Statistical Models for Social Networks

Article *in* Annual Review of Sociology · August 2011

DOI: 10.1146/annurev.soc.012809.102709

CITATIONS
160

READS
183

1 author:

Tom A B Snijders
University of Groningen
168 PUBLICATIONS **19,465** CITATIONS

SEE PROFILE

Statistical Models for Social Networks

Tom A.B. Snijders

University of Oxford, University of Groningen

Key Words Social networks, Statistical modeling, Inference

Abstract

CONTENTS

1 INTRODUCTION

Social network analysis is a branch of social science which seems for a long time to have resisted the integration of empirical research with statistical modeling that has been so pervasive, and fruitful, in other branches. This is perhaps not surprising in view of the nature of social networks. Networks are relational structures, and social networks represent structures of dyadic ties between social actors: examples are friendship between individuals, alliances between firms, or trade between countries. The nature of networks leads to dependence between actors, and also to dependence between network ties. Statistical modeling on the other hand is normally based on assumptions of independence. The complicated nature of network dependencies has delayed the development of statistical models for network structures.

This article is concerned with statistical models for networks as outcome variables, with a focus on models relevant for social networks. This is an area in vigorous current development and intimately embedded in a larger domain, which makes it impossible to approximate a complete overview in a limited number of pages. Some neighboring topics that are not covered are models for the co-evolution of networks and individual outcomes; models that are not probabilistic in nature, or for which methods of statistical inference (such as estimation and testing of parameters) have not been developed; and event networks. The last

section provides some pointers to other literature.

This review is concerned with the models, not with the statistical methods for estimating and testing parameters, assessing goodness of fit, etc. The cited literature contains the details of the statistical procedures necessary for applying these models in practice.

1.1 Notation

A social network is a structure of ties, or relational variables, between social actors. We shall consider mainly a fixed set $\{1, \ldots, n\}$ of actors, and variables X_{ij} representing how actor i is tied to actor j. In some cases these will be directed in nature, so that X_{ij} and X_{ji} are different variables which may assume the same or different values; in other cases they will be nondirectional in nature, so that X_{ij} is necessarily equal to X_{ji}. The most frequently employed and most strongly developed data structure is for binary variables X_{ij}, where the value 1 (or 0) represents that there is (or there is not) a tie from i to j. Then the mathematical object constituted by the set $\{1, \ldots, n\}$ and the variables X_{ij} is called a graph in the nondirected, and a digraph in the directed case. The actors are called the nodes and the ties are usually called arcs or edges, depending on whether the graph is directed or not. It is usual to exclude the possibility of self-ties, so that the variables X_{ii} may be considered to be structural zeros.

The matrix with elements X_{ij} is called the adjacency matrix of the graph. The adjacency matrix as well as the graph or digraph will be denoted by X.

Replacing an index by a plus sign will denote summation over that index: thus, the number of outgoing ties of actor i, also called the *out-degree* of i, is denoted $X_{i+} = \sum_j X_{ij}$, and the *in-degree*, which is the number of incoming ties,

is $X_{+i} = \sum_j X_{ji}$.

1.2 Network dependencies

Social networks are characterized by a number of dependencies which have been found empirically as well as theoretically.

1. *Reciprocation* of directed ties is a basic feature of social networks, found already by Moreno (1934). This will be reflected by dependencies between X_{ij} and X_{ji}. Theoretical accounts have been made for it from the points of view of social exchange theory (Emerson, 1972) and game theory (Axelrod, 1984). Reciprocation need not be confined to pairs, but can circulate in larger groups, see, e.g., Molm et al. (2007). This then can lead to dependence in longer cycles such as X_{ij}, X_{jh}, X_{hi}.

2. *Homophily*, the tendency of similar actors to relate to each other, was discussed and coined by Lazarsfeld and Merton (1954) and has been the subject of much research, reviewed by McPherson et al. (2001). Theoretical arguments can be based, e.g., on opportunity, affinity, ease of communication, reduced transaction costs and break-off risks, and organizational foci (Feld, 1982) composed of similar individuals. This leads to a higher probability of ties being formed between actors with similar values on relevant covariates.

3. *Transitivity* of ties is expressed by the saying 'friends of my friends are my friends', and was proposed as an essential element of networks by Rapoport (1953a,b). Davis (1970) found large-scale empirical support for transitivity in networks. Transitivity is rooted deeply in sociology, going back to authors such as Simmel (1950) and elaborated more recently by Coleman (1990). If there is a tendency toward transitivity, the existence of the two ties

$X_{ij} = X_{jh} = 1$ will lead to an increased probability of the tie $X_{ih} = 1$, the *closure* of the triangle. Concatenation of such closure events then can lead also to the existence of larger connected groups. Transitivity therefore also has been called *clustering* (Watts, 1999).

A natural measure for the transitivity in a graph is the number of transitive triangles, $\sum_{i,j,h} x_{ij}x_{jh}x_{ih}$ (to be divided by 6 in the case of nondirected graphs). A natural normalization is to divide by the number of potentially closed triads, as proposed by Frank (1980),

$$\frac{\sum_{i,j,h} x_{ij}x_{jh}x_{ih}}{\sum_{i,j,h} x_{ij}x_{jh}} \ . \tag{1}$$

In directed graphs, transitivity can have two faces: it may point to a hierarchical ordering or to a clustered structure. These two can be differentiated by the aid of the number of 3-cycles, $\sum_{i,j,h} x_{ij}x_{jh}x_{hi}$. A relatively high number of 3-cycles points toward clustering, a relatively low number toward hierarchy. Davis (1970) found empirically that social networks tend to contain a relatively low number of 3-cycles, indicating the pervasiveness of hierarchies in social networks.

4. *Degree differentials*, some actors being highly connected and others having few connections, was studied since the last 1940s for communication networks by Leavitt, Bavelas, and others, and this led to models for *node centrality* reviewed by Freeman (1979). An important theoretical account was the rich-get-richer phenomenon, or Matthew effect, elaborated in the context of bibliographic references by de Solla Price (1976). This will lead to a high dispersion of the nodal degrees, which then may further lead to core-periphery structures (Borgatti and Everett, 1999) or various other types of hierarchical structures.

5. *Hierarchies* in directed networks, as exhibited by high transitivity and few 3-cycles, may be local or global. A global hierarchy will be indicated by the ordering of the in-degrees and/or out-degrees, where the typical pattern e.g., in esteem or advice asking, is directed from low to high. In a purely global hierarchy, which can be seen, e.g., in some advice networks, in a statistical model the degree differentials will be sufficient to explain the low number of 3-cycles. But local hierarchies are possible in directed networks even when the in-degrees and out-degrees exhibit rather little variability.

There are many other important types of dependencies between ties in networks, not mentioned here because it would require too much space.

The literature contains various ways to represent network dependencies in statistical models. Three broad approaches may be distinguished.

Incorporating network structure through covariates A first approach is to employ a model with independent residuals and to try and represent network dependence in explanatory variables. To the extent that this is feasible, it can be done mainly in longitudinal settings where earlier observations of the network can be used to produce covariates, as for example in Gulati and Gargiulo (1999).

Controlling for network structure A second approach is to control for certain aspects of network dependencies while not explicitly modeling them. The best known example of this approach is a permutational procedure, where nodes in the network are being permuted – one may say also that the rows and columns in the adjacency matrix are permuted simultaneously in such a way that the network structure is left intact. This is called the QAP (Quadratic Assignment Procedure) approach, proposed by Krackhardt (1987) and elaborated to

permutation of regression residuals (Multiple Regression QAP, or MRQAP) by Krackhardt (1988) and Dekker et al. (2007).

Another method was recently developed by Lindgren (2010). He used the idea of a heteroscedasticity-consistent estimator for a covariance matrix elaborated by White (1980) and which has been very fruitful for getting asymptotically correct standard errors for clustered data. It is also known by the affectionate name of sandwich variance estimator. Lindgren (2010) applied this idea to clustering in two dimensions – rows as well as columns of the adjacency matrix – as is seen in network data. The assumption then is that elements X_{ij} and X_{hk} in the adjacency matrix with $\{i, j\} \cap \{h, k\} = \emptyset$ are independent. Below we shall make a remark about this assumption.

A third method in the second approach is to condition on statistics that express network dependencies. Within the set of networks satisfying these constraints, the distribution is assumed to be uniform, hence the name *conditionally uniform models*. In other words, standard deviations, p-values, etc., are calculated in a reference set of networks that obey the same network dependencies as the observed data. This method is explained in Section 2.1.

The MRQAP and heteroscedasticity-consistent approaches are useful where research interest focuses exclusively on the effects of explanatory variables ('predictors', 'covariates') and not on modeling the network as such or on structural dependencies. Conditionally uniform models are useful to provide a statistical control for a few relatively simple network dependencies while testing more complicated structural properties.

Modeling network structure The third approach is to explicitly model the structural dependencies between tie variables. In contrast to the traditionally well known linear and generalized linear models which are the backbone of statistical modeling, such models need a potentially considerable number of parameters to express network structure, as we shall see below. This requires a change of vantage point for researchers because hypotheses may have to be formulated in terms not of relations between variables, such as the regression coefficients, correlation coefficients, or path coefficients of the more commonly used statistical models, but in terms of parameters representing more complex dependencies, such as transitive closure which is a dependency involving three variables at a time.

Of these three approaches, the first may now be regarded as a relict from an earlier period. Models expressing network structure only through covariates while further employing independence assumptions may have had their use in a time when more suitable methods were not available, but are likely to lead to suspicious results because of misspecification and therefore should now be avoided. The MRQAP and heteroscedasticity-consistent approaches are useful, but are further not treated here because they regard network structure as nuisance rather than substance and do not attempt to model network dependencies. The conditionally uniform approach will be treated briefly in Section 2.1. This article will be concerned mainly with models having the aim to model networks by representing network dependency explicitly in a stochastic model.

1.3 The use of probability and statistics to model networks

Many network studies are only about one single network and in that sense are $N = 1$ studies, and it should be argued why statistical methods are applicable at

all to such a kind of data.

The use of probability models in statistical inference can be motivated generally in two ways: based on plausible model assumptions: *model-based inference*; or based on the sampling mechanism used: *design-based inference*. This distinction is explained at length by Sterba (2009), who also shows how the discussion about these traditions goes back to debates in the 1920s and 1930s between two of the founding fathers of statistics, Ronald Fisher and Jerzy Neyman, of whom the former championed model-based and the latter design-based inference.

In *model-based inference* the researcher – explicitly or implicitly – constructs a probability model, makes the assumption that the observed data can be regarded as the outcome of a random draw from this model, and uses this model to derive statistical procedures and their properties. In social science applications, the model will be a stylized representation of social and behavioral theories or mechanisms, and the probabilistic components express behavioral, individual, or other differences of which the precise values are not explicitly determined by the model, and that would assume different values in independent replications of the research. Multiple linear regression models are an example. Molenaar (1988) gives an illuminating discussion of how to link the statistical model to substantive and theoretical considerations, and of different types of replication: for example, new measurements of the same individuals? or a new sample of respondents? and/or other measurement instruments? This idea of replication is linked to the idea of inference from the data to a population. The population is mathematically a probability distribution on the outcome space of the entire data set, and substantively (given that we are discussing statistics in the social sciences) a social and/or behavioral process in some set of individuals in some

social setting. The population of individuals may be described in a very precise or in more general and hypothetical terms. From multilevel analysis (Snijders and Bosker, 2011) we have learned that a given piece of research may generalized simultaneously to several populations, e.g., a population of individuals and a population of workplaces.

Probabilistic statements and properties in *design-based inference* are based on the sampling mechanism which is, in principle, under the control of the researcher. Usually there is some finite population of which the sample is a subset, drawn according to a known probability mechanism. One of the desiderata is that each population element have a positive probability to be included in the sample. The probability distribution of the sample as a subset of the population is known, but the values of the variables of non-sampled population elements are unknown. Statistical inference is from the sample to only this finite population. If the whole population has been observed, statistical inference has become superfluous.

The nice thing about design-based inference is that the probability distribution is under control of the researcher, while in model-based inference it is an assumption which must be based on plausible reasoning, some basic insights into the population and phenomena being studied, and of which the approximate validity should be checked by diagnostic methods. Because of the different nature of the populations to which the data is being generalized, design-based inference often is called descriptive, while model-based inference is called analytical; but mixed forms do exist. Since most of social science aims at studying mechanisms rather than describing specific populations, statistical methods used in the social sciences are mainly design-based. However, textbooks and teaching of statistics often found their probability models on design-based arguments, which leaves

some researchers with the incorrect impression that statistical inference should always be based on a probability sample or a good approximation to such a procedure.

The distinction between model-based and design-based inference applies directly to statistical modeling of networks. This was stressed, e.g., in Frank (2009) (section on Probabilistic Network Models). Design-based methods can be used when a sample is drawn from a larger graph. An important class of designs are link-tracing designs, where the sample uncovers nodes in waves, and the ties found for the nodes uncovered in a given wave will in some way determined the nodes uncovered in the next wave. Examples are snowball designs and random walk designs (Frank, 2009). Such methods are used, e.g., to try and find members of hard-to-reach populations and to get information about their characteristics and the network structure of such populations. An overview of the earlier literature is in Spreen (1992). An overview of the recent method of respondent-driven sampling is given by Gile and Handcock (2010). Model-based inference can be used in the regular case in social network analysis where an entire network has been observed, and it is natural or plausible to consider that the observed network data could also have been different, in the sense of Molenaar's (1988) *"what would happen if you did it again?"*. It could have been observed on a different moment, the individuals could have been different while the social and institutional context could have remained the same, external influences could have been different, etc. The idea of model-based inference is that in such a population of different networks, even though vaguely described, the systematic patterns as expressed in the parameters of the probability model would be the same, while the particular outcome observed (in this case the outcome would be the whole network) could

have been different. Like in all model-based inference, there would thus be a distinction between systematic properties of the social system and random variability, one could say, between signal and noise. The aim of the statistical model is to represent the main features of the data set – in this case, the network – by a small number of parameter estimates, and to express the uncertainty of those estimates by standard errors, posterior distributions, p-values, etc., which give an indication of how different these estimates might be if indeed the researcher would "do it again". Checking the assumptions of the model is important to guard against overlooking important features of the network data set, which also could bias the features that are indeed being reported; and to have confidence in the measures of variability that are being reported.

Inference for networks is potentially even more precarious because the traditional research design is to collect data on $N = 1$ network. The inferential issue here has an internal and an external aspect, corresponding to what Cornfield and Tukey (1956) call the *two spans of the bridge of inference*. The internal issue is that although we have only one 'system' in the sense that potentially everything depends on everything else, we do have large numbers of actors and tie variables, and we can carry out statistical inference because reasonable assumptions of conditional independence or exchangeability can be made. We return to this issue in Section 2.4. The external issue is that from the $N = 1$ observed network we would like to say something about social processes and mechanisms more generally. Whether this is reasonable depends on how 'representative' this network is for other networks – a dirty word in statistics because, as Cornfield and Tukey (1956) argued, this is a question outside of statistics. In some cases it may be argued that indeed one particular network tells us a lot about how social

structure and social constraints operate more generally. But nobody will deny that such knowledge only can be solid if there is a cumulation of results over replications, i.e., studies of broadly the same phenomenon or process in different groups and contexts. (By the way, doing statistical inference for $N = 1$ is not unusual; similar issues arise, e.g., in the economic analysis of long time series.)

More scientific progress can be made when data is available for several networks that may be regarded, in some sense, as replications of each other: several schools, several companies, several villages, etc. Such data sets still are rare but not exceptional. For example, Coleman (1961) collected friendship data for 10 schools, and more recent examples of networks collected in larger numbers of schools are the *Add Health* data (Harris et al., 2003) and the *ASSIST* study (Steglich et al., 2011). Such data sets with multiple networks call for a multilevel study, or meta-analysis, of social networks, enabling generalization to a population of networks. A first step into this direction was made by Snijders and Baerveldt (2003), but as yet this is a thoroughly underdeveloped area of network modeling.

2 SINGLE NETWORKS

This section treats the main various types of statistical models for single, i.e., cross-sectionally observed, networks.

2.1 Conditionally Uniform Models

Conditionally uniform models consider a set of statistics that the researcher wishes to control for, and then assume that the distribution of networks is uniform, conditional on these statistics. Thus, each network satisfying the con-

straints of leading to the desired statistic has the same probability; each network not satisfying these constraints has probability 0. This reflects the notion that the conditioning statistics contain that which is relevant in the studied phenomena, and the rest is randomness. Conditionally uniform distributions are typically used as straw man null hypotheses. They are used in a strategy where network properties that the researchers wishes to control for are put in the conditioning statistic, and the theory that is put to the test is expressed by a different statistic, for which then the p-value is calculated under the conditionally uniform distribution. This strategy has a mathematical basis in the theory of statistical tests that are 'similar' (i.e., have constant rejection probability) on the boundary between null hypothesis and alternative hypothesis, see Lehmann and Romano (2005).

Holland and Leinhardt (1976) initiated the study and application of this type of model, emphasizing the uniform model for directed graphs conditional on the dyad count, i.e., the numbers of mutual, asymmetric, and null dyads, denoted the $U \mid M, A, N$ distribution. They elaborated the strategy where the test statistic is a linear function of the *triad census*, the vector of counts of all possible triads (subgraphs of three nodes contained in the network). A further elaboration is given by Wasserman (1977).

This strategy has two limitations. One is that conditionally uniform models become very complicated, when richer sets of conditioning statistics are considered. For example, since in-degrees and out-degrees are basic indicators of actor position, it is relevant to condition on the in-degrees as well as the out-degrees in the network, leading to the so-called $U \mid (X_{i+}), (X_{+i})$ distribution. This is a distribution with difficult combinatorial properties. Ways to simulate it were developed by Snijders (1991), Rao et al. (1996), Roberts (2000), and Verhelst

(2008). McDonald et al. (2007) studied ways to simulate the $U \mid (X_{i+}), (X_{+i}), M$ distribution, which also conditions on the number of mutual ties in the network. But in practice one would like to go even further in conditioning, which however leads to self-defeating attempts because of combinatorial complexity.

Another limitation is that the rejection of the null hypothesis does not provide a first step toward constructing a model for the phenomenon being studied – the only conclusion is that the observed value for the test statistic is unlikely given the conditioning statistics if all else would be random. Because of these limitations, conditionally uniform distributions are not used a lot currently.

2.2 Latent Space Models

A general strategy to represent dependencies in data is the latent space model of Lazarsfeld and Henry (1968). This model assumes the existence of latent (i.e., unobserved) variables, such that the observed variables have a simple probability distribution given the latent variables. The specification of the latent variables is called the structural model, and the specification of the distribution of the observed variables conditional on the latent variables is the measurement model. Examples are factor analysis (a model proposed long before Lazarsfeld and Henry's book of 1968), item response theory, latent class analysis, and mixture models. In these examples the data are independent across individuals, and the aim of the latent space model is to parsimoniously represent dependencies between multiple variables within individuals.

A number of models for social networks can be subsumed under this category. These models all have latent variables defined for the nodes in the graph which represent the social actors, and assume that the dyads (X_{ij}, X_{ji}) are conditionally

independent, given these nodal variables. In many of them it is assumed even that the tie variables X_{ij} are conditionally independent, given the nodal variables.

This section reviews a number of latent space models, defining them by the type of latent structure and the conditional distribution of the network given the latent structure. The latent variables will be denoted by A_i for node i, with A denoting the vector $A = (A_1, \ldots, A_n)$. The space of values for A_i is denoted $\mathcal{A}$. In all latent structure models, this space will have some topological structure defining the model. In all these models, the assumption is made that the dyads (X_{ij}, X_{ji}) are independent given the vector A in space $\mathcal{A}$. For some models the 'locations' A_i of the nodes are regarded as random variables, in others as estimable deterministic parameters, and in some cases they may be regarded as either and the choice between random and fixed is rather a matter of estimation strategy.

2.2.1 Discrete Space The stochastic block model of Holland et al. (1983), Snijders and Nowicki (1994),Nowicki and Snijders (2001), and Daudin et al. (2008) is a model in which there is a node-level categorical latent variable A_i with K possible values for some $K \geq 2$. Hence the latent space is $\mathcal{A} = \{1, \ldots, K\}$ having no further structure; topologically, it is a discrete space. In the spirit of graph theory, Nowicki and Snijders (2001) refer to these values as *colors*, leading to a model of a colored graph with unobserved colors; one could also call this a latent class model for the nodes. The conditional distribution of the dyads (X_{ij}, X_{ji}) is assumed to depend only on A_i and A_j, the colors (or classes) of i and j. Thus for each pair of colors $(c, d) \in \mathcal{A}^2$, there is a probability vector for the four outcomes $(0,0), (0,1), (1,0), (1,1)$ of the dyad (X_{ij}, X_{ji}). This is a stochastic version of the concept of structural equivalence of Lorrain and White (1971). Such a model

can represent cohesive subgroups but also very different, noncohesive subgroup structures, such as social roles. This model has been extended by Airoldi et al. (2008) to *mixed membership models*, where each node can be a member of several classes, describing situations where the actors may play multiple roles.

2.2.2 DISTANCE MODELS Several variants have been developed of distance models. A function $d : \mathcal{N} \to [0, \infty)$ is called an metric, or distance function, if it satisfies the following axioms:

1. $d(i, i) = 0$ for all $i \in \mathcal{N}$;
2. $d(i, j) = d(j, i) > 0$ for all $i, j \in \mathcal{N}$ with $i \neq j$;
3. $d(i, j) \leq d(i, k) + d(j, k)$ for all $i, j, k \in \mathcal{N}$ (triangle inequality).

In the latent metric space models, it is assumed that the nodes are points in some metric space $\mathcal{A} = \{1, \ldots, n\}$ with distance function $d(i, j)$, and the probability of a tie depends as a decreasing function on the distance between the points, $P\{X_{ij} = 1\} = \pi(d(i, j))$. The closer the points, the larger the probability that they are tied.

The general definition of metric spaces is so wide that it does not lead to very useful models; more constraints are necessary (Hoff et al., 2002). Such more specific latent metric models are defined in the following paragraphs by posing further restrictions on the metric.

It may be noted that the symmetry of distance functions (axiom (2) above) implies here that the tie from i to j has the same probability as the tie in the reverse direction, but the latent distance model does not imply a special tendency to reciprocity.

Ultrametric Space Freeman (1992) proposed ultrametric models as a representation of group structure in social networks. A function $d : \mathcal{N} \to [0, \infty)$ is called an ultrametric, or ultrametric distance, if the triangle inequality is replaced by the stronger requirement

$$d(i,j) \leq \max\{d(i,k), d(j,k)\} \text{ for all } i,j,k \in \mathcal{N} \ .$$

This condition is called the ultrametric inequality. For ultrametric distances on finite spaces, it is not a restriction to assume that the set of values of the distance is a set of consecutive integers $\{0, 1, \ldots, K\}$ for $k \geq 1$.

Ultrametric distances have the property that for every cut-off point k, the graph with edge set $E_k\{(i,j) \mid 0 < d(i,j) \leq k\}$ is a perfectly transitive graph, meaning that it consists of a number of mutually disconnected cliques. Therefore ultrametrics are useful structures for representing the transitivity of social networks. Schweinberger and Snijders (2003) proposed a latent ultrametric model for nondirected graphs as a stochastic implementation of Freeman's (1993) idea of ultrametric spaces for representing networks, and of Pattison and Robins' (2002) concept of social settings. Almost the same model was proposed later by Clauset et al. (2008). For each 'level' k, the graph with edge set E_k can be regarded as a 'smoothed' version of the observed graph, representing the setting structure at level k, where $k = 1$ represents the most fine-grained and the maximum value $k = K$ the most coarse settings structure.

Euclidean Space Hoff et al. (2002) proposed to specify the latent metric as a Euclidean distance – in practice with a dimension $K = 2$, sometimes $K = 3$. For $K = 2$ this means that each node is represented by two real-valued coordinates

(A_{i1}, A_{i2}) and the distance between the two nodes is defined as

$$d(i,j) = \sqrt{(A_{i1} - A_{j1})^2 + (A_{i2} - A_{j2})^2} \ .$$

This model is in line with the graphical representations of networks in two-dimensional pictures, where the points are arranged in the plane in such a way that nearby points are linked more often than points separated by large distances. This model also represents transitivity, following from the triangle inequality and further from the specific structure of Euclidean space.

The model of Hoff et al. (2002) accommodates explanatory variables in addition to the latent distance, letting the probability of a tie between i and j depend on the distance and on explanatory variables through a logistic (or other) link function:

$$\text{logit}\big(P\{X_{ij} = 1\}\big) = \beta' z_{ij} - d(i,j) \ , \tag{2}$$

where z_{ij} is a vector of explanatory variables for the pair (i,j) and β is a vector of regression coefficients.

To represent further group structure in addition to the transitivity already implied by the latent Euclidean distance model, Handcock et al. (2007) complemented this model with the assumption that the locations A_i are outcomes of a mixture model of normal distributions. This is a double-layer latent structure: a first layer consisting of the locations, a second layer consisting of the (normally distributed) subgroups.

2.2.3 Sender and receiver effects Two generations of models have been proposed representing actor differences with respect to sending and receiving ties as well as reciprocation. Holland and Leinhardt (1981) proposed the p_1 model for directed graphs; the name p_1 was chosen because they considered this

the first plausible and viable statistical model for directed networks. In this model dyads (X_{ij}, X_{ji}) are independent; each actor has two parameters α_i, β_i responsible, respectively, for the tendency of the actor to send ties ('activity', influencing the out-degrees) and the tendency to receive ties ('popularity', influencing the in-degrees); in addition there are parameters influencing the total number of ties and the tendency toward reciprocation. The large number of parameters, two for each actor, is a disadvantage of this model, and Fienberg and Wasserman (1981) proposed a model in which this number is reduced by making the parameters dependent on categorical nodal attributes. Various other modifications and extensions have been proposed, as reviewed in Wasserman and Faust (1994).

Another way to reduce the dimensionality of the parameter of this model, without however postulating that actors with the same attributes have identical distributions of their incoming and outgoing ties, was proposed by van Duijn et al. (2004), calling this the p_2 model. In this model the activity and popularity parameters are regressed on nodal and/or dyadic covariates and in addition include random residuals, making this a random effects model. The actor-dependent residuals for the activity and popularity effects are assumed to be correlated. The major advantage of this model over the p_1 model is the possibility to include sender and receiver covariates, which is impossible in the p_1 model because such effects are totally absorbed by the α_i and β_j parameters.

This was generalized by Hoff (2005) by including in the model for the log-odds of a tie not only random sender effects A_i, receiver effects B_j and reciprocity effects $C_{ij} = C_{ji}$, but also bilinear effects $D_i D_j$ where the D_i variables again have multivariate normal distributions.

2.2.4 Ordered Space Hierarchical ordering is another feature that can be exhibited by networks; a linear order, as is usual in the well-known phenomenon of pecking orders of chickens, or more generally a partial order. Procedures for finding linear orders, applied to dominance relations between animals, are reviewed by De Vries (1998). In a partial order not all pairs of points are required to be ordered. The definition of a partial order $\preceq$ on $\mathcal{N}$ is the following:

1. $i \preceq j$ and $j \preceq i$ if and only if $i = j$ ('antisymmetry');
2. $i \preceq j$ and $j \preceq k$ implies $i \preceq k$ ('transitivity').

In the model proposed by Mogapi (2009), it is assumed that the nodes i are points in a latent partially ordered apace, and probabilities of ties depend on how the points are ordered, together with covariates. Given the partial ordering and covariates z_{ij}, the probability of a tie is assumed to be given by

$$\operatorname{logit}\big(P\{X_{ij} = 1\}\big) = \begin{cases} \pi_1 + \beta' z_{ij} \text{ if } i \preceq j \\ \pi_2 + \beta' z_{ij} \text{ if } j \preceq i \\ \pi_3 + \beta' z_{ij} \text{ if } i \not\preceq j \text{ and } j \not\preceq i \ . \end{cases} \tag{3}$$

2.3 Exponential Random Graph Models

A different type of model represents dependencies between ties directly, rather than by conditioning on latent attributes. This line of modeling started with Frank and Strauss (1986). They defined Markov dependence for distributions on network in analogy to distributions for stochastic processes: conditioning on the other random variables, two random variables are independent unless they are tied (where a tie in the stochastic process case would be defined as direct sequentiality in time). For the network case, the definition is given more precisely as follows. The array $X = (X_{ij})$ of random variables, which can also be regarded

as a stochastic graph, is a *Markov graph* if for each set of four distinct nodes $\{i, j, h, k\}$, the random variables X_{ij} and X_{hk} are conditionally independent, given all the other random variables in X. This seems a plausible kind of conditional independence, suitable for social networks.

Frank and Strauss (1986) proved that this Markov dependence for a random nondirected graph, with the additional requirement that the probability distribution is invariant under permutation of the nodes, is equivalent to the possibility to express the probability distribution of the graph by

$$\mathrm{P}\{X = x\} = \frac{\exp\left(\theta L(x) + \sum_{k=2}^{n-1} \sigma_k S_k(x) + \tau T(x)\right)}{\kappa(\theta, \sigma, \tau)} \tag{4}$$

where $L(x) = \sum_{i<j} x_{ij}$ is the edge count, $T = \sum_{i<j<h} x_{ij}x_{jh}x_{ih}$ is the triangle count, and $S_k = \sum_i \sum_{j_1<j_2<\ldots<j_k} x_{ij_1}x_{ij_2}\ldots x_{ij_k}$ is the k-star count (with $S_1(x) = L(x)$).

transitive triangle 6-star

Figure 1. Examples of subgraph structures.

The statistical parameters in this model are $\theta, \sigma_2, \ldots, \sigma_{n-1}$, and τ. Finally, $\kappa(\theta, \sigma, \tau)$ is a normalization constant to let the probabilities sum to 1. The fact that the logarithm of the probability is a linear combination of parameters and statistics makes this into an *exponential family of distributions* (Lehmann and Romano, 2005), an important class of statistical models for which many theoretical properties are known. The statistics are the so-called *sufficient statistics*, as they contain all information in the data about the values of the parameters. The sufficient statistics here are *subgraph counts*, the frequency in the graph of small

configurations: here, edges, k-stars, and triangles. The number of k-stars can be expressed as

$$S_k(x) = \sum_{i=1}^{n} \binom{x_{i+}}{k} , \tag{5}$$

which implies that the vector of k-star counts $S_k(x)$ for $k = 1, \ldots, K$ are a linear combination of the first K moments

$$\frac{1}{n} \sum_{i=1}^{n} X_{i+}^k \quad (k = 1, \ldots, K)$$

of the degree distribution. Thus, if $\sigma_k = 0$ for all k larger than some value K, in a distribution of graphs according to (4) each graph with the same moments of order up to K of the degree distribution, and the same number of triangles, is equiprobable. Often this model is used while including only a few of the σ_k parameters for low k, so that the degree distribution is characterized by a few low-order moments such as the mean, variance, and skewness.

Frank (1991) and Wasserman and Pattison (1996) generalized the Markov graph model by proposing that the exponent in (4) could contain in principle any statistic, thus allowing any kind of dependence between the tie variables:

$$\mathrm{P}_\theta\{X = x\} \;=\; \frac{\exp\left(\sum_k \theta_k s_k(x)\right)}{\kappa(\theta)} , \tag{6}$$

where the $s_k(x)$ can be any statistics depending on the network and observed covariates. They can be specified so as to reflect the research questions and to obtain a good fit between model and data. This model can represent in principle any distribution on the space of graphs that gives positive probability to each possible graph – although such a representation will not necessarily be parsimonious or tractable. This model was called the p^* model by Wasserman and Pattison (1996); more recently it has also been called the Exponential Random Graph Model ('ERGM'). An important subclass is obtained when the sufficient

statistics $s_k(X)$ are subgraph counts (such as is the case for the Markov model). Such models can be obtained from conditional independence assumptions (of which, again, Markovian dependence is one example) based on an application of the Hammersley-Clifford theorem, as proved in Wasserman and Pattison (1996). Examples are the neighbourhood models of Pattison and Robins (2002) and the models excluding 'action at a distance' discussed by Snijders (2010). An overview of the Exponential Random Graph Model is presented in Robins et al. (2007) and in the monograph Koskinen et al. (2011).

Back to the Markov model The two main virtues of the Markov model are its possibility to represent transitivity and the distribution of degrees, as reflected by the parameters τ and σ_k; and its interpretation as being equivalent to the Markovian conditional independence property. One difficulty is related to parameter estimation. The normalizing factor κ in (4) cannot be easily calculated except for uninteresting special cases, which is an impediment for the calculation of likelihoods and for traditional procedures for parameter estimation. Frank and Strauss (1986) proposed a pseudo-likelihood estimation procedure, maximizing the pseudo-log-likelihood defined as

$$\sum_{i,j} \log\left(\mathrm{P}_{\theta,\sigma,\tau}\{X_{ij} = x_{ij} \mid X^{(-ij)} = x^{(-ij)}\}\right) ,$$

where $X^{(-ij)}$ is the random graph X without the information about the tie variable X_{ij}. The pseudo-log-likelihood can be seen to have the same structure as the log-likelihood for a logistic regression model, so that standard software can be used to compute the pseudo-likelihood estimator. The dependence between the tie variables, however, creates problems for this procedure and it later was shown that this procedure and the correspondingly calculated standard errors are

quite unreliable; see, e.g., van Duijn et al. (2009). When Markov chain Monte Carlo procedures for maximum likelihood estimation were developed (Dahmström and Dahmström, 1993; Snijders, 2002), however, a second difficulty came to the surface. This relates to the probability distributions and is not restricted to a particular procedure for parameter estimation. Handcock (2002) found that these distributions can be nearly degenerate in the sense that they concentrate the probability in one or a quite small number of possible outcomes. Snijders (2002) found that these distributions can have a bimodal shape, with the two modes being totally different networks. Such properties are undesirable given that this is a model for a single observation. They will occur when the transitivity parameter τ approaches values required to represent the tendencies toward transitivity observed in real-world networks, unless the number of nodes is quite low (e.g., less than 30) or the average degree is small (less than 2). These degeneracy problems can occur even in models with $\tau = 0$.

The conclusion of these findings is that the Markov model is not a reasonable representation for most empirically observed social networks when they have more than 30 nodes, average degrees more than 2, and transitivity index (1) more than .2. Further elaborations of this problem and relevant references can be found in Snijders et al. (2006) and Rinaldo et al. (2009).

The social circuit model As a less restrictive model, Snijders et al. (2006) proposed the *social circuit assumption*, and a set of statistics satisfying this property. This assumption states that for four distinct nodes i, j, h, k, tie indicators X_{ij} and X_{hk} are conditionally independent, given the rest of the graph, unless the existence of these two ties would imply a 4-cycle in the graph (see Figure 2).

An interpretation (cf. Pattison and Robins, 2002) is that in the latter case, e.g. if $X_{ih} = X_{jk} = 1$ as in Figure 2, the existence of the tie $i - j$ would imply that the four nodes i, j, h, k are jointly included in a social setting, which would affect the conditional probability that also the tie $h - k$ exists.

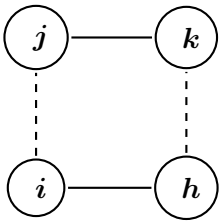

Figure 3. Creation of a 4-cycle by edges $i - j$ and $h - k$.

There are many statistics satisfying the social circuit assumption that could be chosen to represent tendencies toward transitivity as observed in social network data sets. The statistics of the Markov specification (4) are not suitable because for larger n they do not allow probability distributions concentrated around graphs with transitivity indices having any values clearly larger than the density of the graph, and smaller than 1. Snijders et al. (2006) propose statistics that do satisfy this requirement, and that correspond to the social circuit assumption (although they are by no means the only statistics obeying this assumption). Wide experience collected since this proposal has confirmed that these statistics allow to represent quite a large variety of observed social network data sets. Two mathematically equivalent versions have been proposed, indicated by the epithets 'alternating' and 'geometrically weighted', respectively. The relations between these two versions is elaborated by Hunter (2007). Here we give the geometrically weighted versions. There are three statistics: the geometrically weighted degree statistic ('GWD'), the geometrically weighted edgewise shared partner statistics ('GWESP'), and the geometrically weighted dyadic

shared partner statistics ('GWDSP'). The GWD is a function of the degree counts $\mathrm{D}_r = \mathrm{D}_r(X)$ defined as the number of nodes in X with degree r. The GWESP is a function of the edgewise shared partner statistics EP_r defined as the number of unordered linked pairs (i, j) that are both connected to exactly r other nodes,

$$\mathrm{EP}_r = \sum_{i<j} X_{ij}\, I\Big\{ \sum_k X_{ik}X_{jk} = r \Big\} \ .$$

Here $I\{A\}$ is the indicator function of the event A, equal to 1 if A is true and 0 if it is false. The GWDSP is a function of the dyadwise shared partner statistics DP_r defined as the number of unordered pairs (i, j), irrespective of whether they are linked, that are both connected to exactly r other nodes,

$$\mathrm{DP}_r = \sum_{i<j} I\Big\{ \sum_k X_{ik}X_{jk} = r \Big\} \ .$$

The edgewise and dyadwise shared partner statistics reflect tendencies toward transitivity. If there is a tendency toward transitive closure, then for any given pair (i, j), if there are many shared partners, i.e., $\sum_k X_{ik}X_{jk}$ is large, then the conditional probability of the edge $i - j$ will be high. The problem with the Markov specification is that the conditional log-odds of this edge increases linearly with the number of shared partners, which is a too strong dependence. The exponentially weighted statistics depend on a so-called weighting parameter, denoted here by α, and which attenuates the effects of high degrees, or high numbers of shared partners. Usual values of α are nonnegative, and higher values mean stronger attenuation. The exponentially weighted statistics are defined as follows:

$$\begin{aligned} \mathrm{GWD}(X) &= \sum_r w_r(\alpha)\mathrm{D}_r(X) \\ \mathrm{GWESP}(X) &= \sum_r w_r(\alpha)\mathrm{EP}_r(X) \\ \mathrm{GWDSP}(X) &= \sum_r w_r(\alpha)\mathrm{DP}_r(X) \ , \end{aligned}$$

where $w_r(\alpha)$ is given by

$$w_r(\alpha) = e^{\alpha}\left\{1 - \left(1 - e^{-\alpha}\right)^r\right\} .$$

This is an increasing function of r which is nearly linear in r if α is close to 0, and more and more strongly concave is α gets larger. Therefore, for α tending to 0, the model with these statistics approximates the Markov specification, and the degeneracy problems may be expected to become weaker as α becomes larger. The Exponential Random Graph Models with these sufficient statistics are further discussed in Snijders et al. (2006); Hunter (2007); Robins et al. (2007). In data analysis, the value of α can be set at a predetermined value or it can be estimated, as treated in Handcock and Hunter (2006).

There is an empirical interpretation to the fact that Exponential Random Graph Models with the Markov specification cannot be fitted to realistic social network data (assuming that the number of nodes is more than 30, average degree more than 2, and the transitivity coefficient clearly higher than the density), whereas models with the alternating/geometrically weighted specification can. The equivalence of the Markov specification with the conditional independence assumption implies that this assumption must be unrealistic. Thus, it is not reasonable to assume for social networks (under the stated extra conditions) that for any four distinct nodes i, j, k, h, the edge indicators X_{ij} and X_{hk} are conditionally independent, given the rest of the graph; whereas the social circuit assumption, that these edge indicators may be conditionally independent under the additional assumption that the existence of these two edges would not create a 4-cycle, is a more tenable approximation to networks observed in practice. It may be noted that this calls into question the robust standard error estimates of Lindgren (2010) discussed above, these being based on precisely this assumption.

It is unknown whether in practice this is a restriction for the validity of these standard errors.

2.4 Overview; Conditional Independence Assumptions

The three principles presented here for constructing network models all are based on conditionality, but in totally different ways.

Conditionally uniform models condition on observed statistics, and try to assess whether these are sufficient to represent the observed network. One very successful application of this principle, although applied in a statistically informal way, is Bearman et al. (2004). These authors investigated the dating network of an American high school and found that to represent the structure of the large connected component of this network it is sufficient to condition on the degree distribution, the restriction to heterosexual dating, the homophilous preference for dating somebody with similar partnership experience, and a social taboo on 4-cycles. However, the conditionally uniform approach is successful only in rare cases, because of the combinatorial complexity of the resulting distributions.

Latent space models postulate the existence of a space in which the nodes occupy unobserved (latent) positions, such that the tie indicators are independent conditionally on these positions. The estimation of these positions yields a representation of the network which can give a lot of insight, comparable to other visualizations, but with the extra element of a probabilistic interpretation. These models have a combination of rigidity and flexibility – the assumption of a particular type of space is rigid and limits the kinds of dependencies that can be represented, whereas the possibility to position the points anywhere in the space can give a lot of flexibility. The latter flexibility is mirrored by the multimodality

of the likelihood, which may lead to difficulties in estimation and ambiguity of the results. These models have the property that nodes can be dropped from the observations, without affecting the validity of the model assumptions for the remaining nodes. This is convenient for modeling, for example to handle randomly missing data, but may be an unlikely assumption for networks because taking out an actor could well have an impact on the relations between the other actors.

Exponential Random Graph Models, when using subgraph counts as sufficient statistics, are based on conditional independence assumptions between the observed tie variables. There is a large flexibility in specifying the sufficient statistics, and this can give insights in dependence structures between the ties in the network. An example is the elaboration of this type of model for directed networks by Robins et al. (2009), where many different dependence structures are possible because directions of ties can be combined in so many ways. One of the examples treated is a network of negative ties (difficulties in working with the other person), where the specific dependence structures may be of great interest.

The latent space models for categorical and Euclidean spaces, as well as Exponential Random Graph Models, now have been applied in a variety of empirical research articles and may be regarded as being part of the advanced toolkit of the currently up-to-date social scientist. As to choosing between these two types of model, researchers who are interested in detailed dependence structures might profit more from applying Exponential Random Graph Models while those interested in positions of actors might profit more from an appropriate latent space model. For both these types of model, estimation may be difficult for large networks, where "large" could be operationalized as a number of nodes of the order of one thousand or more. For latent space models the difficulty will reside in the

multimodality of the likelihood, and for Exponential Random Graph Models in the difficulty to achieve convergence of the algorithm. This may change as better computational methods become available. On the other hand, the complexity of dependencies in networks is so high that modeling large networks in a way that passes the high requirements of a good statistical fit seems intrinsically difficult to achieve.

3 NETWORK DYNAMICS

Longitudinal social network data, which can also be called data about network dynamics, can be of many different kinds. Some examples, with their salient restrictions, are the following. Friendship networks in a class of school children may be recorded at a few moments in time, but are sure to have been changed in between. Alliances between firms may have their starting points registered, but not their termination dates. Large data sets of email or other electronic communications have been registered, but often with little additional information about the senders and receivers.

Here also there are fundamental questions about dependencies, but now the dependencies are spread out in time, with changes in network ties depending on structures of earlier ties in the network. The arrow of time can make the dependencies much easier to handle, however, than for networks observed at one time point.

Three basic distinctions can be made between statistical models for network dynamics. First, the ties may have the nature of changeable states, like friendship or enduring collaboration, or of events, like sending a message or spending an evening together. Second, there is a distinction between models where the

changes are being driven by the network itself (which is meaningful only for networks of states, not for networks of events) or by a different, perhaps unobserved, entity. Third, the time variable indexing the dynamic network may be discrete or continuous.

A probabilistic model for network dynamics can be represented generally as a stochastic process $X(t)$ $(t \in \mathcal{T})$, where $X(t)$ is the value of the process at time t and the time domain $\mathcal{T}$ may be discrete, such as an interval of consecutive integers, or continuous, such as an interval of real numbers. All proposed models for network dynamics all are based in some way on Markov chains, which are stochastic processes $X(t)$ for which the earlier past can be considered forgotten in the sense that for any 'present' moment $t_0 \in \mathcal{T}$, the conditional probability distribution of $X(t)$ for all future times $t > t_0$, given its values for the entire past $t_0 \leq t$, depends only on the current value $X(t_0)$.

In some of the models proposed for network dynamics the network itself is a Markov chain. This is applicable to networks of states, not to networks of events; for example, it would hardly be meaningful to entertain a model where the network of all phone calls going on at one particular moment depend as a Markov chain on the network of past phone calls. Other models for network dynamics can be represented as Hidden Markov Models (HMM). These are defined (Cappé et al., 2005) as stochastic processes $X(t)$ for which there exists another stochastic process $A(t)$ which itself is a Markov chain, and such that for any fixed t_0, the conditional probability distribution of $X(t_0)$, given $A(t)$ for all t and given $X(t)$ for all $t \neq t_0$, depends only on $A(t_0)$ and on nothing else.

The three distinctions mentioned above can be related in the following way to these definitions. First, dynamic state networks can be represented in principle

by Markov chains as well as by HMMs, whereas dynamic event networks can be represented by HMMs. In the latter case the underlying network $A(t)$ could either be constructed by aggregating the past observations (and hence be directly observable) or be unobserved. Second, for Markov chains the changes in the network are driven by the network itself, as the most direct representation of the network dynamics being a feedback process or a self-organizing system; for HMMs the changes in the network are being driven by the entity $A(t)$. Third, the time domain $\mathcal{T}$ may be discrete or continuous.

OVERVIEW OF THE FOLLOWING

Many longitudinal social network data sets are network panel data, i.e., two or more repeated measurements on the social network existing between a fixed group of social actors (perhaps give or take a few actors who enter into the network or leave it during the period of study). In this overview attention will be given mainly to network panel data, for networks consisting of states.

3.1 Continuous-time models

Holland and Leinhardt (1977) had the important insight that to represent the feedback occurring in network dynamics, generated by, for example, reciprocation, transitive closure, and degree-related processes such as the Matthew effect, it is fruitful to employ a continuous-time Markov process even though the observations are done at a few discrete time points, and to use only the creation and termination of single ties as the basic events in such a process, with the exclusion of simultaneous changes of more than one tie variable. This allows to reduce complicated observations of network change to a few basic processes. The same principle was proposed for non-network data by Coleman (1964). For dis-

crete data it was elaborated by Kalbfleisch and Lawless (1985) and for continuous data in the literature reviewed by Singer (2008).

With continuous-time processes one can explain, for example, the change from a set of isolated points to a highly connected subgroup as a result of the three basic processes of random tie creation, reciprocation, and transitive closure, operating as a feedback process according to a Markov chain. This was applied by Wasserman (1977, 1979); Leenders (1995, 1996) in models where dyads are assumed to be independent, implying that only reciprocation and homophily are the processes that can be represented. Wasserman (1980) presented a model which represents degree-related processes.

3.2 Actor-oriented models

A model which allows the simultaneous representation of an arbitrary array of processes is the actor-oriented model proposed by Snijders and van Duijn (1997) and Snijders (2001), with a recent tutorial presentation in Snijders et al. (2010).

The term 'actor-oriented' refers to the idea of constructing the model as the result of context-dependent choices made by the actors, following up the suggestion by Emirbayer and Goodwin (1994) to combine structure and agency. Actors are thought to control their outgoing ties. In line with the principles of Holland and Leinhardt (1977), the model is a continuous-time model in which ties are changed only one at a time, and the probabilities of changes depend on the total current network configuration. The frequency of tie changes is modeled by the so-called *rate function* $\lambda_i(x;\alpha)$, which indicates the frequency per unit of time with which actor i gets the opportunity to change an outgoing tie, given the current network state x. The choice of which tie variable to change is modeled

using the *objective function* $f_i(x;\beta)$ which can be interpreted as a measure of how attractive the network state x is for actor i. α and β are statistical parameters. To define the probability of a change, we denote, for a given network x, by $x^{(ij\pm)}$ the network which is identical to x in all tie variables except those for the ordered pair (i,j), and for which the tie variable $i \to j$ in $x^{(ij\pm)}$ is just the opposite of this tie variable in x, in the sense that $x_{ij}^{(ij\pm)} = 1 - x_{ij}$. Further we formally define $x^{(ii\pm)} = x$.

The model operates as follows. Suppose that the current network is x. All actors have independent, exponentially distributed waiting times until the next time point where they are allowed to change one of their outgoing tie variables. Let i be the actor with the shortest waiting time, who therefore is the one to make the next change. Then the probability that the change is from network x to network $x^{(ij\pm)}$ is given by

$$P\{X \text{ changes to } x^{(ij\pm)}\} = \frac{\exp\big(f_i(x^{(ij\pm)};\beta)\big)}{\sum_{h=1}^{n}\exp\big(f_i(x^{(ih\pm)};\beta)\big)} \ . \tag{7}$$

In Snijders (2001) this formula is motivated based on *myopic stochastic optimization* of the objective function, as is often used in game-theoretical models of network formation (e.g., Bala and Goyal, 2000). When this change has been made (if there was a change, which in this model has probability less than 1), the process starts all over again but now from the new state.

Model specification The choice of the rate function $\lambda_i(x;\alpha)$ and the objective function $f_i(x;\beta)$ will reflect the research questions, underlying theory, and substantive knowledge. The rate function often is constant, or dependent on measures of the amount of activity and resources put by actor i in determining or optimizing her network position. The objective function is usually specified as

a linear combination

$$f_i(x;\beta) = \sum_k \beta_k \, s_{ki}(x) \ , \tag{8}$$

where the $s_{ki}(x)$, called 'effects', are functions of the personal network of i. Examples are

- $s_{ki}(x) = \sum_j x_{ij}$ (out-degree), reflecting average degrees;
- $s_{ki}(x) = \sum_j x_{ij} \, x_{ji}$ (number of reciprocated dyads of actor i), reflecting reciprocation;
- $s_{ki}(x) = \sum_{j,h} x_{ij} \, x_{jh} \, x_{ih}$ (number of transitive triplet of actor i), reflecting transitivity;
- $s_{ki}(x) = \sum_j x_{ij} \, x_{+j}$ (sum of in-degrees of actor i's network contacts), reflecting the Matthew effect.

Extensive list of possible effects are given in Snijders (2001) and Snijders et al. (2010).

Use of continuous-time models to represent network panel data Given an initial network for, say, time $t_1 = 0$, the process described above defines a continuous-time Markov chain with time parameter $\{t \geq 0\}$. When network panel data have been observed at time points $t_1 = 0 < t_2 < \ldots < t_M$, for some $M \geq 2$, the dynamics of the process between consecutive time points t_m and t_{m+1} is unobserved, which can be accounted for in the estimation procedure by simulating this dynamics, cf. Snijders (2001). The distribution of $X(t)$ does not need to be stationary, but the transition probability distribution is assumed to be stationary except for any time-changing parameters that may be incorporated in the parameter vectors α and β. It is usual always to include in the rate function a multiplicative parameter that depends on the time interval (t_m, t_{m+1}),

to reflect the total amount of change observed when going from observation $x(t_m)$ to observation $x(t_{m+1})$.

3.3 Dynamic Exponential Random Graph Models

Discrete-time extensions of the Exponential Random Graph Model for observations $x(t_1), x(t_2), \ldots, x(t_M)$ can be formulated by the model

$$\mathrm{P}_\theta\{X(t_1) = x(t_1), \ldots, X(t_M) = x(t_M)\} = \frac{\exp\left(\sum_k \theta_{1k} s_{1k}\left(x(t_1)\right) + \sum_{m=1}^{M-1} \sum_k \theta_{2k}\, s_{2k}(x(t_m), x(t_{m+1}))\right)}{\kappa(\theta)},$$

where the effects $s_{1k}(x(t_1))$ and parameters θ_{1k} are used to represent the distribution of the network at $X(t_1)$, while $s_{2k}(x(t_m), x(t_{m+1}))$ and θ_{2k} represent the conditional distribution of $X(t_{m+1})$ given $X(t_m)$. Here again the distribution of $X(t)$ is not necessarily stationary in t, but the conditional distribution of $X(t_{m+1})$ given $X(t_m)$ is stationary, unless some of the components in s_{2k} depend also on m.

This model was proposed by Robins and Pattison (2001) and further elaborated by Hanneke et al. (2010). It has in principle the same generality and the same difficulties as the Exponential Random Graph Model for single observed networks. It does not have the parsimonious approach of the actor-oriented model which, due to its definition in continuous time, represents network change in terms of its most simple building block: simple tie changes. This will lead to greater model complexity to obtain a good fit, and hence more complicated interpretation, unless the successive networks $x(t_m)$ and $x(t_{m+1})$ are very close to each other.

3.4 Hidden Markov Models

Several kinds of discrete-time Hidden Markov Models have been proposed where the underlying variables $A(t_m)$ are Markov chains for which the marginal distributions as well as the conditional distributions of $A(t_{m+1})$ given $A(t_m)$ are multivariate normal.

One type of such a model is proposed by Xing et al. (2010). This is a dynamic version of the mixed membership model of Airoldi et al. (2008). Their model has two sets of latent variables: probabilities of class membership which may change over time, and probabilities of ties between various classes, which also may change over time. For both of these multivariate normal distributions are being assumed, which are transformed to the required domain of probability vectors.

Sarkar and Moore (2005) generalize the latent Euclidean distance model of Hoff et al. (2002) to longitudinal network data. They use a random walk model for the changes in the latent locations of the nodes.

Another Hidden Markov Model was proposed by Westveld and Hoff (2011) (although they do not use this term). They extend the random effects model with sender, receiver, and reciprocity effects of van Duijn et al. (2004) and Hoff (2005) to a dynamic model with random effects also for sender, receiver, and reciprocity effects over time. For the random effects an autoregressive normal model is assumed – since this is a Markov chain, the resulting model is a HMM.

4 REVIEW AND FORWARD LOOK

During the last 10 years, tremendous developments have taken place in network modeling in general, including statistical inference for network modeling. Most of the models reviewed here have been applied fruitfully in diverse areas of social

science. The challenge of dealing with complex network dependencies in statistical inference now is starting to be met by the models and methods developed recently – of which the models were reviewed here, but not the methods.

This article has focused on two of the basic types of network data: single and longitudinal observations of graphs (directed or nondirected, but this is not a major distinction), interpretable as states. This limitation is argued by noting that this already covers a large domain, it allows illustrating important issues in the representation of network dependencies, and it has many applications. In this last section some connections to other models and other literature are mentioned.

4.1 Other network models

A closely related stream of network models have been developed by researchers with a background in statistical physics and computer science. Widely known models are the Watts-Strogatz small world model (Watts, 1999), which is a model for large networks that combines the features of transitivity, limited degrees, and limited path lengths (geodesics); and the scale-free network model (de Solla Price, 1976; Barabási and Albert, 1999) which yields networks where the degree distribution has a power law distribution, implying that some nodes will have very large degrees – the 'hubs' in the network. This stream of literature is reviewed, e.g., in Newman et al. (2002), Watts (2004), and more recently in Toivonen et al. (2009). These models may be regarded as micro-macro models in the sense that they are built on simple rules for the formation of ties and study the network-level structures that are generated. The resulting insights have percolated into the literature on statistical modeling of social networks, as illustrated by Robins et al. (2005) who showed how models with small world properties can be ob-

tained from Exponential Random Graph Models, and the importance attached to degree-related effects in stochastic actor-oriented models in Snijders et al. (2010). Models from the physics as well as statistical backgrounds are treated in depth by Kolaczyk (2009).

Network modeling in economics has focused on optimal network structures when actors have relatively simple utility functions, e.g., with a cost on links and a benefit for reaching other nodes indirectly. A review of network formation models in economics is given by Jackson (2005), and the extensive work in this area has led to three recent books: Goyal (2007), Vega-Redondo (2007), and Jackson (2008). The book by Vega-Redondo also contains much material on the techniques from statistical mechanics that are used extensively in the physics literature on networks.

A particular feature of many published models for network dynamics assume that nodes are added sequentially, making some ties to earlier created nodes, while ties remain forever once they exist. This helps tractability for deriving mathematical properties, but makes them unsuitable for modeling dynamics of networks involving tie creation as well as tie deletion on a given, fixed node set.

There exist many more statistical models for data with a network structure. A part of this literature is labeled 'machine learning', which is the name used by computer scientist when referring to inferential problems. Extensive reviews are given in Kolaczyk (2009) and Goldenberg et al. (2009).

4.2 Further work

The field reviewed here is in a state of vigorous development, and the models treated are being extended in various ways. One type of extension is for other

types of network structure: valued graphs, signed graphs, bipartite graphs, etc. For bipartite networks, for example, Exponential Random Graph Models were developed by Wang et al. (2009) and actor-oriented models by Koskinen and Edling (2010). Multivariate Exponential Random Graph Models were discussed in Koehly and Pattison (2005), while actor-oriented models for multivariate networks were proposed by Snijders et al. (2011). Network structure can also be combined with other structures as dependent variables, as in the actor-oriented model for the co-evolution of networks and behavior of Steglich et al. (2010).

Another type is the combination of several of the principles reviewed here. An example is the combination by Krivitsky et al. (2009) of latent Euclidean distances, latent sender and receiver effects, and covariate effects, which are different elements in the framework of latent space models. Latent space elements could also be combined with Exponential Random Graph Models or actor-oriented models.

5 REFERENCES

References

Airoldi, E., D. Blei, S. Fienberg, and E. Xing (2008). Mixed membership stochastic blockmodel. *Journal of Machine Learning Research 9*, 1981–2014.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*, 1181–1229.

Barabási, A. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*, 509–512.

Bearman, P., J. Moody, and K. Stovel (2004). Chains of affection. *American Journal of Sociology 110*, 44–91.

Borgatti, S. and M. Everett (1999). Models of core/periphery structures. *Social Networks 21*, 375–395.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. New York: Springer.

Clauset, A., C. Moore, and M. E. J. Newman (2008). Hierarchical structure and the prediction of missing links in networks. *Nature 453*, 98–101.

Coleman, J. (1961). *The Adolescent Society*. New York: The Free Press of Glencoe.

Coleman, J. (1964). *Introduction to Mathematical Sociology*. New York: The Free Press of Glencoe.

Coleman, J. (1990). *Foundations of Social Theory*. Cambridge/London: Belknap Press of Harvard University Press.

Cornfield, J. and J. W. Tukey (1956). Average values of mean squares in factorials. *Annals of Mathematical Statistics 27*, 907–949.

Dahmström, K. and P. Dahmström (1993). Ml-estimation of the clustering parameter in a markov graph model. Technical report, Department of Statistics, University of Stockholm, Stockholm.

Daudin, J.-J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistical Computing 18*, 173–183.

Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review 35*, 843–852.

de Solla Price, D. (1976). A general theory of bibliometric and other advantage processes. *Journal of the American Society for Information Science 27*, 292–306.

De Vries, H. (1998). Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. *Animal Behaviour 55*, 827–843.

Dekker, D., D. Krackhardt, and T. A. B. Snijders (2007). Sensitivity of mrqap tests to collinearity and autocorrelation conditions. *Psychometrika 72*, 563–581.

Emerson, R. M. (1972). Exchange theory, part ii: Exchange relations and networks. In J. Berger, M. Z. Jr., and B. Anderson (Eds.), *Sociological Theories in Progress*, Volume 2, pp. 58–87. Boston: Houghton-Mifflin.

Emirbayer, M. and J. Goodwin (1994). Network analysis, culture, and the problem of agency. *American Journal of Sociology 99*, 1411–1454.

Feld, S. (1982). Structural determinants of similarity among associates. *American Sociological Review 47*, 797–801.

Fienberg, S. and S. Wasserman (1981). Categorical data analysis of single sociometric relations. In S. Leinhardt (Ed.), *Sociological Methodology*, pp. 156–192. San Francisco: Jossey - Bass.

Frank, O. (1980). Transitivity in stochastic graphs and digraphs. *Journal of Mathematical Sociology 7*, 199–213.

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

Frank, O. (2009). Estimation and sampling in social network analysis. In B. Meyers (Ed.), *Encyclopedia of Complexity and System Science*. Springer.

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

Freeman, L. C. (1979). Centrality in networks: I. conceptual clarification. *Social Networks 1*, 215–239.

Freeman, L. C. (1992). The sociological concept of group: An empirical test of two models. *American Journal of Sociology 98*, 152–166.

Gile, K. J. and M. S. Handcock (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology 40*, xxx–xxx.

Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoldi (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning 2*, 129–233.

Goyal, S. (2007). *Connections; An Introduction to the Economics of Networks*. Princeton: Princeton University Press.

Gulati, R. and M. Gargiulo (1999). Where do interorganizational networks come from? *American Journal of Sociology 104*, 1439–1493.

Handcock, M. S. (2002). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, and P. E. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pp. 229–240. National Research Council of the National Academies. Washington, DC: The National Academies Press.

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

Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based clus-

tering for social networks (with discussion). *Journal of the Royal Statistical Society, Series A 170*, 301–354.

Hanneke, S., W. Fu, and E. P. Xing (2010). Discrete temporal models for social networks. *Electronic Journal of Statistics 4*, 585–605.

Harris, K., F. Florey, J. Tabor, P. Bearman, J. Jones, and J. Udry (2003). The national longitudinal study of adolescent health: Research design. Technical report, University of North Carolina.

Hoff, P. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association 100*, 286–295.

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

Holland, P., K. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: Some first steps. *Social Networks 5*, 109 – 137.

Holland, P. W. and S. Leinhardt (1976). Local structure in social networks. *Sociological Methodology 6*, 1–45.

Holland, P. W. and S. Leinhardt (1977). A dynamic model for social networks. *Journal of Mathematical Sociology 5*, 5–20.

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

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks 29*, 216–230.

Jackson, M. O. (2005). A survey of models of network formation: stability and

efficiency. In G. Demange and M. Wooders (Eds.), *Group Formation in Economics; Networks, Clubs and Coalitions*, pp. 11–57. Cambridge, U.K.: Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.

Kalbfleisch, J. and J. Lawless (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association 80*, 863–871.

Koehly, L. M. and P. Pattison (2005). Random graph models for social networks: multiple relations or multiple raters. In P. Carrington, J. Scott, and S. Wasserman (Eds.), *Models and methods in social network analysis*, Chapter 9, pp. 162–191. New York: Cambridge University Press.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.

Koskinen, J. and C. Edling (2010). Modelling the evolution of a bipartite network–peer referral in interlocking directorates. *Social Networks In Press, Corrected Proof*. http://dx.doi.org/10.1016/j.socnet.2010.03.001.

Koskinen, J., D. Lusher, and G. Robins (2011). *Exponential Random Graph Models*. Cambridge: Cambridge University Press.

Krackhardt, D. (1987). Qap partialling as a test of spuriousness. *Social Networks 9*, 171–186.

Krackhardt, D. (1988). Predicting with networks - nonparametric multiple-regression analysis of dyadic data. *Social Networks 10*(4), 359–381.

Krivitsky, P. N., M. S. Handcock, A. E. Raftery, and P. D. Hoff (2009). Repre-

senting degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks 31*, 204–213.

Lazarsfeld, P. and N. Henry (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lazarsfeld, P. F. and R. K. Merton (1954). Friendship as social process. In M. Berger, T. Abel, and C. Page (Eds.), *Freedom and Control in Modern Society*, pp. 18–66. New York: Van Nostrand.

Leenders, R. (1996). Evolution of friendship and best friendship choices. *Journal of Mathematical Sociology 21*, 133 – 148.

Leenders, R. T. A. J. (1995). Models for network dynamics: A Markovian framework. *Journal of Mathematical Sociology 20*, 1–21.

Lehmann, E. and J. Romano (2005). *Testing Statistical Hypotheses* (3rd ed.). New York: Springer.

Lindgren, K.-O. (2010). Dyadic regression in the presence of heteroscedasticity – an assessment of alternative approaches. *Social Networks 32*, 279–289.

Lorrain, F. and H. White (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology 1*, 49–80.

McDonald, J. W., P. W. Smith, and J. J. Forster (2007). Markov chain monte carlo exact inference for social networks. *Social Networks 29*, 127–136.

McPherson, J. M., L. Smith-Lovin, and J. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 415–444.

Mogapi, O. (2009). *Partial order models for social networks*. Ph. D. thesis, Department of Statistics, University of Oxford.

Molenaar, I. (1988). Formal statistics and informal data analysis or why laziness should be discouraged. *Statistica Neerlandica 42*, 83–90.

Molm, L. D., J. L. Collett, and D. R. Schaefer (2007). Building solidarity through generalized exchange: A theory of reciprocity. *American Journal of Sociology 113*, 205–242.

Moreno, J. (1934). *Who shall survive? A new approach to the problem of human inter-relations*. New York: Beacon House.

Newman, M. E. J., D. J. Watts, and S. H. Strogatz (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences USA 99*, 2566–2572.

Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association 96*, 1077–1087.

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

Rao, A., R. Jana, and S. Bandyopadhyay (1996). A markov chain monte carlo method for generating random (0,1) matrices with given marginals. *Sankhya, ser. A 58*, 225–242.

Rapoport, A. (1953a). Spread of information through a population with sociostructural bias: I. assumption of transitivity. *Bulletin of Mathematical Biophysics 15*, 523–533.

Rapoport, A. (1953b). Spread of information through a population with sociostructural bias: Ii. various models with partial transitivity. *Bulletin of Mathematical Biophysics 15*, 535–546.

Rinaldo, A., S. E. Fienberg, and Y. Zhou (2009). On the geometry of discrete

exponential families with application to exponential random graph models. *Electronic Journal of Statistics 3*, 446–484.

Roberts, Jr., J. (2000). Simple methods for simulating sociomatrices with given marginal totals. *Social Networks 22*, 273–283.

Robins, G. and P. Pattison (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology 25*, 5–41.

Robins, G., P. Pattison, Y. Kalish, and D. Lusher (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks 29*, 173–191.

Robins, G., P. Pattison, and P. Wang (2009). Closure, connectivity and degree distributions: Exponential random graph (p^*) models for directed social networks. *Social Networks 31*, 105–117.

Robins, G., T. A. B. Snijders, P. Wang, M. Handcock, and P. Pattison (2007). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks 29*, 192–215.

Robins, G., J. Woolcock, and P. Pattison (2005). Small and other worlds: Global network structures from local processes. *American Journal of Sociology 110*, 894–936.

Sarkar, P. and A. W. Moore (2005). Dynamic social network analysis using latent space models. *SIGKDD Explorations 7*, 31–40.

Schweinberger, M. and T. A. B. Snijders (2003). Settings in social networks: A measurement model. *Sociological Methodology 33*, 307–341.

Simmel, G. (1917 [1950]). Individual and society. In K. Wolff (Ed.), *The Sociology of Georg Simmel*. New York: The Free Press.

Singer, H. (2008). Nonlinear continuous time modeling approaches in panel research. *Statistica Neerlandica 62*, 29–57.

Snijders, T. and K. Nowicki (1994). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification 14*, 75 – 100.

Snijders, T. A. B. (1991). Enumeration and simulation methods for 0-1 matrices with given marginals. *Psychometrika 56*, 397–417.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. In M. E. Sobel and M. P. Becker (Eds.), *Sociological Methodology – 2001*, Volume 31, pp. 361–395. Boston and London: Basil Blackwell.

Snijders, T. A. B. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure 3*(2). Available from http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html.

Snijders, T. A. B. (2010). Conditional marginalization for exponential random graph models. *Journal of Mathematical Sociology 34*.

Snijders, T. A. B. and C. Baerveldt (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology 27*, 123–151.

Snijders, T. A. B. and R. Bosker (2011). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Snijders, T. A. B., A. Lomi, and V. Torlò (2011). Multiplex dynamics of one-mode and two-mode networks, with an application to friendship and employment preference. (in press).

Snijders, T. A. B., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006).

New specifications for exponential random graph models. *Sociological Methodology 36*, 99–153.

Snijders, T. A. B., G. G. van de Bunt, and C. E. G. Steglich (2010). Introduction to actor-based models for network dynamics. *Social Networks 32*, 44–60.

Snijders, T. A. B. and M. A. J. van Duijn (1997). Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, and P. Terna (Eds.), *Simulating Social Phenomena*, pp. 493–512. Berlin: Springer.

Spreen, M. (1992). Rare populations, hidden populations, and link-tracing designs: What and why? *Bulletin de Méthodologie Sociologique 36*, 34–58.

Steglich, C., P. Sinclair, J. Holliday, and L. Moore (2011). Actor-based analysis of peer influence in a stop smoking in schools trial (assist). *Social Networks 33*, xxx–xxx.

Steglich, C. E. G., T. A. B. Snijders, and M. Pearson (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology 40*, 329–393.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate Behavioral Research 44*, 711 – 740.

Toivonen, R., L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki, and K. Kaski (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks 31*, 240–254.

van Duijn, M. A., K. J. Gile, and M. S. Handcock (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks 32*, 52–62.

van Duijn, M. A. J., T. A. B. Snijders, and B. H. Zijlstra (2004). p_2 : A random effects model with covariates for directed graphs. *Statistica Neerlandica 58*, 234–254.

Vega-Redondo, F. (2007). *Complex Social Networks*. Cambridge, U.K.: Cambridge University Press.

Verhelst, N. D. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika 73*(4), 705–728.

Wang, P., K. Sharpe, G. L. Robins, and P. E. Pattison (2009). Exponential random graph (p^*) models for affiliation networks. *Social Networks 31*, 12–25.

Wasserman, S. (1977). Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology 5*, 61–86.

Wasserman, S. (1979). A stochastic model for directed graphs with transition rates determined by reciprocity. In K. F. Schuessler (Ed.), *Sociological Methodology 1980*. San Francisco: Jossey-Bass.

Wasserman, S. (1980). Analyzing social networks as stochastic processes. *Journal of the American Statistical Association 75*, 280–294.

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

Wasserman, S. and P. Pattison (1996). Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika 61*, 401–425.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology 105*, 493–527.

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology 30*, 243–270.

Westveld, A. H. and P. D. Hoff (2011). A mixed effects model for longitudinal network (relational) data. *Annals of Applied Statistics (in press)*.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–830.

Xing, E. P., W. Fu, and L. Song (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Annals of Applied Statistics 4*, 535–566.