Exponential Random Graph (*p) Models for Social Networks**

Author:
Garry Robins,
School of Behavioural Science,
University of Melbourne,
Australia

Article outline:

Exponential random graph models, also known as *p** models, constitute a family of statistical models for social networks. The importance of this modeling framework lies in its capacity to represent social structural effects commonly observed in many human social networks, including general degree-based effects as well as reciprocity and transitivity, and at the node-level, homophily and attribute-based activity and popularity effects. The models can be derived from explicit hypotheses about dependencies among network ties. They are parameterized in terms of the prevalence of small subgraphs (configurations) in the network and can be interpreted as describing the combinations of local social processes from which a given network emerges. The models are estimable from data and readily simulated. Versions of the models have been proposed for univariate and multivariate networks, valued networks, bipartite graphs and for longitudinal network data. Nodal attribute data can be incorporated in social selection models, and through an analogous framework for social influence models.

The modeling approach was first proposed in the statistical literature in the mid-1980s, building on previous work in the spatial statistics and statistical mechanics literature. In the 1990s, the models were picked up and extended by the social networks research community. In this century, with the development of effective estimation and simulation procedures, there has been a growing understanding of certain inadequacies in the

original form of the models. Recently developed specifications for these models have shown a substantial improvement in fitting real social network data, to the point where for many network data sets a large number of graph features can be successfully reproduced by the fitted models.

Glossary:

Alternating independent-2-paths: a parameter (and statistic) in new specification models; a particular combination of *k-independent-2-path* counts into the one statistic;

Alternating k-stars: a Markov parameter (and statistic) in the new specification models; a particular combination of Markov *k*-star counts into the one statistic; equivalent to *geometrically weighted degree counts*; useful for modeling the degree distribution.

Alternating k-triangles: a parameter (and statistic) in the new specification models; a particular combination of *k-triangle* counts into the one statistic; equivalent to *weighted shared partners.*

Cyclic triad: a Markov graph configuration: in a directed network, ties *ij, jk* and *ki* are observed among actors *i, j,* and *k*.

Degeneracy (or near-degeneracy): when a model implies that very few distinct graphs are probable, often only empty or complete graphs; degenerate models cannot be good models for social network data.

Dependence assumption: Theoretical assumption about dependencies among possible network ties; determines the type of parameters in the model.

Dyad independence: assumes that dyads are independent of one another; the model includes edge and reciprocity parameters, and possibly also node or dyad attributes.

Dyad-wise shared partners: a parameter (and statistic) in the higher order models; equivalent to *alternating independent 2-paths*.

Edge-wise shared partner distribution: Distribution of the number of dyads who are themselves related and who have a fixed number of shared partners.

Edge-wise shared partners: a parameter (and statistic) in the higher order models; equivalent to *alternating k-triangles*.

Geometrically weighted degree counts: a statistic (and parameter) in the new specification models: a sum of degree counts with geometrically decreasing weights; equivalent to *alternating k-stars.*

Homogeneity assumption: Assumption about which parameters to equate, to make a model identifiable.

k-independent-2-paths: configurations in the higher order models; equivalent to *k*-triangles but without the base.

k-in-star: a Markov graph configuration: in a directed graph, *k* arcs are directed to the one actor.

k-out-star: a Markov graph configuration: in a directed graph, *k* arcs are expressed by the one actor.

k-triangle: a configuration in higher order models; in a non-directed graph, the combination of *k* triangles, each sharing the one edge (the base of the *k*-triangle).

k-star: a Markov graph configuration: in a non-directed graph, *k* edges are expressed by the one actor.

Markov dependence assumption: introduced by Frank and Strauss (1986), proposes that, conditional on the rest of the graph, two possible ties are independent of each other unless they share an actor.

Mixed-star: a Markov graph configuration: a two path in a directed graph.

Monte Carlo Markov Chain maximum likelihood estimation (MCMCMLE): Method of estimation based on computer simulation; more principled than pseudolikelihood.

Network Configuration: A small subgraph that may be observed in the data and that is represented by parameters in the model: e.g. reciprocated ties, triangles.

Parameters: relate to specific network configurations that may be observed in the graph; a large positive parameter is interpreted as the presence of more of the configurations than might be expected from chance (given the other effects in the model); a large negative parameter signifies the relative absence of the configuration.

Partial dependence assumption: assumption for dependencies among possible ties created by the presence of other ties; permits models with higher order configurations than Markov configurations.

Pseudo-likelihood estimation: an approximate method of estimation using logistic regression; does not produce reliable standard errors.

p_1 *models:* an early dyad independence model, including popularity and expansiveness effects.

p_2 *model:* elaboration of $p1$ model, where popularity and expansiveness effects are random, and independent variables may be used to predict ties.

Simple random graphs, Bernoulli graphs, Erdos-Renyi graphs: assume that edges are independent of one another and are observed with a given probability.

Social circuit dependence: Two possible ties are conditionally dependent when, if observed, they would create a 4-cycle.

Transitive triad: a Markov graph configuration: in a directed network, ties *ij, jk* and *ik* are observed among actors *i, j,* and *k.*

Triangle: a Markov graph configuration: in a non-directed network, a clique of three actors, ties *ij, jk* and *ik* are observed among actors *i, j,* and *k*.

I. Definition:

Exponential random graph models, also known as p^* models, constitute a family of statistical models for social networks. These models take the form of a probability distribution of graphs:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp\{\boldsymbol{\eta}' \mathbf{g}\}$$

for a set of tie indicator variables $\mathbf{X}$ on a network of fixed node size n, where $\mathbf{x}$ is a realization, with a parameter vector $\boldsymbol{\eta}$ and a vector of network statistics $\mathbf{g}$. Each value of the parameter vector corresponds to a probability distribution on the set of all graphs with n nodes.

II. Introduction:

As noted in [33], statistical approaches to social networks have quite a long history, stretching at least back to the work of Moreno in the 1930s [24]. Yet, that history is rather sparsely scattered across various literatures and different eras and, as a result, even today

we see that older techniques and approaches are reinvented and repackaged as new. The *statistical modeling* of a social network – as distinct from the use of statistically-based approaches to understanding particular properties of a social network – has a shorter and somewhat more coherent pedigree.

Criteria for a successful statistical model of a social network have been proposed by [33]: it should be possible to estimate the parameters of the model from data in a principled way, and the fitted model should be a good representation of that data; the model should provide theoretically plausible interpretations about the type of effects that might have produced the network; and using the estimated model parameters it should be possible to draw inferences about competing explanations for the data. Most, perhaps all, models for social networks fall short of completely meeting these requirements. Some models are "thought experiments", intended to illuminate but with an uncertain link to data. Some models are estimable from data but still cannot adequately represent important features of the network. Models may imply interpretations that are just not theoretically plausible in terms of social science; and some models cannot include different effects simultaneously in order to test one against the other. Of course, by definition all models are imperfect (else, they are not "models"). So these criteria direct the aim in model development, even under the knowledge that we cannot always hit the target exactly.

A major insight in the statistical modeling of social network structure has been that structural effects can be detected as some form of deviation from what would be expected as "randomness". This was in fact the original proposal of Moreno and colleagues [24] who were the first to compare observed network data to what would be expected from null distributions. More explicitly, Rapoport's biased net theory [31, 32] proposed certain structural "biases" away from random tie formation, including the important notions of transitivity (later to be described as clustering), the propensity for human social networks to exhibit a high proportion of triangulated ties.

Of course, in the graph theory literature the best-known model for randomness is the *simple random graph model*, often known as the Erdös-Rényi graph [7] or uniform Bernoulli graph distribution. This model proposes that for a given number of nodes network ties are observed between pairs of nodes independently and with a fixed probability p. Properties of this model have been examined extensively, utilizing its analytic tractability. It has also been used as a null model for various sampling distributions (e.g. see the summary in [8].) For an observed network, p can be readily estimated (the maximum likelihood estimate is simply the density of the network), but unfortunately this model is not a good representation of almost any human social network. There are no structural effects here, only randomness.

For directed networks, [15] extended the Bernoulli model in the early 1980s by parameterizing structural effects for reciprocity and for differential node-level activity (out-degree) and popularity (in-degree). Possible network ties within dyads were dependent on one another, but were independent between dyads. This dyad-independence assumption permitted the model to be estimated as a straight-forward loglinear model. Holland and Leinhardt called this model p_1, the subscripted '1' implying a program of

further research, with progressively enlarged dependence assumptions (within arcs, within dyads, within triads, etc.) However, Holland and Leinhardt were uncertain how to progress beyond dyads. The problem was that standard statistical estimation required some level of independence and triads, unlike dyads, overlapped. (A subsequent, more sophisticated extension, the p_2 model, also has dyadic independence at its heart, but conditional on random node-level effects, so that the random effects indirectly introduce dependencies that may extend beyond dyads [52]).

Ove Frank and his colleague, David Strauss, provided the crucial insight: in the world of complex networks, it was dependence that mattered, and assumptions involving traditional statistical independence, although helpful for fitting models, were likely to be inadequate empirically [9]. In contrast, their approach centered on *conditional* independence, whereby contingencies among network ties may transmit across the network in the sense that the presence of one network tie may affect the presence of any other, but most pairs of ties were in some sense "remote", so that the contingencies affecting a given network tie had to be transmitted through "neighboring" ties. So once one had a concept of "neighboring ties", that is, some notion of which possible ties were conditionally independent after taking into account other observed network ties, the form of a model could be specified. Approaches from spatial statistics and statistical mechanics, with notions of dependence within neighborhoods (broadly defined) could be translated in this way into social network models. Frank and Strauss proposed Markov random graph models, based on an explicit argument about conditional independence among networks ties, that could be regarded as theoretically plausible for the first time. These models incorporated reciprocity, degree-based effects and various forms of triangulation or clustering.

In the 1990s, Markov random graph models became the accepted form of exponential random graph models and they have only recently been superseded. They could be estimated through the rough and ready procedure of pseudo-likelihood [50], a not particularly sophisticated approach but enough to get the field started empirically. Wasserman and Pattison popularized these models among social networks researchers as p^* models [54] and added further generalizations, with extensions to multivariate [28] and valued networks [38]. Nodal attributes were introduced in social selection [35] and social influence models [36].

In the last decade, especially with more principled methods of estimation and simulation, it has been shown that homogeneous Markov random graph models face considerable, often insurmountable, difficulties in dealing with real network data. Data with inhomogeneities – for instance, very high degree nodes, or dense regions of multiple triangulation, both of which are not uncommon in real social networks – typically result in *degeneracy* for Markov random graph models, where no set of parameter estimates can adequately represent the data. Newer specifications have been proposed by Snijders and colleagues [47] that help considerably with these issues of degeneracy, and have the capacity to enhance our ability to model small-scale to medium-sized social networks, in some cases remarkably well. These advances are based on new theorizations of

dependence among network ties, echoing the original project of Holland and Leinhardt with a notion of increasingly extended dependence assumptions.

This article begins with a presentation of notation and terminology and then discusses how dependence hypotheses lead to the general form of the model. Specific examples of dependence hypotheses, and the resulting models, are then presented: Bernoulli graph distributions; dyadic independence models; and Markov random graph models. Methods of simulation are briefly introduced before a discussion of degeneracy issues, particularly as they apply to Markov random graph models. Recently proposed dependence hypotheses are introduced, including the so-called "social circuit" dependence, leading to additional specifications for exponential random graph models that substantially improve model performance. Estimation and goodness of fit approaches are discussed before a short empirical example is presented. Extensions are briefly described followed by some concluding remarks about future directions.

III. Notation and terminology

A network comprises a set of relational ties between pairs of individual *actors* (be they people or other social entities). A network can be represented as a graph G with node set $N = \{1, 2, \ldots, n\}$ representing the individuals and edge set E representing the relational ties. For the statistical models of this article, relational ties are construed as a set of binary random variables X_{ij} such that $X_{ij} = 1$ if a tie is observed from node i to node j, for $i \neq j$, and $X_{ij} = 0$, otherwise. For a nondirected network X_{ij} and X_{ji} are equivalent, whereas for a directed network the two variables are distinct. X_{ii} is undefined (or may be considered as a structural zero.) A nondirected tie is an *edge* and a directed tie an *arc*. We specify x_{ij} as the observed value of the variable X_{ij} and we let $\mathbf{X}$ be the matrix of all variables with $\mathbf{x}$ being a realization (where the diagonals of the matrices are forced to be zero.) $\mathbf{X}$ and $\mathbf{x}$ are necessarily symmetric for nondirected networks, but not so for directed networks. More generally, $\mathbf{x}$ may be valued but in this article we restrict attention to binary ties.

There are natural extensions to this basic notation, depending on the data structures. For instance, suppose the data is in the form of a bipartite network, with two distinct sets of nodes and ties between nodes of different types but not between nodes of the same type (e.g. such a data structure can represent people's membership of clubs). Then $\mathbf{X}$ may comprise a set of variables X_{pa} indicating the presence or absence of a tie between a node p of the first type and a node a of the second type. The data may involve *multiple* or *multivariate networks*, with r different types of relations among the one set of nodes. In that case $\mathbf{X}$ may be thought of as a three-way array of variables X_{ijr} such that $X_{ijr} = 1$ indicates that presence of a tie of type r from node i to node j. An analogous three-way array may represent network data collected for the same set of nodes at different time points such that $X_{ijt} = 1$ indicates that presence of a tie at time t from node i to node j.

Nodes themselves can have certain properties (*attributes*) that may be measured as binary, categorical or continuous variables. We denote an attribute variable with the vector $\mathbf{Y}$, where $Y_i = y_i$ indicates that for attribute Y node i has a value y_i.

IV. Dependence hypotheses

Following work in spatial statistics [1], Frank and Strauss introduced the notion of a *dependence graph* into social network modeling [9] in order to represent possible dependencies among network variables X_{ij}. The nodes of the dependence graph are the network variables X_{ij} and an edge between two nodes indicates dependence between the respective variables, even when the observed values for all remaining variables are known. These edges specify a *neighborhood* relationship between pairs of variables and the cliques of the dependence graph can be thought of as *local social neighborhoods* for the set of tie variables [26]. Conversely, the absence of an edge in the dependence graph indicates that two network variables are conditionally independent, and are not neighbors, so that their interaction need not be taken into account in a model based on local social neighborhoods. (A more technical discussion on dependence graphs for social network models can be found in [34].)

The previous paragraph describes one important result from the Hammersley-Clifford theorem [1], that the form of a probability distribution for a set of interacting variables relates solely to the neighborhood structure (or clique structure) of the dependence graph (with a single network variable also taken as a clique.) For a given node set N, once a dependence hypothesis is specified and the dependence graph is defined, it follows *necessarily* from the Hammersley-Clifford theorem that:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa(\boldsymbol{\eta})) \exp\{\Sigma_A \eta_A g_A(\mathbf{x})\} \qquad (1)$$

where:

(i) the summation is over all neighborhoods A;

(ii) η_A is a parameter corresponding to the neighborhood A (and there is no non-zero parameter for any set of ties that do not constitute a neighborhood);

(iii) $g_A(\mathbf{x}) = \prod_{x_{ij} \in A} x_{ij}$ is the *network statistic* corresponding to neighborhood A and indicates whether all the ties in A are observed in the network $\mathbf{x}$;

(iv) κ is a normalizing quantity, that is a function of the parameter vector $\boldsymbol{\eta}$, to ensure that (1) is a proper probability distribution.

All exponential random graph models take this general form, describing a probability distribution of graphs on n nodes. Neighborhoods A are subsets of possible ties, so they represent possible subgraphs that may or may not be observed in the network $\mathbf{x}$. There is one, and no more than one, parameter for each distinct neighborhood (although one neighborhood may be a subgraph of another and so both may have separate parameters.)

The form of the model in (1) is too general to be identifiable and some homogeneity constraint needs to be imposed on what is otherwise a large number of neighborhoods A (and hence a large number of parameters). The original proposal [9] was that parameters should be equated across subsets of ties that were isomorphic to each other (that is, subgraphs that are indistinguishable once node labels are removed). With some further constraints (noted below), this assumption can produce identifiable models that can be fitted to data. (Other possible ways to constrain the number of parameters are also described below.) Following the terminology first used by Moreno [24], we term these

isomorphic neighborhoods as *network configurations* (although some of the network literature of the last decade has adopted the term *motif*.) Counts of configurations become the sufficient statistics of the model.

Examination of the parameters for the various configurations of a fitted model can provide insight into the social processes that may underpin the network. The strength and direction of any particular parameter value will affect how frequently the corresponding configuration is observed. If the parameter is large and positive, we expect to observe the corresponding configuration to occur more frequently than if the parameter value were zero [39].

Obviously, good dependence hypotheses are necessary for this approach to work, as they crucially shape the nature of the model by defining the possible configurations.

V. Bernoulli random graph (Erdös-Rényi) models

Suppose we hypothesize that there are no dependencies within the network, so that the dependence graph has no edges and the only neighborhoods are single tie variables X_{ij}. Then (1) becomes:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp(\Sigma_{i,j}\, \eta_{ij} x_{ij})$$

Constraining parameters to be equal for isomorphic neighborhoods here implies equating all the parameters $\eta_{ij} = \theta$, as there is only one type of neighborhood. Then we have:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp(\Sigma_{i,j}\, \theta x_{ij}) = (1/\kappa) \exp(\theta L(\mathbf{x}))$$

where $L(\mathbf{x})$ is the number of edges (arcs) in the network $\mathbf{x}$ and θ is a *density* or *edge* parameter (sometimes called a *choice* parameter in older literature.) This graph distribution is equivalent to that of a distribution of simple random graphs where the probability of a tie between a pair of edges is $\exp\theta/(1+\exp\theta)$.

The dependence assumption here is unrealistic and this model will fit few, if any, real networks, but it is often a useful null or baseline model.

VI. Dyadic independence models

For directed networks, a dyadic independence hypothesis implies edges in the dependence graph between variables X_{ij} and X_{ji}. Neighborhoods then take the form of single edges and dyadic pairs $\{X_{ij}, X_{ji}\}$. Applying constraints in the form of the one density parameter θ to all single edge neighborhoods and the one *reciprocity parameter* ρ to all dyadic neighborhoods results in the following two parameter model:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp(\theta L(\mathbf{x}) + \rho M(\mathbf{x}))$$

where $L(\mathbf{x})$ is the number of arcs and $M(\mathbf{x})$ the number of mutual dyads (i.e. where $X_{ij} = X_{ji} = 1$) in the network $\mathbf{x}$. This model has the capacity to model reciprocity effects in

directed networks. Slightly less sweeping homogeneity constraints result in the p_1 model [15].

The interpretation of a single parameter in an exponential random graph model is relative to other effects in the model, so that in this case, the presence of a large and positive ρ parameter is relative to the density effect. This is an advantage: in this model, for instance, we can infer that there is substantial reciprocity *given the density of the graph.* It is of little use to talk about a large reciprocity effect in absolute terms, because the number of arcs in the graph (i.e. the density) establishes the preconditions for reciprocity. In other words, graphs of very low density have very little opportunity to demonstrate reciprocity anyway, and the absence of mutual dyads in a low density graph may simply be a feature of the low density. Similarly the presence of many mutual dyads in a high density graph may simply be explained by the high density, without the need to infer a reciprocity effect. A large positive ρ, on the other hand, shows that the number of arcs in a graph are arranged in sufficiently many mutual dyads to suggest that a separate and substantial reciprocity process is required to explain the structure of this network, over and above any density effect.

Once again, the dyadic independence assumption does not do a good job of reproducing real networks, because the model has no capacity for triangulation. The model may be useful as a baseline model.

VII. Markov random graphs

Frank and Strauss [9] proposed Markov dependence, by postulating that a possible tie from *i* to *j* is assumed to be contingent on any other possible tie involving *i* or *j*, even if all other ties in the network are fixed. Markov dependence implies that two possible network ties are conditionally independent unless they share a common actor. They showed that this assumption resulted in models with configurations for nondirected graphs of single edges, star-like structures and triangles (or 3-cycles). For directed graph models, configurations include directed counterparts of these as well as reciprocity (see [54] for a fuller description of directed graph parameters). Frank and Strauss termed these *Markov random graph models*.

Once homogeneity is imposed across isomorphic neighborhoods, we obtain configurations (and related parameters) for nondirected networks as depicted in Figure 1. The resulting model is:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp(\theta L(\mathbf{x}) + \Sigma_{r=2,n-1} \sigma_r S_r(\mathbf{x}) + \tau T(\mathbf{x})) \qquad (2)$$

where:
(i) θ is a density parameter, and $L(\mathbf{x})$ the number of edges in $\mathbf{x}$, as before;
(ii) σ_r is a parameter for a *star of size r*, and $S_r(\mathbf{x})$ is the number of stars of size *r* in $\mathbf{x}$;
(iii) a star of size *r* is a configuration centered on a single node *i* such that there are *r* edges emanating from *i* (note that stars of size larger than *r* contain many stars of size *r*,

so that this is not simply a partition into nodes of different degrees, although the full set of counts $S_r(\mathbf{x})$ can be converted into the degree distribution, and vice versa);
(iv) τ is a triangle or clustering parameter and $T(\mathbf{x})$ is a count of the number of triangles in $\mathbf{x}$.

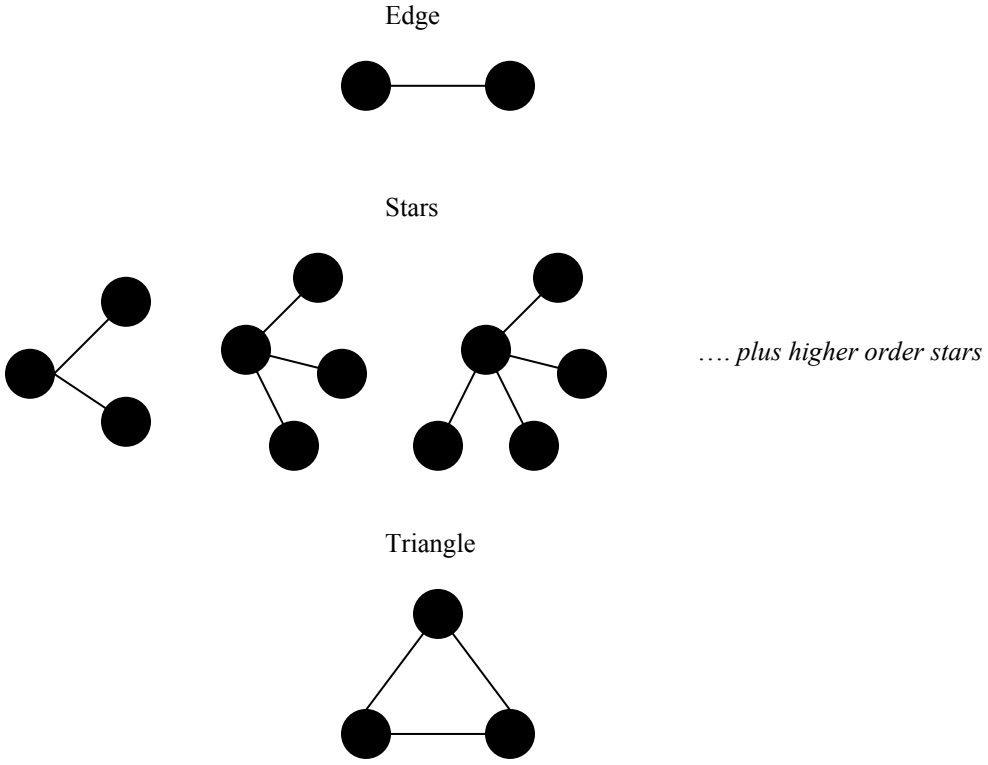

Figure 1: Configurations for a nondirected Markov random graph model

This model still has too many star parameters to be identifiable. Note that the expression $\theta L(\mathbf{x}) + \Sigma_{r=2,n-1} \sigma_r S_r(\mathbf{x})$ completely parameterizes the degree distribution (so the τ parameter can be interpreted as the strength of transitivity conditional on the degree distribution of the graph). An identifiable model can be obtained by restricting the number of star parameters to be substantially less than $n - 2$, for instance, to less than four (an alternative approach is described below). Then the model becomes:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp(\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \tau T(\mathbf{x})) \qquad (3)$$

The parameters can be interpreted in ways that represent plausible social processes [37]:
(i) θ is a baseline propensity for social ties to form;
(ii) σ_2 represents a tendency for individuals to seek multiple partners (if it is positive);
(iii) whereas σ_3 (if negative as it often is when this model is fitted to real data) represents a ceiling effect against having too many partners; so, together the balance between σ_2 and

σ_3 represents the benefits of multiple contacts against the costs of maintaining too many contacts;
(iv) while τ represents tendencies towards clustering.

An alternative interpretation is to consider the θ, σ_2 and σ_3 parameters as controlling for the first three moments of the degree distribution, so that while model (2) in effect presents a full parameterization, model (3) provides a more parsimonious control over the degree distribution (saturated models of the degree distribution are discussed by [42] and [46], see also [11].) Of course, for any data set it is an empirical question whether the first three moments are sufficient to capture the degree distribution adequately. Below, we describe ways to investigate that question.

Extensions of this basic Markov random graph model include models for multivariate networks [28], for valued networks [38] and for affiliation networks [41] (see also [27]).

By tuning parameter values even the simple four parameter model (3) can represent many different types of networks, including small world networks, networks with long paths, and highly clustered ("caveman") networks [37]. For some parameter values, however, the models became "frozen" into certain unchanging highly structured patterns. This effect is an illustration of *degeneracy* for these models, which was a research issue given important attention in the early years of the 21st century. The upshot of this body of research was that homogeneous Markov random graph models were recognized as frequently inadequate to deal with real social network data.

VIII. Simulation and model degeneracy

The problems of degeneracy became apparent through a number of simulation studies. Strauss was the first to describe a rather straightforward application of standard statistical simulation techniques (e.g. Metropolis or Metropolis-Hastings algorithms) to simulate Markov random graph models for a fixed set of nodes and given set of parameter values [49]. More recent treatments and some important results are given in [44, 13, 14, 17, 37]. These methods provide a principled statistical means to produce a distribution of graphs with probabilities of each graph being observed in the distribution consistent with any particular model in the form of (1). Typically the simulation is run for a large number of iterations and a sample of graphs is extracted and examined to understand typical properties of the graphs in the distribution.

A graph distribution is termed as *near degenerate* [13, 14] if it implies only a very few (possibly only one or two) distinct graphs with substantial non-zero probabilities. For instance, certain parameter values for Markov random graph models place almost all of the probability mass on either the empty or the full graph. These cannot be good models for human social networks. For certain parameter values, there may be two quite separate regions where graphs are most likely (possibly one region of low density and another of high density graphs), with the possibility of a dramatic phase transition from one region to the other (for instance, as parameter values change very slightly.) Such phase transitions and other aspects of degeneracy have been studied for a variety of simple

Markov random graph models by a number of authors [2, 12, 13, 14, 19, 25, 37, 40, 44, 47].

Issues of degeneracy call into question whether Markov random graph models can adequately represent most network data. It would not matter if some regions of the parameter space produced degenerate models, as long as those regions applicable to empirical data were non-degenerate. Experience shows the opposite applies: for some data, Markov random graph models turn out to be well-behaved, but for many empirical networks they are degenerate. When applied to social networks with a few very high degree nodes, or with some regions of high triangulation, degeneracy frequently occurs for Markov models. Models such as (3) reflect homogeneity constraints, and they appear to have difficulty when the data presents inhomogeneity in the distribution of degrees or triangles. It is not entirely certain whether the problem with the models arises because of the Markov dependence assumption or the assumption of homogeneity. In any event, to deal more effectively with real networks, a different approach to these models seems necessary. To date, attention has been directed to revisiting the dependence assumption.

Snidjers and colleagues [47] drew two conclusions from the tendency for Markov random graph models to be degenerate when fitted to real networks with a high level of clustering: (1) the Markov dependence assumption may be too restrictive; (2) the representation of the social phenomenon of transitivity by the total number of triangles might be too simplistic. They proposed specifications that drew on both these possibilities. Before we introduce these new specifications, however, we turn to higher order network dependencies.

IX. Social circuit dependence: Partial conditional dependence hypotheses

There are several ways in which network dependence assumptions could be extended beyond Markov dependence [26]. One approach is of *partial conditional dependence*, whereby the presence of certain network ties *created* dependence among other possible network ties. This assumption permits the emergence of dependence structures as ties come into and go out of existence. This type of dependence can be incorporated into a version of the Hammersley-Clifford theorem and hence into exponential random graph models [26].

A particular hypothesis about this type of emergent dependence was used by [47] and described by [40] as *social circuit dependence.* Social circuit dependence is defined as two possible network ties being conditionally dependent if their observation would lead to a 4-cycle. This type of dependence is depicted in Figure 2. Here $X_{rs} = X_{uv} = 1$, that is, in the network data there are observed edges between nodes r and s, and between nodes u and v. The presence of these edges leads to conditional dependence between variables X_{ru} and X_{sv} because if edges were also observed between r and u, and between s and v, a 4-cycle would result in the graph.

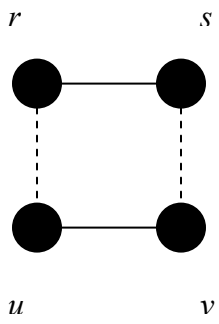

Figure 2
Social circuit dependence
(Broken lines represent possible edges; full lines represent observed edges
X_{ru} is conditionally dependent on X_{sv})

Interpretation of this hypothesis is discussed by [40]. Generally we would expect two distinct possible edges (*r,u*) and (*s,v*) to be conditionally independent. If, however, person *r* knows person *s*, and person *u* knows person *v*, then the presence of a tie between *r* and *u* can make the presence of a tie between *s* and *v* more likely, that is, they are conditionally dependent. It is simple to think of real social circumstances where this happens: for instance, in families the presence of a friendship between two children increases the chances of the two mothers coming to know one another; or in a business, cooperation between two bosses may lead to their employees working together; or in research, two postdoctoral researchers might discuss issues with each other because their academic mentors have been collaborators. In all of these cases, what makes the dependence come into effect is the presence of ties that constitute part of the 4-cycle. Without those ties, the dependence does not arise. So, the mothers are the mothers of the two particular children, not the mothers of entirely different children; the employees are employed by those particular bosses; and the two post-docs are supported by those two mentors.

X. Social circuit specifications

Snijders and colleagues [47] proposed three new statistics for exponential random graph models for nondirected networks: *alternating k-stars*, *alternating k-triangles*, and *alternating independent two-paths*. To derive these statistics they use: (1) social circuit dependence *in addition to* Markov dependence; and (2) a non-linear functional form combining various configurations into the one statistic.

Alternating *k*-stars

The alternating *k*-star parameter uses the second idea alone and does not extend dependence beyond the class of Markov random graph models. Rather than adopt the usual practice of limiting the number of higher order stars – as for example in model (3) where the star parameters are deliberately limited to no more than 3-stars – Snidjers and colleagues returned to model (2), keeping all σ_r in the model but imposing constraints among these parameter values. Specifically, the *alternating k-star assumption* proposes that for all $k \geq 2$, $\sigma_{(k+1)} = -\sigma_k/\lambda$ for some λ greater than 1. Then in (1) there is one parameter σ for all star effects, with an associated statistic:

$$u = \sum_{k=2}^{n-1} (-1)^k \frac{S_k}{\lambda^{k-2}} \tag{4}$$

where the parameter σ is referred to as the *alternating k-star parameter*. As λ is greater than 1, the direct impact of higher order stars (that is, very high degree nodes) is reduced for higher *k*. Of course, as noted above, high degree nodes still have substantial effect in the model as they produce many 2- and 3-stars, and so on. The alternating sign helps balance these overlapping star counts for high degree nodes. In [47] λ is set at 2 but [17] show how to estimate an optimal value of λ (see also [16]). The alternating *k*-star parameter is equivalent to a *geometrically weighted degree parameter* that explicitly models the degree distribution but puts more weight on the numbers of nodes with lower degrees, with weights decreasing geometrically as the degrees increase [16].

The interpretation of the parameter is as follows [47, 40]. If the alternating *k*-star parameter is positive, then highly probable networks are likely to contain some higher degree nodes. A positive alternating *k*-star parameter (together with a negative density parameter) implies graphs that exhibit preference for connections between a larger number of low degree nodes and a smaller number of higher degree nodes, akin to a core-periphery structure.

Alternating *k*-triangles

The alternating k-triangle assumption incorporates both social circuit and Markov dependence. Snijders and colleagues show that with both these dependence assumptions operating simultaneously, configurations more complex than simple triangles are possible [47]. In particular, they introduced the notion of a *k-triangle*, a combination of *k* individual triangles that all share one edge (the common *base* of the *k* triangles), as represented in Figure 3.

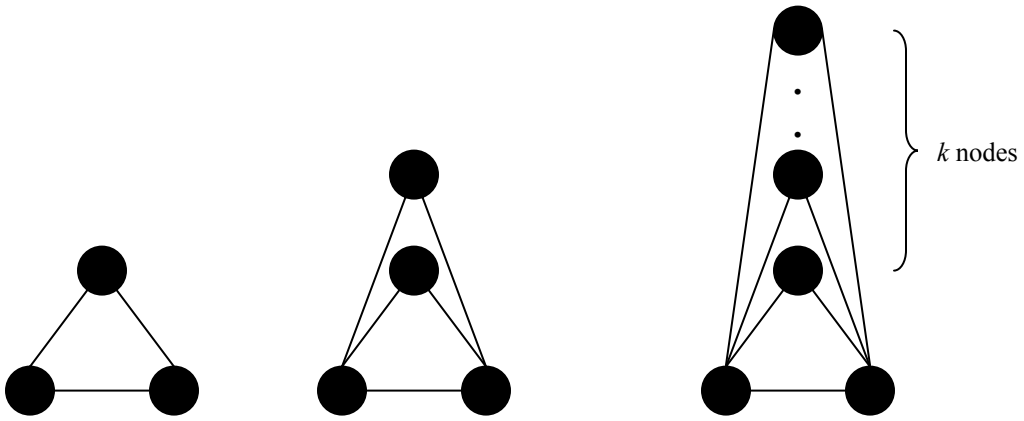

Figure 3
Various *k*-triangles

Let T_k be the count of *k*-triangles in a graph. Then the *alternating k-triangles assumption* combines these counts into the one statistic in an analogous way as for the alternating *k*-stars, that is:

$$t = 3T_1 + \sum_{k=1}^{n-3} (-1)^k \frac{T_{k+1}}{\lambda^k} \tag{5}$$

where the factor of 3 for T_1 is due to symmetry considerations in a non-directed triangle [47].

This is the *alternating k-triangle statistic* with an associated *alternating k-triangle parameter* τ. A positive *k*-triangle parameter indicates triangulation in the network but also tendencies for triangles themselves group together in larger higher order "clumps". A positive alternating *k*-triangle effect combined with a negative alternating *k*-star effect can produce a web of multiple smaller regions of triangulation [40, 47].

[16] shows that the alternating *k*-triangle parameter is equivalent to the *edge-wise shared partner (*or ESP) parameter that models the distribution of shared partners of tied actors, but with weights decreasing geometrically as the number of shared partners increase.

Alternating *k*-two-paths

Snijders and colleagues also proposed a parameter that represents a lower order configuration for a *k*-triangle, namely a *k-two-path* which is a *k*-triangle without the base [47]. These configurations, represented in Figure 4, describe the number of distinct two-paths between a pair of nodes. The motivation was to provide a parameter that, in

conjunction with *k*-triangles, would distinguish between tendencies to form edges at the base, or at the sides of a *k*-triangle.

Again the counts of these configurations are combined into the one alternating statistic in an analogous way as in the previous two cases, to produce an alternating 2-path statistic and parameter for the model. [16] shows that this new parameter is equivalent to the *dyad-wise shared partner* (or DSP) parameter that models the distribution of shared partners of actors who may or may not be tied, but with weights decreasing geometrically as the number of shared partners increase.

The parameter can be interpreted as representing localized multiple connectivity between nodes. When this parameter is negative, together with a positive alternating k-triangle parameter, there is a tendency against 4-cycles in the network, unless those cycles include triangles (alternatively, the presence of many 2-paths between nodes is related to the formation of triangles.)

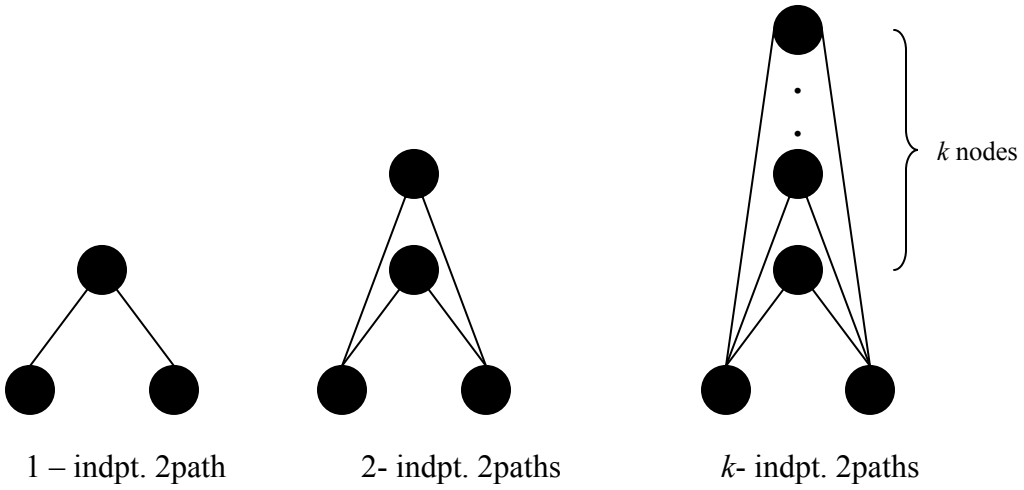

Figure 4
Various alternating *k*-independent 2path configurations

Specifications for directed graphs

Counterpart statistics and parameters for directed graphs have also been proposed. We do not discuss these here but refer readers to [47].

XI. Estimation

Pseudo-likelihood estimation [50], based on logistic regression techniques, was commonly used until more principled methods became available in recent years.

Preferred methods of estimation involve simulation procedures (Monte Carlo Markov chain maximum likelihood estimation - MCMCMLE) which for these models has been discussed by a number of authors [4, 5, 6, 17, 44, 46, 55]. Software for Monte Carlo estimation techniques has recently become publicly available ([40] provides a review). The observed graph statistics are compared with those expected under a set of provisional parameter values, using a stochastic simulation based on those values. The goal is to find a set of parameter values whereby the observed statistics are equal to the expected statistics. Parameter values are refined until the observed and expected statistics can be equated, when the model is said to converge. If the parameter estimates never converge, the model is degenerate. Standard errors can also be estimated.

XII. Goodness of fit and comparisons with Markov models

Simulation also provides an innovative means to assess how well the model can represent the data. By simulating the model from the parameter estimates, and extracting a sample of graphs it is possible to examine any graph statistic of interest (whether there is an associated parameter in the model or not). Any graph statistic for the observed graph can be compared to the distribution of such graph statistics from the simulated graphs. If the observed graph statistic is not extreme in the distribution of simulated statistics, then it is plausible that that particular feature of the observed graph could have come from the distribution of graphs implied by the model. In this way, it is possible to check which graph features the model successfully captures. This procedure is described in detail by [18] (see also [40]).

[11] provides a compelling example of how this procedure may be used to make decisions about model selection. On the basis of applying the new parameterization to a large school-based network of over 1000 nodes, [11] concludes that for this data Markov models were degenerate, but the newer parameterization had similar underlying interpretations, avoided degeneracy and was empirically validated. Other comparisons with Markov models demonstrate substantial improvements for the new parameterization in overcoming degeneracy [40].

A simple empirical example

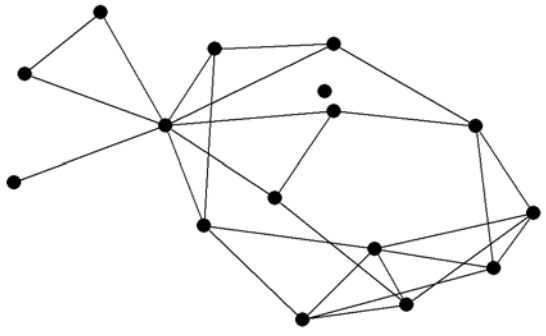

Figure 5
Social relationships among 16 actors

Figure 5 presents an empirical network of social relationships among 16 actors. Markov models do not converge for this network. Table 1 provides parameter estimates and standard errors for models involving the social circuit parameters. The third column in the table provides a convergence statistic which indicates good convergence if it has an absolute value less than 0.1 [40]. In this case, the model exhibits excellent convergence.

Table 1
Parameter estimates: Florentine families network
(NB: $\lambda = 2$)

Parameter	Estimate	Standard error	Convergence
Edge	-0.04	2.09	0.01
Alternating *k*-star	-1.01	0.78	0.02
Alternating *k*-triangle	0.68	0.33	0.02
Alternating *k*-2paths	0.18	0.17	0.03

A parameter may be inferred as important if its estimate in absolute terms is more than twice its standard error. So in this simple model, there is a substantial effect for alternating *k*-triangles. In Figure 5, regions of triangulation can be seen, consistent with the alternating *k*-triangle effect, perhaps combined with the negative alternating *k*-star effect (although the reliability of that negative effect is not certain, given that the estimate is not more than twice the standard error.) In sum, there is good evidence here that this network differs substantially from a simple random network, in particular, that there is substantially more triangulation in this network than would be expected in a simple random graph.

Table 2 shows the goodness of fit analysis for graph features other than those in the model. This analysis was based on extracting 1,000 sampled graphs from a simulation of 1,000,000 (after a burn-in of 100,000), using the parameter estimates in Table 1. The goodness of fit statistic takes the form of a *t*-ratio, calculated as the difference between the graph statistic and the mean from the simulated graphs divided by the standard deviation as estimated from the sample. For graph features not in the model, the model is considered to reproduce the data well if the goodness of fit statistic is not extreme (for instance, with *t*-ratios of less than 2). As can be seen, on the features examined, including counts of various Markov configurations, aspects of the degree distribution, and clustering, the model reproduces the data well. There is one isolate in the data, whereas 71% of simulated graphs had no isolates, hence the rather larger goodness of fit statistic. Nevertheless, with 29% of graphs in the sample having at least one isolate, the data cannot be said to be extreme in this regard.

Table 2
Goodness of fit analysis: Florentine families network model

Graph feature	Goodness of fit
Number of 2-stars	-0.10
Number of 3-stars	-0.25
Number of triangles	-0.17
Standard deviation, degree distribution	-0.10
Skew, degree distribution	-0.12
Number of isolates	1.06
Clustering coefficient	-0.09

The clustering coefficient in the Table refers to the proportion of actual to possible triangles and is calculated as thrice the number of triangles as a ratio of the number of two-paths (the factor of three applies because there are three two-paths in any one triangle.)

Table 2 does not show features of the geodesic distribution but a separate analysis reveals that the observed geodesic distribution is consistent with those from the simulated graphs: none of the geodesic quartiles of the observed graph is extreme compared to the distribution of quartiles from the simulated graphs. Overall, the model can be said to fit the data well.

XIII. Further extensions and future directions

Multiple networks

It is possible to use the same general framework as (1) when **X** is a three-way array representing multiple relations on the one set of nodes. The parameterization, however, becomes more complex. [28] showed how to develop Markov models along these lines, with these models subsequently used in some interesting empirical work, principally in organizations [22, 23]. This empirical research, however, utilized pseudo-likelihood estimation and did not take into account possible model degeneracy. See also [20].

Early work on adapting the new specifications proposed by Snijders and colleagues to multiple networks has concentrated on the case of two networks, by using the alternating *k*-star, -triangle, and -2path parameters for effects within networks and then investigating dyadic associations between networks [56].

Bipartite networks

Again (1) may be used as a basis for models for bipartite networks. [41] were the first to do this with Markov dependence. [27] proposed additional parameters that derived from partial conditional dependence assumptions. [53] incorporated these and other parameters, combined with an alternating combination of configuration counts analogous to those in the newer specifications, into a model for bipartite networks.

Longitudinal networks

Perhaps the major approach to modeling network data collected at multiple time points is the actor-oriented models of Snijders and colleagues [43] which utilize a continuous time Markov chain technique. Certain specifications of actor-oriented dynamic models have exponential random graph models as their stationary distribution [43], so the links between the two approaches are quite strong. Accordingly, it is possible to develop longitudinal versions of exponential random graph models to investigate network dynamics. The conceptual difference with the actor-oriented approach is quite subtle: exponential random graph longitudinal models suppose that ties change in response to particular social neighborhoods of other ties, rather than in response to actors seeking to optimize particular structural positions. Early work along these lines, including the evolution of multiple networks, has been reported by [29]. For further discussion on the links between the actor-oriented and tie-based versions of the models, see [45].

Nodal attributes

One of the important issues that researchers often wish to examine is whether and how nodal attributes relate to network structure. For instance, there are good empirical and theoretical grounds for expecting that people who have social relationships are more likely to be similar to each other (known as *homophily*). Conceptually, two possible processes apply: individuals may develop ties because they are similar (*social selection*) or individuals who are tied may influence each other to be similar (*social influence*). Of course, both of these processes may occur simultaneously and it is a methodological challenge for a given set of data to disentangle the two to determine the respective strengths of selection and influence effects, respectively. Snijders and colleagues [30, 48, 51] have recently developed innovative methods for investigating this question for longitudinal data within actor-oriented longitudinal models.

[35] proposed social selection models within the framework of (1), having the form:

$$\Pr(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y}) = (1/\kappa) \exp\{\Sigma_A \eta_A g_A(\mathbf{x},\mathbf{y})\} \qquad (6)$$

where **Y** is a vector of attribute variables (binary, categorical or continuous). Here the configurations *A* relate to combinations of nodal attributes and network subgraphs, and the network is modeled conditional on a fixed distribution of attributes. The intent of such models may be various: accounting for heterogeneity (that may otherwise cause difficulties in fitting models) as well as assesing selection, homophily, and covariate activity and popularity, while controlling for dependencies naturally occurring in the network setup

An analogous approach to social influence models and network contagion was proposed by [36] where the patterns of attributes were modeled conditional on a fixed network:

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = (1/\kappa) \exp\{\Sigma_A \eta_A g_A(\mathbf{x},\mathbf{y})\} \quad (7)$$

Future Directions

Not only have the new specifications shown a remarkable improvement in the successful modeling of real networks, they have opened up a wide range of network modeling possibilities. The discussion of model extensions mentioned above gives some indication of current work in this area. Of course, the possibility of further improvements in model specification has not been closed. Future directions may include methods to examine very large, community-based networks, not only in terms of data collection (for instance, the development of model-based sampling methods) but also theoretical issues based on possible differences in dependence between very large scale and smaller scale structures. Model-based approaches to missing network data, perhaps within a Bayesian framework, are also in prospect [10, 21].

The use of exponential random graph models in empirical research, especially in combination with multiple measures on individuals, opens a wide prospect for network statistical modeling to make a substantial contribution in many applied contexts.

Bibliography

Primary literature (in text):

1. Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B 36: 96-127.
2. Burda Z, Jurkiewicz J, Kryzwicki A (2004) Network transitivity and matrix models. Physical Review E69: 026106.
3. Corander J, Dahmström K, Dahmström P (1998) Maximum likelihood estimation for Markov graphs, Research report. Stockholm: Department of Statistics, Stockholm University.
4. Corander J, Dahmström K, Dahmström P (2002) Maximum likelihood estimation for exponential random graph models. In: Hagberg J (ed.), Contributions to social network analysis, information theory and other topics in statistics: A festschrift in honour of Ove Frank. Stockholm: Department of Statistics, Stockholm University.

5. Crouch B, Wasserman S (1998) Fitting p^*: Monte Carlo maximum likelihood estimation. International Conference on Social Networks, Sitges, Spain, May 28-31.
6. Dahmström K, Dahmström P (1993) ML-estimation of the clustering parameter in a Markov graph model, Research report. Stockholm: Department of Statistics, Stockholm University.
7. Erdös P, Rényi A (1959) On random graphs. I. Publicationes Mathematicae (Debrecen) 6: 290-297.
8. Frank O (1981) A Survey of Statistical Methods for Graph Analysis. Sociological Methodology 12: 110-155.
9. Frank O, Strauss D (1986) Markov graphs. Journal of the American Statistical Association 81: 832-842.
10. Gile K, Handcock M. (2006) Model-based assessment of the impact of missing data on inference for networks. CSS working paper 66: University of Washington.
11. Goodreau S (2007) Advanced in exponential random graph (*p**) models applied to a large social network. Social Networks 29: 231-248.
12. Häggström O, Jonasson, J (1999) Phase transition in the random triangle model. Journal of Applied Probability 36: 1101-1115.
13. Handcock M (2002) Statistical models for social networks: Degeneracy and inference. In: Breiger R, Carley K, Pattison P (eds) Dynamic social network modeling and analysis. Washington DC: National Academies Press, pp. 229-240
14. Handcock M (2003) Assessing degeneracy in statistical models of social networks. Center for Statistics and the Social Sciences, Working Paper no 39.
15. Holland P, Leinhardt S (1981) An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76: 33-65.
16. Hunter D (2007) Curved exponential family models for social networks. Social Networks 29: 216-230.
17. Hunter D, Handcock M (2006) Inference in curved exponential families for networks. Journal of Computational and Graphical Statistics 15: 565-583.
18. Hunter D, Goodreau S, Handcock, M (in press) Goodness of fit of social network models. Journal of the American Statistical Association.
19. Jonasson J (1999) The random triangle model. Journal of Applied Probability 36: 852-867.
20. Koehly L, Pattison P (2005) Random graph models for social networks: multiple relations or multiple raters. In: Carrington P, Scott J, Wasserman S (eds.), Models and methods in social network analysis. New York: Cambridge University Press pp.162-191.
21. Koskinen J (2004) Essays on Bayesian inference for social networks. Department of statistics: Stockholm University.
22. Lazega E, Pattison P (1999) Multiplexity, generalized exchange and cooperation in organizations. Social Networks 21: 67-90.
23. Lomi A, Pattison P (2006) Manufacturing relations: An empirical study of the organization of production across multiple networks. Organization Science 17: 313-332.
24. Moreno J, Jennings H (1938) Statistics of social configurations. Sociometry 1: 342-374.

25. Park J, Newman M (2004) Solution of the 2-star model of a network. Physical Review E70: 066146.
26. Pattison P, Robins G (2002) Neighborhood-based models for social networks. Sociological Methodology 32: 301-337.
27. Pattison P, Robins G (2004) Building models for social space: Neighbourhood based models for social networks and affiliation structures. Mathematiques des science humaines 168: 11-29.
28. Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks, II. Multivariate relations. British Journal of Mathematical and Statistical Psychology 52: 169-194.
29. Pattison P, Robins G, Wang P, Snijders T.A.B., Koskinen J (2006) The co-evolution of multiple networks. Sunbelt XXVI International Social Networks Conference, Vancouver, April 2006.
30. Pearson M, Steglich C, Snijders T.A.B. (2006) Homophily and assimilation among sport-active adolescent substance users. Connections 27: 51-67.
31. Rapoport A (1953) Spread of information through a population with a socio-structural bias: I. assumption of transitivity. Bulletin of Mathematical Biophysics 15: 523-533.
32. Rapoport A (1957) Contributions to the theory of random and biased nets. Bulletin of Mathematical Biophysics 19: 257-277.
33. Robins G, Morris M (2007) Advances in exponential random graph (*p**) models. Social Networks 29: 169-172.
34. Robins G, Pattison P (2005) Interdependencies and social processes: Generalized dependence structures. In: Carrington P, Scott J, Wasserman S. (eds) Models and Methods in Social Network Analysis. New York: Cambridge University Press pp. 192-214.
35. Robins G, Elliott P, Pattison P (2001) Network models for social selection processes. Social Networks 23: 1-30.
36. Robins G, Pattison P, Elliott (2001) Network models for social influence processes. Psychometrika 66: 161-190.
37. Robins G, Pattison P, Woolcock J (2005) Social networks and small worlds. American Journal of Sociology 110: 894-936.
38. Robins G, Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks, III. Valued relations. Psychometrika 64: 371-394.
39. Robins G, Pattison , Kalish Y, Lusher D (2007) An introduction to exponential random graph (*p**) models for social networks. Social Networks 29: 173-191.
40. Robins G, Snijders T.A.B., Wang P, Handcock M, Pattison P (2007) Recent developments in exponential random graph (*p**) models for social networks. Social Networks 29: 192-215.
41. Skvoretz J, Faust K (1999) Logit models for affiliation networks. Sociological Methodology 29: 253-280.
42. Snijders T.A.B. (1991) Enumeration and simulation methods for 0-1 matrices with given marginals. Psychometrika 56: 397-417.
43. Snijders T.A.B. (2001) The statistical evaluation of social network dynamics. Sociological Methodology 31: 361-395.
44. Snijders T.A.B. (2002) Markov chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3: 2.

45. Snijders T.A.B. (2006) Statistical methods for network dynamics. In: Luchini S et al (eds) Proceedings of the XLIII Scientific Meeting, Italian Statistical Society pp 281-296.
46. Snijders T.A.B., van Duijn M.A.J. (2002) Conditional maximum likelihood estimation under various specifications of exponential random graph models. In: Hagberg J (Ed), Contributions to social network analysis, information theory, and other topics: A festschrift in honour of Ove Frank. University of Stockholm: Department of Statistics pp117-134.
47. Snijders T.A.B., Pattison P, Robins G, Handcock M (2006) New specifications for exponential random graph models. Sociological Methodology 36: 99-153.
48. Snijders T.A.B., Steglich C, Schweinberger M (2007) Modeling the co-evolution of networks and behavior. In: van Montfort K, Oud H, Satorra A (eds) Longitudinal models in the behavioral and related sciences. Hillsdale, NJ: Lawrence Erlbaum pp.41-71.
49. Strauss D (1986) On a general class of models for interaction. SIAM Review 28: 513-527.
50. Strauss D, Ikeda M (1990) Pseudo-likelihood estimation for social networks. Journal of the American Statistical Association 85: 204-212.
51. Steglich C, Snijders T.A.B., West P (2006) Applying SIENA: An illustrative analysis of the coevolution of adolescents' friendship networks, taste in music, and alcolhol consumption. Methodology 2: 48-56.
52. van Duijn M, Snijders T.A.B., Zijlstra B (2004) p2: a random effects model with covariates for directed graphs. Statistica Neerlandica 58: 234-254.
53. Wang P (2006) Exponential random graph (p*) models for affiliation networks. Postgraduate diploma thesis: University of Melbourne.
54. Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and *p**. Psychometrika 61: 401-425.
55. Wasserman S, Robins G (2005) An Introduction to Random Graphs, Dependence Graphs, and p*. In: Carrington P, Scott J, Wasserman S (eds.) Models and Methods in Social Network Analysis. New York: Cambridge University Press pp.148-161.
56. Zhao Y, Robins G (2006) Multiple networks: Comparing QAP and exponential random graph (p*) models. Sunbelt XXVI International Social Networks Conference, Vancouver, April 2006.

Other relevant literature

Agneessens F, Roose H, Waege H (2004) Choices of theatre events: *p** models for affiliation networks with attributes. Metodoloski Zvezki 1: 419-439.
Anderson C, Wasserman S, Crouch B (1999) A *p** primer: Logit models for social networks. Social Networks 21: 37-66.
Baddeley A, Möller J (1989) Nearest-neighbour Markov point processes and random sets. International Statistical Review 57: 89-121.
Boer P, Huisman M, Snijders T.A.B., Zeggelink E (2003) StOCNET: an open software system for the advanced analysis of social networks. Groningen: ProGAMMA/ICS.
Brieger R (1981) Comment on "An exponential family of probability distributions for directed graphs". Journal of the American Statistical Association 76: 51-53.
Butts C (in press) Social network analysis: A methodological introduction. Asian Journal

of Social Psychology.
Contractor N, Wasserman S, Faust K (2006) Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. Academy of Management Journal 31: 681-703.
Frank O (1981) A survey of statistical methods for graph analysis. Sociological methodology 11: 110-155.
Frank O (1991) Statistical analysis of change in networks. Statistica Neerlandica 95: 283-293.
Frank O, Nowicki K (1993) Exploratory statistical analysis of networks. Quo Vadis, Graph Theory? Annals of Discrete Mathematics 55: 349-366.
Handcock M, Hunter D, Butts C, Goodreau S, Morris M (2006) Statnet: An R package for the statistical analysis and simulation of social networks. University of Washington.
Koehly L, Goodreau S, Morris M. (2004) The link between exponential random graph models and loglinear models for networks. Sociological Methodology 34: 241-270.
Lazega E, van Duijn M (1997) Position in formal structure, personal characteristics and choices of advisors in a law firm: A logistic regression model for dyadic network data. Social Networks 19: 375-397.
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chkovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298: 824-827.
Monge P, Contractor N (2003) Theories of communication networks. New York: Oxford University Press.
Newman M (2003) The structure and function of complex networks. SIAM Review 45: 167-256.
Pattison P, Robins G (in press) Probabilistic network theory. In: Rudas, T (ed), Handbook of Probability Theory with Applications.Sage Publications.
Pattison P, Wasserman S (2002) Multivariate random graph distributions: applications to social network analysis. In: Hagberg J (ed) Contributions to social network analysis, information theory and other topics in statistics: A festschrift in honour of Ove Frank. Stockholm: Department of Statistics, University of Stockholm pp. 74-100.
Robins G, Pattison P (2001) Random graph models for temporal processes in social networks. Journal of Mathematical Sociology 25: 5-41.
Robins G, Pattison P, Woolcock J (2004) Models for social networks with missing data. Social Networks 26: 257-283.
Schweinberger M, Snijders T.A.B. (2003) Settings in social networks: A measurement model. Sociological Methodology 33: 307-341.
Wang P, Robins G, Pattison P (2006) PNet: Program for the estimation and simulation of p*exponential random graph models, User Manual. Department of Psychology, University of Melbourne.