-triangle distribution (edge parameter = –1.35, resulting in a mean 89.5 edges); and a Markov random graph model with positive two-star, negative three-star, and positive triangle parameters, with parameter values chosen to produce the same mean number of edges (edge parameter = –2.7; two-star parameter = 0.5; three-star parameter = –0.2; triangle parameter = 1.0; mean number of edges = 88.8). We can see from the figure that for the same number of edges the alternating k-triangle distribution is clearly differentiated both from its comparable Bernoulli model as well as the Markov model in having higher numbers of triangles. The Markov model also tends to have more triangles than the Bernoulli model, reflecting its positive triangle parameter.

For an edge-plus-alternating-k-triangle model applied to the graph Figure 10 (a), SIENA produced Monte Carlo maximum likelihood estimates that converged satisfactorily and were consistent with the original parameter values: edge –3.74 (S.E. 0.30), alternating k-triangles 1.06 (S.E. 0.20).

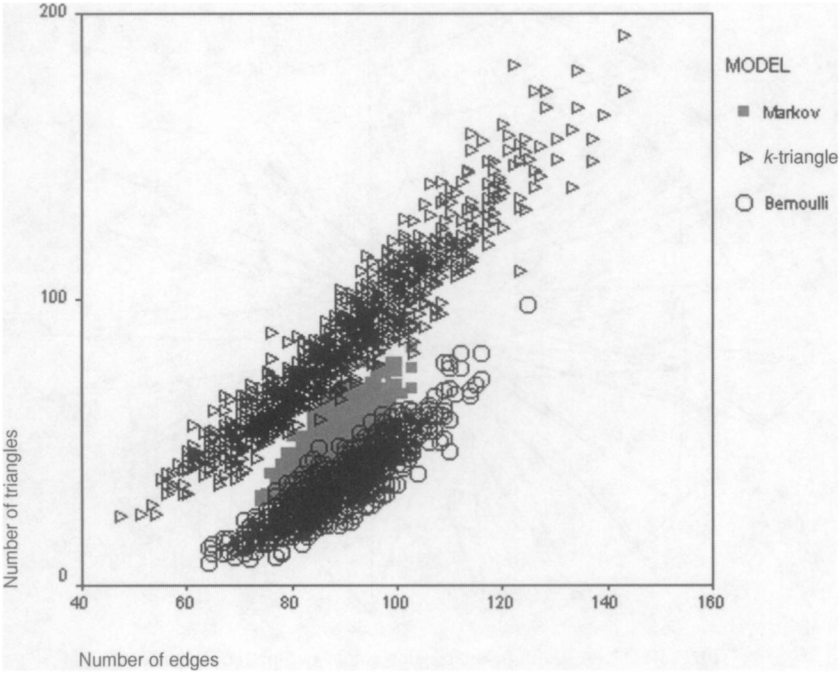

FIGURE 11. Number of triangles against number of edges for three different graph distributions.

Estimates for a Markov model with two-star, three-star, and triangle parameters do exist for this graph (as can be shown using results in Handcock 2003). However it is very difficult to obtain them using SIENA or statnet as the dense core of triangulation produced in graphs from this distribution take us into nearly degenerate regions of the parameter space of Markov models.

4.3. *Independent Two-Paths*

Some of the distinctive features of independent two-path distributions are as follows. A simple way to achieve many independent two-paths is to have cycles through two high-degree nodes. This is what we see in Figure 12, which is a graph from a distribution with edge parameter -3.7 and independent two-paths parameter 0.5. Compared to a Bernoulli graph distribution with the same mean number of edges, this

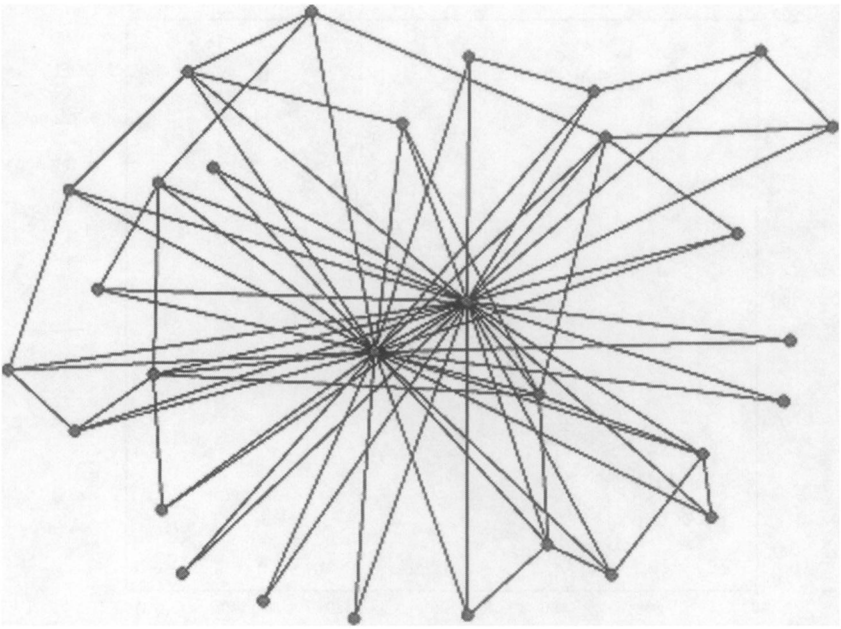

FIGURE 12. A graph from an independent two-path distribution.

graph distribution has substantially more stars, triangles, k-stars, k-triangles, and of course independent two-paths. The graph in Figure 12 is dramatically different from graphs generated under a Bernoulli distribution.

With increasing independent two-paths parameters, the resulting graphs tend to have two centralized nodes, but with more edges among the noncentral nodes. For lower (but positive) independent two-paths parameters, however, only one centralized node appears, resulting in a single starlike structure, with several isolates. We know of no set of Markov graph parameters that can produce such large starlike structures, without conditioning on degrees.

5. EXAMPLE: COLLABORATION BETWEEN LAZEGA'S LAWYERS

Several examples will be presented based on a data collection by Lazega, described extensively in Lazega (2001), on relations between lawyers in a New England law firm (also see Lazega and Pattison 1999). As a first example, the symmetrized collaboration relation was used between the 36 partners in the firm, where a tie is defined to be present if both partners indicate that they collaborate with the other. The average degree is 6.4, the density is 0.18, and degrees range from 0 to 13. Several actor covariates were considered: seniority (rank number of entry in the firm), gender, office (there were three offices in different cities), years in the firm, age, practice (litigation or corporate law), and law school attended (Yale, other Ivy League, or non–Ivy League).

The analysis was meant to determine how this collaboration relation could be explained on the basis of the three structural statistics introduced above (alternating combinations of two-stars, alternating k-stars, and alternating independent two-paths), the more traditional other structural statistics (counts of k-stars and triangles), and the covariates. For the covariates X with values x_i, two types of effect were considered as components of the statistic $u(y)$ in the exponent of the probability function. The first is the main effect, represented by the statistic

$$\sum_i x_i y_{i+}.$$

A positive parameter for this model component indicates that actors i high on X have a higher tendency to make ties to others, which will contribute to a positive correlation between X and the degrees. This main effect was considered for the numerical and dichotomous covariates. The second is the similarity effect. For numerical covariates such as age and seniority, this was represented by the statistic

$$\sum_{i,j} \mathrm{sim}_{ij}\, y_{ij} \tag{30}$$

where the dyadic similarity variable sim_{ij} is defined as

$$\mathrm{sim}_{ij} = 1 - \frac{|\, x_i - x_j \,|}{d_x^{\max}},$$

with $d_x^{\max} = \max_{i,j} |x_i - x_j\,|$ being the maximal difference on variable X. The similarity effect for the categorical covariates, office and law school, was represented similarly using for sim_{ij} the indicator function $I\{x_i = x_j\}$ defined as 1 if $x_i = x_j$ and 0 otherwise. A positive parameter for the similarity effect reflects that actors who are similar on X have a higher tendency to be collaborating, which will contribute to a positive network autocorrelation of X.

The estimations were carried out using the SIENA program (Snijders et al. 2005), version 2.1, implementing the Metropolis-Hastings algorithm for generating draws from the exponential random graph distribution, and the stochastic approximation algorithm described in Snijders (2002). Since this is a stochastic algorithm, as is any MCMC algorithm, the results will be slightly different, depending on the starting values of the estimates and the random number streams of the algorithm. Checks were made for the stability of the algorithm by making independent restarts, and these yielded practically the same outcomes. The program contains a convergence check (indicated in the program as "Phase 3"): after the estimates have been obtained, a large number of Metropolis-Hastings steps is made with these parameter values, and it is checked if the average of the statistics $u(Y)$ calculated for the generated graphs (with much thinning to obtain approximately independent draws) is indeed very close to the observed values of the statistics. Only results are reported for which this stochastic algorithm converged well, as reflected by t-statistics less than 0.1 in absolute value for the deviations between all components of the observed $u(y)$ and the average of the simulations, which are the estimated expected values $\mathrm{E}_{\hat{\theta}} u(Y)$ (cf. (5) and also equation (34) in Snijders 2002).

The estimation kept the total number of ties fixed at the observed value, which implies that there is a not a separate parameter for this statistic. This conditioning on the observed number of ties is helpful for the convergence of the algorithm (for the example reported here, however, good convergence was obtained also without this conditioning). Effects were tested using the t-ratios defined as parameter estimate divided by standard error, and referring these to an approximating standard normal distribution as the null distribution. The effects are considered to be significant at approximately the level of $\alpha = 0.05$ when the absolute value of the t-ratio exceeds 2.

Some explorative model fits were carried out, and it turned out that of the covariates, the important effects are the main effects of seniority and practice, and the similarity effects of gender, office, and practice. In Model 1 of Table 1, estimation results are presented for a model that contains the three structural effects: (1) geometrically weighted degrees for $\alpha = \ln(1.5) = 0.405$ (corresponding to alternating combinations of

TABLE 1
MCMC Parameter Estimates for the Symmetrized Collaboration Relation Among Lazega's Lawyers

	Model 1		Model 2	
Parameter	Est.	S.E.	Est.	S.E.
Geometrically weighted degrees, $\alpha = \ln(1.5)$	−0.711	2.986	—	—
Alternating k-triangles, $\lambda = 3$	0.588	0.184	0.610	0.094
Alternating independent two-paths, $\lambda = 3$	−0.030	0.155	—	—
Number of pairs directly and indirectly connected	0.430	0.512	—	—
Number of pairs indirectly connected	−0.014	0.184	—	—
Seniority main effect	0.023	0.006	0.024	0.006
Practice (corporate law) main effect	0.383	0.111	0.373	0.109
Same practice	0.377	0.103	0.382	0.095
Same gender	0.336	0.124	0.354	0.116
Same office	0.569	0.105	0.567	0.103

two-stars for $\lambda = 3$), (2) alternating k-stars and (3) alternating independent two-paths, both for parameter $\lambda = 3$ and in addition, the last two of these effects for parameter $\lambda = 1$; as indicated above, the latter effects are equal to the number of pairs of nodes both directly and indirectly connected, and the number of pairs indirectly connected.

The results show that none of the structural effects except the alternating k-triangles has a t-ratio greater than 2. There is quite some collinearity between these effects. For example, the estimated correlation between the estimate for the alternating independent two-paths and that for the number of pairs indirectly connected is −0.94. All of the retained covariate effects have t-ratios larger than 2. With a backward selection procedure, nonsignificant effects were stepwise deleted from the model. The result is presented as Model 2 in Table 1. The only remaining structural effect is the alternating k-triangles effect ($\hat{\theta}_j = 0.610$, $t = 6.5$). This indicates that there is strong evidence for transitivity, as represented by the k-triangles effect, and not for any other structural effects except what is already represented by the covariates.

It appears that this model is successful in also representing other structural characteristics of this network, such as the numbers of two-, three-, and four-stars and the number of triangles. The observed number of triangles is 120, while simulations of the ERGM (carried out also using the SIENA program) show that the expected number of triangles under Model 2 is 128.5 with standard deviation 13.2. So the difference is less than 1 standard deviation. Fitting the model extended with the number of triangles did indeed lead to a nonsignificant effect for the number of triangles. The observed number of four-stars is 6091, and under Model 2 the expected number of four-stars is equal to 6635 and the standard deviation is 1042; so here also the difference between observed and expected value is less than 1 standard deviation. Thus, even though these statistics are not directly fitted, the representation of the network structure by the alternating k-triangles together with the covariate effects also gives an adequate representation of these graph statistics.

The estimates for the covariates are hardly different from those in Model 1. Note that these can be interpreted as estimates of the covariate effects on a log-odds scale, similar to effects in logistic regression models, except that they are controlled for the structural effects. This implies that $\exp(\theta_j d)$ is the multiplicative effect that a difference d on variable j has on the estimated odds of a tie. Seniority ranges from 1 to 36, so the more senior partners collaborate more with others, the odds ratio for the greatest difference of 35 being $\exp(0.024 \times 35) = 2.32$. Corporate lawyers have an odds of collaboration that is $\exp(0.373) = 1.45$ higher than those doing litigation, and having the same specialty makes collaboration $\exp(0.382) = 1.47$ more likely. The odds ratio related to having the same gender is $\exp(0.354) = 1.42$, and that related to working in the same office is $\exp(0.567) = 1.76$. All these odds ratios are controlled for the structural transitivity effect. Summarizing, there are especially large effects of seniority and of working in the same office, and slightly smaller but still large effects of doing corporate law, having the same specialty, and having the same gender; in addition, there is a strong transitivity effect.

The latter effect, represented by the alternating k-triangles, can be interpreted as evidence that there are organizing principles in this network that go beyond homophilous selection in creating triangles. The nonsignificance of the weighted degrees effect suggests that there are no other important effects distinguishing the partners in their level of collaborative activity beside the effects of seniority and specialty; and the significance of the k-triangle effect while controlling for the weighted degrees effect indicates that the transitivity is not the result of popularity selection effects alone. In this law firm, it seems that collaborative structures arise not just because of lawyers' personal backgrounds; nor do they arise because of popular collaborators attracting less popular followers; rather, next to the covariate effects, there is a distinct balance effect of self-organizing team formation, resulting in close-knit transitive structures. This conclusion is in line with the conclusions obtained in earlier analyses of this data set (see Lazega 2001), but the results are not directly comparable because analyses including effects of covariates as well as structural transitivity effects were not published before.

The good fit in the sense of good reproduction of a variety of other network statistics is not strongly dependent on the value of the parameter λ. Values $\lambda = 2, 4$, and 5 also yielded good results. The values $\lambda = 1$ and 6 were not satisfactory in this sense.

Model specifications containing the number of two-stars and of transitive triangles also yielded convergence of the estimation algorithm, but it did not succeed well in reproducing the observed number of pairs of nodes tied directly and indirectly; this implies that the number of pairs at a geodesic distance equal to two was not reproduced adequately. Thus, we can conclude that this example is a case where the traditional Markov random graph model for transitivity can be practically applied but that our new model specification yields a better fit to the data.

5.1. *Parameter Sensitivity in Various Models*

To illustrate the differences between the various model specifications and the difficulties of some specifications, we present some simulation results where a parameter is varied in fitted models. We contrast the specification based on the number of triangles to the one using the alternating k-triangles. In addition, we compare models with and without conditioning on the total number of edges.

A model similar to Model 2 in Table 1 was fitted to this data set, and then a long sequence of graphs was simulated by the MCMC procedure, starting with the empty graph, where all parameters except one kept their fixed value, and one designated parameter slowly increased from a low to a high value, and then decreased again to the low value, with 40,000 MCMC iteration steps for each value of this parameter. The designated parameter was the one representing transitivity, being the number of triangles, or the alternating k-triangles. The figures present the generated values of the associated statistic after the 40,000 iterations for each single parameter value.

Figure 13 gives the generated values for the model without conditioning on the number of edges (which for the number of triangles

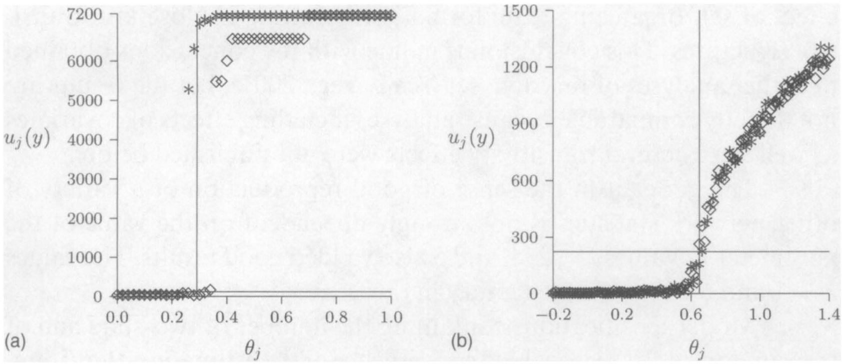

FIGURE 13. Generated statistics $u_j(y)$ for unconditional models, as a function of triangle parameter (a) and alternating k-triangles parameter (b).
(Symbol ◇ indicates simulated values generated for increasing parameter values, $*$ those generated for decreasing values.)

did not lead to satisfactory convergence in the algorithm for parameter estimation but was used anyhow). A vertical line is plotted at the estimated parameter value, and a horizontal line at the observed value of the statistic. This implies that the curve of expected values should exactly go through the intersection of these two lines.

For the number of triangles, an almost discontinuous jump is observed, exactly at the intersection point that was the target for the estimation procedure. In Figure 13 (a), for the increasing parameter values, the jump up is made at a somewhat higher parameter value than the jump down for the decreasing parameter values. This path-dependence, or hysteresis, was also observed in Snijders (2002, p. 9), and it is well-known for Ising models (Newman and Barkema 1999), which show a similar kind of degeneracy. In a small interval of parameter values where this jump occurs, the distribution of the statistic (and of the graph density) has a bimodal shape. The suddenness of the jump, and the fact that the observed statistic is in the region of the jump, is associated with the large practical difficulties in fitting this model to realistic data.

For the alternating k-triangles in Figure 13 (b), a smoother sequence of values is obtained, but the slope is still quite large, especially near the parameter value of about 0.6 where the simulated values "take off" from their starting values close to 0. This is similar to what is shown for the expected values of the density in Figures 8 and 9.

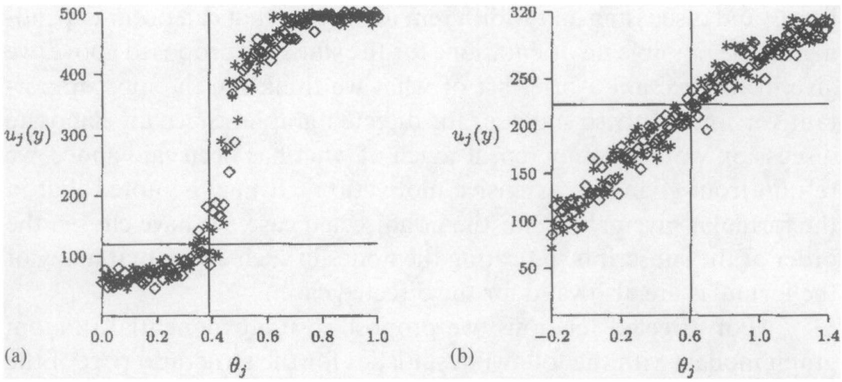

FIGURE 14. Generated statistics $u_j(y)$ for conditional models, as a function of triangle parameter (a) and alternating k-triangles parameter (b).
(Symbol ◇ indicates values generated for increasing parameter values, $*$ those generated for decreasing values.)

Analogous simulated values, but for models conditioning on the observed number of edges, are given in Figure 14. Here we see in both cases a much smoother process. The hysteresis effect for the number of triangles is not observed any more, suggesting good mixing of the MCMC procedure, but this model still has quite a strong slope near the observed value of the statistic. In Figure 14 (b), which combines conditional estimation with the alternating k-triangles statistic, the generated values form a smooth pattern that confirms that estimation can proceed smoothly for this model. The figures together illustrate that both working with the newly proposed statistics rather than with the number of triangles and conditioning on the number of edges contribute to better possibilities for parameter estimation. Combining both elements yields especially good results.

6. DIRECTED RELATIONS

Directed relations are perhaps more frequent in social network research than nondirected relations. For directed relations, the requirement $Y_{ij} = Y_{ji}$ for the adjacency matrix is dropped, and the oriented nature of ties is reflected by using the term "arcs" rather than "edges." Except for the number of ties, all statistics discussed above have multiple analogues for the directed case, depending on the orientation of the ties. Rather than listing and discussing all the different analogues that can occur depending on the possible tie orientations for the statistics proposed above, we give in this section a brief list of what we think are the most important versions of these statistics for directed graphs. Since an elaborate discussion would mainly repeat much of what has been said above, we refrain from giving an extensive motivation. (It may be noted that in the formulas given above for the nondirected case, we have chosen the order of the subscripts indicating the nodes in such a way that many of the formulas are also valid for the directed case.)

For directed relations, we propose to use exponential random graph models with the following statistics for the structural part of the model.

1. The total number of arcs

$$\sum_{i \neq j} Y_{ij} \; ;$$

this is a superfluous element of the sufficient statistics if the analysis is done (as advised) conditional on the number of arcs.

2. The number of mutual dyads

$$\sum_{i<j} Y_{ij} \, Y_{ji} \, .$$

3. Geometrically weighted out-degrees

$$u_\alpha^{(\mathrm{od})}(y) = \sum_{k=0}^{n-1} e^{-\alpha k} d_k^{(\mathrm{out})}(y) = \sum_{i=1}^{n} e^{-\alpha y_{i+}} , \tag{31}$$

and geometrically weighted in-degrees

$$u_\alpha^{(\mathrm{id})}(y) = \sum_{k=0}^{n-1} e^{-\alpha k} d_k^{(\mathrm{in})}(y) = \sum_{i=1}^{n} e^{-\alpha y_{+i}} , \tag{32}$$

where $d_k^{(\mathrm{out})}(y)$ and $d_k^{(\mathrm{in})}(y)$, respectively, are the numbers of nodes with out-degrees or in-degrees equal to k. Similar to (14), these statistics can also be expressed as alternating out-k-star combinations and alternating in-k-star combinations.

4. Next to, or instead of, the alternating k-star combinations: the number of in-two-stars and the number of out-two-stars, reflecting the in-degree and out-degree variances.
5. The alternating transitive k-triangles defined by

$$\lambda \sum_{i,j} y_{ij} \left\{ 1 - \left(1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\} \tag{33}$$

and the alternating independent two-paths defined by

$$\lambda \sum_{i,j} \left\{ 1 - \left(1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\}, \tag{34}$$

where L_{2ij} is still the number of two-paths defined by (10); the orientations implied by these formulas are illustrated in Figure 15.
6. Next to, or instead of the alternating transitive k-triangles: the count of transitive triads

$$\sum_{i,j,h} Y_{ij} Y_{jh} Y_{ih}$$

and the number of two-paths, the latter reflecting the covariance between in-degrees and out-degrees.

The change statistic for the geometrically weighted out-degrees is

$$z_{ij} = -\left(1 - e^{-\alpha}\right) e^{-\alpha \tilde{y}_{i+}}. \tag{35}$$

The change statistic is still given for the alternating transitive k-triangles by (22), and for the alternating independent two-paths by (27).

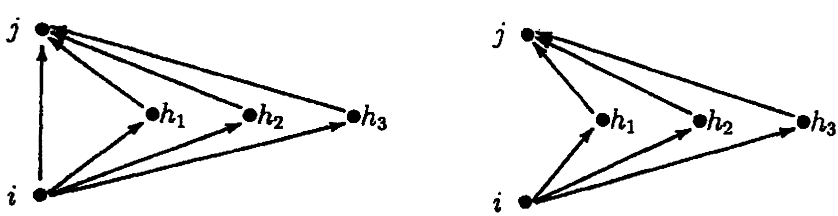

FIGURE 15. Transitive three-triangle (a) and three-independent two-paths (b).

6.1. *Example: Friendship Between Lazega's Lawyers*

As an example of modeling a directed relation, we use the friendship relation between the 36 partners in the law firm studied by Lazega (2001). This is a network with density 0.21 and average degree 7.4. In-degrees vary from 2 to 16, out-degrees from 0 to 21. The larger variability of the degrees and skewness of the distribution of the out-degrees indicates that here it may be more difficult to obtain a well-fitting model than in the preceding example. The same covariates are used as in the earlier example. For effects of actor-level covariates X on directed relations, instead of the main effect we distinguish between (1) the activity effect, represented in $u(y)$ by the statistic

$$\sum_i x_i y_{i+},$$

for which a positive parameter will tend to increase the correlation between the covariate and the out-degrees; and (2) the popularity effect, represented by

$$\sum_i x_i y_{+i},$$

which contributes to the correlation between the covariate and the in-degrees. The similarity effect connected to an actor-level covariate is defined as for the undirected case.

Preliminary analyses showed that the most important effects of covariates are the similarity effect of working in the same office, and the effects associated with seniority (rank number of entry in the firm) and practice (litigation versus corporate law). The same procedure for estimation was used as in the preceding example. A forward selection procedure using the effects listed above, with estimation conditional on the total number of ties, led to the results presented in Table 2. Model 1 contains, next to various covariate effects, the four structural effects proposed above. This appeared not to give a good fit with respect to the number of out-k-stars for $k = 2, 3, 4$. Therefore the model was extended with the the numbers of in-two-stars, out-two-stars, and two-paths. This means that the observed covariance matrix of the in- and out-degrees is fitted exactly. The results are presented in Table 2 as Model 2.

TABLE 2
MCMC Parameter Estimates for the Friendship Relation Between Lazega's Lawyers

	Model 1		Model 2	
Parameter	Est.	S.E.	Est.	S.E.
Mutual dyads	1.659	0.241	2.217	0.303
Out-two-stars	—	—	0.129	0.016
In-two-stars	—	—	0.147	0.026
Two-paths	—	—	−0.089	0.020
Geometrically weighted out-degrees, $\alpha = \ln(2)$	0.956	1.185	−1.364	1.361
Geometrically weighted in-degrees, $\alpha = \ln(2)$	−4.550	2.249	−8.367	3.893
Alternating transitive k-triangles, $\lambda = 2$	0.665	0.136	0.709	0.146
Alternating independent two-paths, $\lambda = 2$	−0.139	0.031	−0.068	0.036
Same office	0.570	0.121	0.839	0.180
Seniority popularity	−0.003	0.008	0.001	0.007
Seniority activity	0.013	0.007	0.010	0.006
Seniority similarity	0.038	0.008	0.041	0.008
Practice (corporate law) popularity	−0.049	0.159	0.066	0.122
Practice (corporate law) activity	0.320	0.134	0.205	0.104
Same practice	0.283	0.130	0.291	0.125

The estimation procedure for both models presented in this table converged well, but these results were obtained only after repeated runs of the estimation algorithm, always using the previously obtained results as the initial values for the new estimation. In this case, without conditioning on the total number of ties it was not possible to obtain convergence of the estimation algorithm. Model specifications including the total number of transitive triplets as a separate statistic did not lead to converging estimates. The strong correlations between the structural statistics lead to very strongly correlated parameter estimates, so that for a good reproduction of the model actually more decimal places for the parameter estimates are required than given in Table 2 (cf. Snijders 2002, p. 32). The parameter estimates for the geometrically weighted out-degrees and in-degrees, and the alternating independent two-paths effects differ strongly in Models 1 and 2, due to the inclusion in Model 2 of the out-two-stars, in-two-stars, and two-paths effects. There is a strong transitivity effect, represented by alternating transitive k-triangles with $\lambda = 2$ ($t = 0.709/0.146 = 4.8$ in Model 2). Further, there is evidence that—controlling for these structural effects—friendship is more likely between partners working in the same office, between those similar in seniority (this can be interpreted in part as a cohort effect), and those with the same practice. Those doing corporate law mention more friends than those doing litigation. The other covariate effects are not significant. Other effects, such as the number of transitive triplets and the numbers of three-stars and four-stars, also are represented adequately, each with a difference between observed and estimated expected value of less than 1.5 standard deviation.

A positive effect of alternating k-triangles in the presence of a negative independent two-paths effect suggests that the friendship network tends to be cliquelike, with possibly several different denser clusters of friends. Because the geometrically weighted in-degree parameters are negative, high in-degrees and high order in-stars are less likely in this network, unless of course they are implied by the transitive structure and therefore are involved in cliques of friends. So popular friends tend to be popular within clusters of dense friendships rather than between clusters.

Compared with the collaboration relation, modeling the friendship relation requires a much more complicated structural model. This network is an example where modeling transitivity by the number of transitive triplets by itself is not successful, whereas modeling transitivity by the alternating transitive k-triangles is successful and also does provide a good fit for the number of transitive triplets.

6.2. *Parameter Sensitivity in Two Models*

For this model also, the sensitivity of generated statistics to the parameter representing transitivity can yield insights in the possibility of modeling by using a particular model specification. Two models were considered, both with conditioning on the observed number of ties: the model in Table 2 and the corresponding model with the four new statistics replaced by the number of transitive triplets. For the latter model the parameter estimation did not converge satisfactorily, but the obtained parameter values were used anyway. Figure 16 gives simulated statistics for a continuous MCMC chain of graphs, where all parameter values were fixed at these estimated values, except for the number of transitive triplets (Figure 16 a) and the alternating k-triangles (Figure 16 b), which started at a very low value, increased in little steps to a very high value, and then decreased again to the low value. Again, 40,000 iteration steps

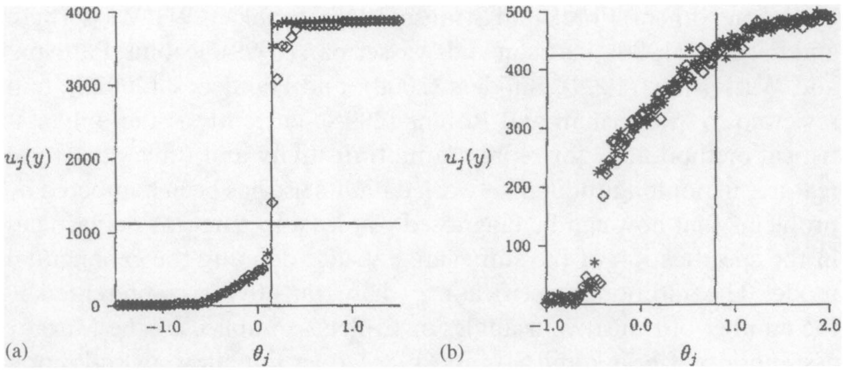

FIGURE 16. Generated statistics $u_j(y)$ for two models, as a function of a transitive triplets parameter (a) and an alternating k-triangles para-meter (b).
(Symbol ◇ indicates values generated for increasing parameter values, $*$ those generated for decreasing values.)

of the MCMC algorithm were made for each parameter value and the resulting value of the statistic corresponding to the changing parameter is represented in the figure.

For the transitive triplets model, there is a very strong jump that occurs right at the observed value of the statistic, indicating the impossibility to adequately model this data set using this particular model—even when conditioning on the observed number of ties. A hysteresis effect can again be discerned (for iteration runs of more than 40,000, this would decrease and eventually disappear). For the alternating k-triangles model Figure 16 (b), there is a jump but it occurs in a region representing antitransitivity (at a negative parameter value of about −0.5), and does not invalidate the application of this model for this data set. The generated statistics for this model vary stochastically about a smooth function of this parameter in a wide region comprising the observed value of the statistic, which corresponds to the possibility to indeed obtain adequate parameter estimates for this specification.

7. DISCUSSION

The methodology based on exponential random graph models, also called p^* models, of which the principles were introduced and elaborated by (among others) Frank and Strauss (1986), Frank (1991), Wasserman and Pattison (1996), Pattison and Wasserman (1999), Robins, Pattison, and Wasserman (1999), Snijders (2002), and Handcock (2002b), and reviewed in Wasserman and Robins (2005), currently is the only statistical methodology for representing transitivity and other structural features in nonlongitudinal network data. Its use has been hampered by problems that now can be diagnosed—at least in part—as deficiencies in the specification of the sufficient statistics defining the exponential model. The traditional specification, where transitivity is represented by the number of transitive triangles or triplets as implied by the Markov assumption of Frank and Strauss (1986), does not allow a good representation of the quite strong but far from complete tendency to transitivity that is commonly observed in social networks. A symptom of these problems is the difficulty to find maximum likelihood estimates, but the near degeneracy and poor fit of the implied model is the more fundamental issue.

In this paper we have proposed a new specification of network statistics defining the exponential random graph model. The new statistics are geometrically weighted degrees to represent degree heterogeneity; alternating k-triangles to represent transitivity; and alternating independent two-paths to represent the preconditions for transitive configurations. Section 3.4 summarizes for nondirected graphs the resulting approach to model specification, and Section 6 gives the similar approach for directed graphs. The new statistics are defined in such a way that the model specification based on them does not imply, for moderately positive values of the parameter representing transitivity, the drive toward complete graphs that is inherent in the traditional specification. Therefore the new specification may be expected to avoid the large degeneracy problems associated with the traditional specification. This is related to the fact that the new specification does not satisfy the Markov conditional independence assumption of Frank and Strauss (1986), which often is too stringent. The new specification satisfies only a weaker type of partial conditional independence defined here as assumption [CD], a specification of dependence concepts discussed in Pattison and Robins (2002). However, the new statistics are not a panacea: in the new models phase changes occur less often but they are not completely excluded, as shown in the near discontinuity in Figure 16 (b). In this figure, however, the near discontinuity occurs at a negative value of the transitivity parameter, which in sociological applications is a relatively unimportant part of the parameter space. Our experience up to now, illustrated by examples presented above, suggests that the models defined by these statistics can be used to give a good representation of the transitivity and degree heterogeneity in many observed social networks, also for data sets in which modeling on the basis of the traditional specification, using only stars and the number of transitive triplets or triangles, was not feasible. Further work is in progress that confirms the wider modeling possibilities opened up by the new specifications. Obtaining maximum likelihood estimates under these specifications by MCMC algorithms is relatively uncomplicated for many data sets. The readily available computer programs SIENA (Snijders et al. 2005) and statnet (Handcock et al. 2005) can be used for this purpose.

The statistics proposed here may look rather contrived at first reading, but they are nevertheless a means to express that regions of incomplete cliquelike structures will occur in social networks, but that these cannot be expressed merely by noting that the network contains many triangles. The k-triangle parameters should therefore be interpreted somewhat differently from the traditional transitivity parameters based on only the triangle count; in our first example we argued that the estimates give evidence for the emergence of close-knit transitive structures that might be interpretable as self-organizing team formation. Such a "higher order" interpretation is not really available in models containing only Markov transitivity based completely on the prevalence of single triangles in the network. The greater difficulty in interpretation of the new statistics, as compared to the traditional specification, seems unavoidable to us, given the complexity of empirically observed social networks. Further experience with this model will be conducive to enhancing the interpretability of the parameters.

Other network statistics might also be possible to achieve these modeling possibilities. We do think that any statistics that achieve this will have some arbitrary elements, or seem contrived to some degree. This may have to do with the fact that network structure, as observed at one given moment, in most cases is the result of many different forces and mechanisms that operated in a period—often of long duration—before the observation of the network, precluding a simple representation of the dependencies within the network. One of the arguments supporting the statistics proposed here is that they lead to a model satisfying the comparatively simple conditional dependence assumption [CD].

The statistics contain parameters (α or λ) that are supposed to be fixed in the treatment given here, and they can be estimated in practice by trying out some reasonable values. In the examples discussed earlier, the sensitivity of the conclusions for the precise value of this parameter was small. However, it is preferable to estimate this parameter statistically. The methods to do this are presented in Hunter and Handcock (2005).

Depending on the observed data set, it is still possible that degeneracy problems occur even with these new specifications. More experience with these models, and the further development of new models, is required before a satisfactory and well-balanced methodology for the statistical modeling of networks will be attained. We think that the representation of social settings in particulars (Pattison and Robins 2002) will need more attention and will be possible by incorporating extra model elements. In some cases such extra model elements will be individual and dyadic covariates, or interactions of covariates with the structures proposed above, which is easy to carry out as it remains within the framework of the exponential random graph model. In other cases the model would have to be compounded with additional elements, such as latent structure models like the Euclidean models of Hoff, Raftery, and Handcock (2002), the ultrametric models of Schweinberger and Snijders (2003), or the model-based clustering version of Tantrum, Handcock, and Raftery (2005); in still other cases, in the exponential model one could include other complicated statistics in addition to those proposed here, along the lines of Pattison and Robins (2002).

REFERENCES

Albert, Réka, and Albert-László Barabási. 2002. "Statistical Mechanics of Complex Networks." *Review of Modern Physics* 74:47–97.

Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society*, Series B 36:192–225.

Boer, Peter, Mark Huisman, Tom A. B. Snijders, and Evelien P. H. Zeggelink. 2003. *StOCNET: An Open Software System for the Advanced Statistical Analysis of Social Networks. Version 1.4*. Groningen: ProGAMMA/ICS.

Burda, Zdzislaw, Jerzy Jurkiewicz, and André Krzywicki. 2004." Network Transitivity and Matrix Models." *Physical Review* E69, 026106.

Davis, James A. 1970. "Clustering and Hierarchy in Interpersonal Relations: Testing Two Theoretical Models on 742 Sociograms." *American Sociological Review* 35:843–52.

Davis, James A., and Samuel Leinhardt. 1972. "The Structure of Positive Interpersonal Relations in Small Groups." Pp. 218–51 in *Sociological Theories in Progress*, vol. 2., edited by Joseph Berger. Boston: Houghton-Mifflin.

Efron, Bradley. 1975. "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *Annals of Statistics* 3:1189–242.

Frank, Ove. 1991. "Statistical Analysis of Change in Networks." *Statistica Neerlandica* 45:283–93.

Frank, Ove, and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81:832–42.

Geman, Stuart, and Donald Geman. 1983. "Stochastic Relaxation, gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transaction on Pattern Analysis and Machine Intelligence* 6:721–41.

Geyer, Charles J., and Elizabeth A. Thompson. 1992. "Constrained Monte Carlo maximum likehood calculations (with discussion)." *Journal of the Royal Statistical Society,* Series C, 54:657–99.

Häggström, Olle, and Johan Jonasson. 1999. "Phase Transition in the Random Triangle Model." *Journal of Applied Probability* 36:1101–15.

Handcock, Mark S. 2002a. "Specifying Models for Social Networks." Presented at Contact Networks and Transmission Dynamics of STDs, Centers for Disease Control and Prevention, National Center for HIV, STD, and TB Prevention Meeting, April 17–20, Washington, DC.

———. 2002b. "Statistical Models for Social Networks: Inference and Degeneracy." Pp. 229–40 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by Ronald Breiger, Kathleen Carley, and Philippa E. Pattison. Washington, DC: National Academies Press/National Research Council of the National Academies.

———. 2003. "Assessing Degeneracy in Statistical Models of Social Networks." Working Paper No. 39, Center for Statistics and the Social Sciences, University of Washington, Seattle.

Handcock, Mark S., and James H. Jones. 2004. "Likelihood-Based Inference for Stochastic Models of Sexual Network Formation." *Theoretical Population Biology* 65:413–22.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2005. statnet*: An* R *package for the Statistical Analysis and Simulation of Social Networks*. Manual, version 1.0. Center for Statistics and the Social Sciences, University of Washington, Seattle.

Hoff, Peter D. 2003. "Bilinear Mixed Effects Models for Dyadic Data." Working Paper No. 32, Center for Statistics and the Social Sciences, University of Washington, Seattle.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97:1090–98.

Holland, Paul W., and Samuel Leinhardt. 1975. "Local Structure in Social Networks." Pp. 1–45 in *Sociololgical Methodology*, vol. 6, edited by David R. Heise, Boston, MA: Blackwell Publishing.

Hunter, David R., and Mark S. Handcock. (Forthcoming). "Inference in Curved Exponential Family Models for Networks." *Journal of Graphical and Computational Statistics.*

Jonasson, Johan. 1999. "The Random Triangle Model." *Journal of Applied Probability* 36:852–67.

Lazega, Emmanuel. 2001. *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership.* Oxford, England: Oxford University Press.

Lazega, Emmanuel, and Philippa E. Pattison. 1999. "Multiplexity, Generalized Exchange and Cooperation in Organizations: A Case Study." *Social Networks* 21:67–90.

Lehmann, Erich L. 1983. *Theory of Point Estimation.* New York: Wiley.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Newman, Mark E. J. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45:167–256.

Newman, Mark E. J., and Gerard T. Barkema. 1999. *Monte Carlo Methods in Statistical Physics.* Oxford, England: Clarendon Press.

Park, Juyong, and Mark E. J. Newman. 2004. "Solution of the 2-star Model of a Network." *Condensed Matter Abstracts*, cond-mat/0405457.

Pattison, Philippa E., and Garry L. Robins. 2002. "Neighbourhood Based Models for Social Networks." Pp. 301–37 in *Sociological Methodology*, vol. 22, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Pattison, Philippa E., and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology* 52:169–93.

Robins, Garry L., Philippa E. Pattison, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: III. Valued Relations." *Psychometrika* 64:371–94.

Robins, Garry L., and Philippa E. Pattison. 2005. "Interdependencies and Social Processes: Generalized Dependence Structures." Pp. 192–214 in *Models and Methods in Social Network Analysis*, edited by Peter Carrington, John Scott, and Stanley Wasserman. Cambridge, England: Cambridge University Press.

Robins, Garry L., Philippa E. Pattison, and Jodie Woolcock. 2005. "Small and Other Worlds: Global Network Structures from Local Processes." *American Journal of Sociology* 110:894–936.

Schweinberger, Michael, and Tom A. B. Snijders. 2003. "Settings in Social Networks: A Measurement Model." Pp. 307–41 in *Sociological Methodology*, vol. 23, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Snijders, Tom A. B. 1991. "Enumeration and Simulation Methods for 0-1 Matrices with Given Marginals." *Psychometrika* 56:397–417.

———. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure*, 3.2. Web journal available from http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html.

Snijders, Tom A. B., Christian E. G. Steglich, Michael Schweinberger, and Mark Huisman. 2005. *Manual for SIENA version 2.1*. Groningen, Netherlands: ICS.

Snijders, Tom A. B., and David A. Kenny. 1999. "The Social Relations Model for Family Data: A Multilevel Approach." *Personal Relations* 6:471–86.

Snijders, Tom A. B., and Marijtje A. J. van Duijn. 2002. "Conditional Maximum Likelihood Estimation under Various Specifications of Exponential Random Graph Models." Pp. 117–34 in *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank*, edited by Jan Hagberg. Stockholm, Sweden: University of Stockholm, Department of Statistics.

Strauss, David. 1986. "On a General Class of Models for Interaction." *SIAM Review* 28:513–27.

Tantrum, Jeremy, Mark S. Handcock, and Adrian E. Raftery. 2005. "Model Based Clustering for Social Networks." Working Paper No. 46, Center for Statistics and the Social Sciences, University of Washington, Seattle.

Tierney, Luke. 1994. "Markov Chains for Exploring Posterior Distributions" (with discussion). *Annals of Statistics* 22:1701–62.

Wasserman, Stanley, and Philippa E. Pattison. 1996. "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*." *Psychometrika* 61:401–25.

Wasserman, Stanley, and Garry L. Robins. 2005. "An Introduction to Random Graphs, Dependence Graphs, and p^*." Pp. 148–61 in *Models and Methods in Social Network Analysis*, edited by Peter Carrington, John Scott, and Stanley Wasserman. Cambridge, England: Cambridge University Press.

Watts, Duncan J. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105:493–527.