The Statistical Evaluation of Social Network Dynamics

Tom A.B. Snijders
ICS, Department of Statistics and Measurement Theory
University of Groningen *

January 2001

To be published, Sociological Methodology

*Tom A.B. Snijders, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, email <t.a.b.snijders@ppsw.rug.nl>.

Abstract

A class of statistical models is proposed for longitudinal network data. The dependent variable is the changing (or evolving) relation network, represented by two or more observations of a directed graph with a fixed set of actors. The network evolution is modeled as the consequence of the actors making new choices, or withdrawing existing choices, on the basis of functions, with fixed and random components, that the actors try to maximize. Individual and dyadic exogenous variables can be used as covariates. The change in the network is modeled as the stochastic result of network effects (reciprocity, transitivity, etc.) and these covariates. The existing network structure is a dynamic constraint for the evolution of the structure itself. The models are continuous-time Markov chain models that can be implemented as simulation models. The model parameters are estimated from observed data. For estimating and testing these models, statistical procedures are proposed which are based on the method of moments. The statistical procedures are implemented using a stochastic approximation algorithm based on computer simulations of the network evolution process.

1. INTRODUCTION

Social networks represent relations (e.g., friendship, esteem, collaboration, etc.) between actors (e.g., individuals, companies, etc.). This paper is concerned with network data structures in which all relationships within a given set of n actors are considered. Such a network can be represented by an $n \times n$ matrix $x = (x_{ij})$, where x_{ij} represents the relation directed from actor i to actor j $(i, j = 1, ..., n)$. Only dichotomous relations are considered here: the relation from i to j either is present, denoted $x_{ij} = 1$, or absent, denoted $x_{ij} = 0$. Self-relations are not considered, so that the diagonal values x_{ii} are meaningless. They are formally defined as $x_{ii} = 0$. This x is the adjacency matrix of the directed graph by which the network can be represented, and it is also called the *sociomatrix*.

More specifically, we consider longitudinal data on entire networks. It is supposed that the data available is a time-series $x(t), t \in \{t_1, ..., t_M\}$ of social networks for a constant set $\{1, ..., n\}$ of actors. The observation times are ordered, i.e., $t_1 < t_2 < ... < t_M$. The number M of time points is at least 2. The purpose of the statistical analysis is to obtain an insight in the evolution of the network, where the initial state $x(t_1)$ is taken for granted.

Longitudinal social network data are a complex data structure, requiring complex methods of data analysis for a satisfactory treatment. Holland

and Leinhardt (1977a, 1977b) and Wasserman (1977) already proposed to use continuous-time Markov chains as a model for longitudinal social networks. In a continuous-time model, time is assumed to flow on continuously, although observations are available only at the discrete time points t_1 to t_M, and between the observations the network is assumed to change unobserved at random moments as time progresses. Continuous-time models offer, in principle, greater flexibility than the discrete-time Markov chain models elaborated, e.g., by Katz and Proctor (1959), Wasserman (1987), and Wasserman and Iacobucci (1988).

A basic continuous-time Markov chain model for dichotomous social networks, the reciprocity model, was elaborated by Wasserman (1977, 1979, 1980) and further investigated by Leenders (1995a, 1995b) and Snijders (1999). This model is limited because it assumes *dyad independence*. A dyad is defined as the pair (x_{ij}, x_{ji}) of relations between two actors i and j. Dyad independence means that the dyads $(X_{ij}(t), X_{ji}(t))$ evolve as mutually independent Markov chains. This assumption effectively allows to change the analysis from the level of the network to the level of the dyad. This is computationally attractive, but does not leave much room for realistic statistical modeling. Effects related to dependence in the relations between sets of three or more actors, e.g., transitivity ("a friend of my friend is my friend"), cannot be represented by models with dyad independence. Other continuous-time models for social network evolution were proposed by Wasserman (1980) and Mayer (1984), but to allow parameter estimation these models also were very restrictive.

Markov chain Monte Carlo ("MCMC") methods can be used to develop statistical procedures for quite general probability models for the evolution of social networks, provided that these models can be implemented as stochastic simulation models. This was proposed by Snijders (1996) for data defined by sociometric rankings. Snijders and Van Duijn (1997) sketched how this approach can be used for dichotomous social network data. They also indicated how such an actor-oriented model must be specified in order to obtain the dyad-independent models of Wasserman and Leenders. Empirical applications of these stochastic actor-oriented models were presented in Van de Bunt (1999) and Van de Bunt, Van Duijn, & Snijders (1999). The present paper extends this method to data observed at more than two time points, specifies a more efficient and simpler stochastic approximation algorithm, and presents a wider array of effects that can be included in the model.

The basic idea for our model for social network evolution is that the actors in the network may evaluate the network structure and try to obtain a "pleasant" (more neutrally stated, "positively evaluated") configuration

of relations. The actors base their choices in the network evolution on the present state of the network, without using a memory of earlier states. However, they are assumed to have full knowledge of the present network. This represents the idea that actors pursue their own goals under the constraints of their environment, while they themselves constitute each others' changing environment (cf. Zeggelink, 1994). It is immaterial whether this "network optimization" is the actors' intentional behavior; the only assumption is that the network can be modeled *as if* each actor strives after such a positively evaluated configuration. This evaluation is defined as a function of the network as regarded from the perspective of the focal actor, and depends on parameters which are to be estimated from the data. This approach to network evolution is in line with the theoretical sociological principle of methodological individualism, and was referred to by Snijders (1996) as a *stochastic actor-oriented model*. The evaluation includes a random element to account for the deviation between theoretical expectation and observed reality, which leads to a kind of random utility model (cf. random utility models commonly used in econometrics and treated, e.g., in Maddala, 1983). The models can be implemented as stochastic simulation models, which is the basis for the MCMC procedure for parameter estimation. This is a frequentist procedure, using the method of moments. The MCMC implementation of the method of moments uses a stochastic approximation algorithm which is a descendant of the Robbins-Monro (1951) algorithm.

2. CONTINUOUS-TIME MARKOV CHAINS

This section gives a brief introduction to continous-time Markov chains. Karlin and Taylor (1975) and Norris (1997) give general treatments of this kind of stochastic process models. More elaborate introductions to continuous-time Markov chain models for social networks are given by Leenders (1995b) and Wasserman (1979, 1980).

The available data are assumed to be two or more observations of social networks; but the present section is phrased, more generally, in terms of an arbitrary finite outcome space $\mathcal{Y}$. The finitely many observation times t_1 to t_M are embedded in a continuous set of time points $\mathcal{T} = [t_1, t_M] = \{t \in \mathbb{R} \mid t_1 \leq t \leq t_M\}$. Thus it is assumed that changes can take place unobserved between the observation moments. This is not unrealistic and allows a more versatile and natural mathematical treatment.

Suppose that $\{Y(t) \mid t \in \mathcal{T}\}$ is a stochastic process where the $Y(t)$ have a finite outcome space $\mathcal{Y}$ and the time parameter t assumes values in a bounded or unbounded interval $\mathcal{T} \subset \mathbb{R}$. Such a stochastic process is a Markov process

or Markov chain if for any time $t_a \in \mathcal{T}$, the conditional distribution of the future, $\{Y(t) \,|\, t > t_a\}$ given the present and the past, $\{Y(t) \,|\, t \leq t_a\}$, is a function only of the present, $Y(t_a)$. This implies that for any possible outcome $x \in \mathcal{Y}$, and for any pair of time points $t_a < t_b$,

$$\begin{aligned} &\mathrm{P}\{Y(t_b) = x \mid Y(t) = y(t) \text{ for all } t \leq t_a\} \\ &\quad = \mathrm{P}\{Y(t_b) = x \mid Y(t_a) = y(t_a)\} \ . \end{aligned} \tag{1}$$

The Markov chain is said to have a stationary transition distribution if the probability (1) depends on the time points t_a and t_b only as a function of the elapsed time in between, $t_b - t_a$. It can be proven that if $\{Y(t) \,|\, t \in \mathcal{T}\}$ is a continuous-time Markov chain with stationary transition distribution, then there exists a function $q : \mathcal{Y}^2 \to \mathbb{R}$ such that

$$\begin{aligned} q(x,y) &= \lim_{dt \downarrow 0} \frac{\mathrm{P}\{Y(t+dt) = y \mid Y(t) = x\}}{dt} \text{ for } y \neq x \\ q(x,x) &= \lim_{dt \downarrow 0} \frac{1 - \mathrm{P}\{Y(t+dt) = x \mid Y(t) = x\}}{dt} \ . \end{aligned} \tag{2}$$

This function q is called the intensity matrix or the infinitesimal generator. The element $q(x,y)$ is referred to as the *rate* at which x tends to change into y. More generally, an event is said to happen at a rate r, if the probability that it happens in a very short time interval $(t, t+dt)$ is approximately equal to $r\,dt$.

The simultaneous distribution of the Markov chain $\{Y(t) \mid t \geq t_a\}$ with stationary transition distribution is determined completely by the probability distribution of the initial value $Y(t_a)$ together with the intensity matrix. Specifically, the transition matrix

$$P(t_b - t_a) = \Big(\mathrm{P}\{Y(t_b) = y \mid Y(t_a) = x\} \Big)_{x,y \,\in\, \mathcal{Y}}$$

is defined by

$$P(t) = e^{Qt} \ ,$$

where Q is the matrix with elements $q(x,y)$ and the matrix exponential is defined by

$$e^{Qt} = \sum_{h=0}^{\infty} \frac{Q^h t^h}{h!} \ .$$

The reasons for specializing the model to Markov processes with stationary transition distributions are that such models often are quite natural, and that they lend themselves well for computer simulation. The resulting dynamic computer simulation models can be regarded as a type of discrete event simulation models as discussed by Fararo and Hummon (1994).

3. STOCHASTIC ACTOR-ORIENTED MODELS FOR NETWORK EVOLUTION: SIMPLE SPECIFICATION

The specification of the model developed in this paper has three ingredients: the rate function, the objective function, and the gratification function. A simple specification is determined by only the objective function, with a constant rate function and a gratification function equal to zero. The model is explained first for this simple specification. The rate and gratification functions are treated in a later section.

3.1. Basic model ingredients

The class of all sociomatrices, i.e., of all $n \times n$ matrices of 0-1 elements with a zero diagonal, is denoted by $\mathcal{X}$. Note that $\mathcal{X}$ has $2^{n(n-1)}$ elements, a number which is so huge that analytical calculations based on the intensity matrix will be out of the question for most purposes.

It is assumed that each actor "controls" his outgoing relations, which are collected in the row vector $(X_{i1}(t), ..., X_{in}(t))$ of the sociomatrix. Each actor has the opportunity to change his outgoing relations at stochastic times; in the interval between the observation moments t_m and t_{m+1} these opportunities occur at a rate ρ_m. When an actor changes his outgoing relations, he is assumed to strive after a rewarding configuration for himself in the network. This goal is modeled in the so-called *objective function* f discussed below, to which a random component is added, representing the actor's drives that are not explicitly modeled. The actors are assumed to have all information required to calculate their own objective function. This information can be extensive or limited, depending on the model.

At any single time point, at most one actor may change his outgoing relations. Furthermore, he may change only one relation at the time. Of course, many small changes between two observation times can result in a big difference between the two observed networks. The fact that the model specification focuses on changes of single relations is the major reason why continuous time modeling is relatively straightforward. (An example of a continuous-time model for social networks where more than one relation can change at one time point is given by Mayer, 1984.) It should be noted that the fact that the actors take into account the present network structure which is common to them all, introduces a high degree of interdependence between them (when one marginalizes out, rather than conditions upon, the current network structure).

3.2. Objective function

The objective function for actor i is denoted by

$$f_i(\beta, x)\,, \quad x \in \mathcal{X}\,, \tag{3}$$

and indicates the degree of satisfaction for actor i inherent in the relational situation represented by x. This function depends on a parameter vector β. In the simple model specification of this section, the parameter of the statistical model is $\theta = (\rho, \beta)$, where $\rho = (\rho_1, ..., \rho_{M-1})$ is the vector of change rates during the time periods from t_m to t_{m+1} $(m = 1, ..., M-1)$.

Suppose that at some moment t, actor i has the opportunity to change his outgoing relations. At this moment, actor i determines the other actor j with whom he will change his relation x_{ij}. If immediately before time t actor i does have a relation to actor j, then a change implies withdrawing the relation; if immediately before time t actor i does *not* have a relation to actor j, then a change implies initiating the relation. Given the present state x of the network, the network that results when the single element x_{ij} is changed into $1 - x_{ij}$ (i.e., from 0 to 1 or from 1 to 0), is denoted by $x(i \rightsquigarrow j)$. Note that $x(i \rightsquigarrow j)$ refers to an entire adjacency matrix. When the current network is x, actor i has the choice between $x(i \rightsquigarrow j)$ for all possible $j = 1, ..., n$, $j \neq i$. It is assumed that actor i chooses the j that maximizes the value of his objective function $f_i(\beta, x(i \rightsquigarrow j))$ plus a random element,

$$f_i(\beta, x(i \rightsquigarrow j)) \; + \; U_i(t, x, j)\,. \tag{4}$$

The term $U_i(t, x, j)$ is a random variable, indicating the part of the actor's preference that is not represented by the systematic component f_i. It is assumed that these random variables are independent and identically distributed for all i, t, x, j. The assumption that the actor tries to maximize (4), which refers to the state obtained immediately after making this single choice, can be regarded as an assumption of myopia: the actor does not consider the longer-term, or indirect, effects of his choices.

3.3. Markov chain with random utility component

These functions are used in the following way to define a continuous-time Markov chain $X(t)$ with the finite outcome space $\mathcal{X}$.

Events, i.e., changes of the network structure, take place at discrete time points; in between these points, the network structure remains constant. The process is modeled as being right-continuous: if a change takes place from state x_0 to state x_1 at time t_0, then there is an $\epsilon > 0$ such that $X(t) = x_0$ for $t_0 - \epsilon < t < t_0$, while $X(t) = x_1$ for $t_0 \leq t < t_0 + \epsilon$.

The actions of the n actors depend only on the current state of the network, not on the history of how this network came into being. Each actor changes his relations one-at-the-time at stochastic moments at a rate ρ_m. This means that at each time point $t \in (t_m, t_{m+1})$, the time until the next change by *any* actor has the negative exponential distribution with parameter $n\rho_m$ and the expected waiting time until the next change by any actor is $1/(n\rho_m)$. When an event occurs, all actors have the same probability $1/n$ to be the one to change one of his outgoing relations. Given that actor i may change an outgoing relation, he chooses to change his relation to that actor j ($j \neq i$) for whom the value of (4) is highest.

It is convenient to let the $U_i(t, x, j)$ have the type 1 extreme value distribution (or Gumbel distribution) with mean 0 and scale parameter 1 (Maddala, 1983). This assumption is commonly made in random utility modeling in econometrics. When this distribution is used, the probability that the given actor i chooses the other actor j for changing the relation x_{ij}, is the multinomial logit expression, cf. Maddala (1983, p. 60),

$$p_{ij}(\theta, x) = \frac{\exp(f_i(\beta, x(i \rightsquigarrow j)))}{\sum_{h=1, h\neq i}^{n} \exp(f_i(\beta, x(i \rightsquigarrow h)))} \qquad (j \neq i). \tag{5}$$

3.4. Intensity matrix

It was mentioned in Section 2 that stationary transition distributions of continuous-time Markov chains are characterized by their intensity matrix. In our case, where relations are allowed to change only one at a time, the intensity matrix can be represented by functions $q_{ij}(x)$, indicating the change rates of x to $x(i \rightsquigarrow j)$ for $j \neq i$. All other change rates are 0. These functions are defined for $i, j = 1, ..., n,\ i \neq j$, as

$$q_{ij}(x) = \lim_{dt \downarrow 0} \frac{\mathrm{P}\{X(t + dt) = x(i \rightsquigarrow j) \mid X(t) = x\}}{dt} \quad . \tag{6}$$

The intensity matrix $q(x, y)$ defined in (2) is related to $q_{ij}(x)$ by

$$q(x, y) = \begin{cases} q_{ij}(x) & \text{if } y = x(i \rightsquigarrow j) \\ 0 & \text{if } x \text{ and } y \text{ differ in more than one element} \\ -\sum_{i \neq j} q_{ij}(x) & \text{if } x = y. \end{cases} \tag{7}$$

Note that directed graphs x and y differ in exactly one element (i, j) if and only if $y = x(i \rightsquigarrow j)$ and $x = y(i \rightsquigarrow j)$.

For the Markov chain in the simple model specification of the present section, $q_{ij}(x)$ is given for time period (t_m, t_{m+1}) by

$$q_{ij}(x) = \rho_m \, p_{ij}(\theta, x) \, . \tag{8}$$

3.5. Specification of the model

The objective function must contain the substantive ingredients of the model, including, e.g., actor attributes and structural properties of the directed graph. Since the actor has direct control only of his outgoing relations, only the dependence of f_i on row i of the adjacency matrix has an influence on the behavior of the model.

A convenient choice for the objective function is to define it as a sum

$$f_i(\beta, x) = \sum_{k=1}^{L} \beta_k \, s_{ik}(x) \,, \tag{9}$$

where the weights β_k are statistical parameters indicating the strength of the corresponding effect $s_{ik}(x)$, controlling for all other effects in the model, and the $s_{ik}(x)$ are relevant functions of the digraph that are supposed to play a role in its evolution. All formulae given below for possible components s_{ik} refer to a contribution to the objective function of actor i, while the other actors to whom i could be related are indicated by j.

Effects can be distinguished according to whether they depend only on the network x – in which case they can be regarded as endogenous network effects – or also on covariates, which are supposed to be determined exogenously. Covariates can be of two kinds: actor-dependent covariates V with values v_i for actor i, or pair-dependent (dyadic) covariates W with values w_{ij} for the ordered pair (i, j). Only constant (i.e., time-independent) covariates are considered.

The following list is a collection of network effects, as possibilities for the functions s_{ik} in (9).

1. *density effect*, defined by the out-degree
   $s_{i1}(x) = x_{i+} = \sum_j x_{ij}$;

2. *reciprocity effect*, defined by the number of reciprocated relations
   $s_{i2}(x) = \sum_j x_{ij} \, x_{ji}$;

3. *popularity effect*, defined by the sum of the in-degrees of the others to whom i is related,
   $s_{i3}(x) = \sum_j x_{ij} \, x_{+j} = \sum_j x_{ij} \sum_h x_{hj}$

4. *activity effect*, defined by the sum of the out-degrees of the others to whom i is related,
   $s_{i4}(x) = \sum_j x_{ij} \, x_{j+} = \sum_j x_{ij} \sum_h x_{jh}$

5. *transitivity effect*, defined by the number of transitive patterns in i's relations (ordered pairs of actors (j, h) to both of whom i is related, while also j is related to h),
   $s_{i5}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{jh}$;

6. *indirect relations effect*, defined by the number of actors to whom i is indirectly related (through one intermediary, i.e., at sociometric distance 2),
   $s_{i6}(x) = \#\{j \mid x_{ij} = 0,\, \max_h(x_{ih}\, x_{hj}) > 0\}$;

7. *balance*, defined by the likeness between the out-relations of actor i to the out-relations of the other actors j to whom i is related,

$$s_{i7}(x) = \sum_{j=1}^{n} x_{ij} \sum_{\substack{h=1 \\ h\neq i,j}}^{n} (b_0 - |x_{ih} - x_{jh}|) \;, \tag{10}$$

   where b_0 is a constant included for convenience. If the density effect is included in the model (which normally will be the case), the number b_0 can be chosen so as to obtain the clearest interpretation without essentially changing the model specification.

   E.g., to have a balance effect that is not too strongly correlated with the density effect, the number b_0 in (10) can be chosen so that the average of the second sum in this equation over all actors and over the first $M - 1$ time points is 0, i.e.,

$$b_0 = \frac{1}{(M-1)n(n-1)(n-2)} \sum_{m=1}^{M-1} \sum_{i,j=1}^{n} \sum_{\substack{h=1 \\ h\neq i,j}}^{n} |x_{ih}(t_m) - x_{jh}(t_m)| \,. \tag{11}$$

This list can be extended, in principle, indefinitely. Potentially important additional types of effect are *non-linear effects*, i.e., non-linear functions of s_{ik} defined above, the out-degree x_{i+} being the primary candidate for such a non-linear transformation; and other *subgraph counts* in which actor i is involved, of which the reciprocity and transitivity effects are examples.

In practically all applications it will be advisable to include the density effect, because the other effects listed above should be controlled for the density effect. The reciprocity effect is so fundamental in social relations that it is advisable also to include this effect in most applications.

The transitivity and balance effects, and the indirect relations effect when it has a negative weight, all are different mathematical specifications of the intuitive idea that actor i has a 'closed' or transitive personal network, i.e.,

the others to whom i is related tend to have comparatively many relations among themselves. Verbal theories will not often be detailed enough to distinguish between these effects. It can be determined empirically if one or some of these three effects succeed better than the others in accounting for the observed degree of closure, or transitivity, in the data.

For each actor-dependent covariate V there are the following three basic potential effects. (The notation for the functions s_{ik} does not explicitly indicate their dependence on the covariate values v_j.)

8. *covariate-related popularity*, defined by the sum of the covariate over all actors to whom i has a relation,
   $s_{i8}(x) = \sum_j x_{ij}\, v_j$;

9. *covariate-related activity*, defined by i's out-degree weighted by his covariate value,
   $s_{i9}(x) = v_i\, x_{i+}$;

10. *covariate-related dissimilarity*, defined by the sum of absolute covariate differences between i and the others to whom he is related,
    $s_{i10}(x) = \sum_j x_{ij}\, |\, v_i - v_j\, |$.

Positive covariate-related popularity or activity effects will lead to associations between the covariate and the in-degrees and out-degrees, respectively. A negative covariate-related dissimilarity effect will lead to relations being formed especially between actors with similar values on the covariate.

This list can be extended, e.g., by including covariate values in the definitions of the network effects listed above. This represents interactions between the covariate and the network effect.

The main effect for a pair-dependent covariate is

11. *covariate-related preference*, defined by the sum of the values of w_{ij} for all others to whom i is related,
    $s_{i11}(x) = \sum_j x_{ij}\, w_{ij}$.

Here also, the list can be extended by including covariate values in the definition of network effects.

Theoretical insights in the relational process and experience with modeling this type of data have to determine the effects that are included.

4. MOMENT ESTIMATORS

Let the objective function be given by (9), so that the parameter of the statistical model is $\theta = (\rho, \beta)$. The dimensionality of β is denoted L and the total number of dimensions for θ is $K = M - 1 + L$. Analogous to what was proposed for a similar model by Snijders (1996), this parameter can be estimated by the method of moments (explained for general statistical models, e.g., by Bowman and Shenton, 1985). This means that a statistic $Z = (Z_1, ..., Z_K)$ is used, for which θ is determined as the solution of the K-dimensional moment equation

$$\mathcal{E}_\theta Z = z\,, \tag{12}$$

where z is the observed outcome. This moment equation will be specified further by certain ways of conditioning on the initial and intermediate outcomes $x(t_1)$ to $x(t_{m-1})$.

First the choice of the statistic Z is discussed, and then a MCMC algorithm that can be used to approximate the solution of the moment equation.

For the estimation, no assumptions whatsoever are made about the initial state $x(t_1)$. Therefore, the estimation is carried out conditional on this initial state, and this state is not used to obtain any information about the value of the parameter.

In the absence of a formal method such as a reduction to sufficient statistics, the statistics Z_k should be chosen so that they are relevant for the components of the parameter θ in the sense that the expected values of Z_k $(k = 1, ..., K)$ are sensitive to changes in the components of θ. One way to specify this is to require that

$$\frac{\partial \mathcal{E}_\theta Z_k}{\partial \theta_k} > 0 \text{ for all } k.$$

A more stringent specification is to require that this property hold not only for all separate coordinates of the parameter vector, but also for all linear combinations:

$$a' \left(\frac{\partial \mathcal{E}_\theta Z}{\partial \theta} \right) a > 0 \text{ for all } a \in \mathbb{R}^K\,, a \neq 0, \tag{13}$$

where $(\partial \mathcal{E}_\theta Z / \partial \theta)$ is the matrix of partial derivatives. This requirement is far from implying the statistical efficiency of the resulting estimator, but it confers a basic credibility to the moment estimator and it ensures the convergence of the stochastic approximation algorithm mentioned below.

The components of $\theta = (\rho, \beta)$ are the rates of change ρ_m in the time interval (t_m, t_{m+1}) and the weights β_k in the objective function (9). The motivation for the statistics Z_i, at this moment, is of a heuristic nature, based on their obvious connection to the parameters and supported by sufficiency considerations in certain special cases.

For ρ_m, a relevant statistic is the total amount of change in the m'th time period measured by the number of differences between two consecutive observation moments,

$$C_m = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} |X_{ij}(t_{m+1}) - X_{ij}(t_m)| \ . \tag{14}$$

This choice for the statistic relevant for ρ_m can be supported by noting that if $\beta = 0$, which reduces the model to the trivial situation where the $X_{ij}(t)$ are randomly changing 0-1 variables, C_m is a sufficient statistic for ρ_m.

For β_k, a relevant statistic is the sum over all actors i of the digraph statistics s_{ik}, observed at time t_{m+1},

$$S_{mk} = \sum_{i=1}^{n} s_{ik}(X(t_{m+1})) \ . \tag{15}$$

This statistic has an immediate intuitive appeal: if β_k is larger, then the actors strive more strongly to have a high value of s_{ik}, so that it may be expected that S_{mk} will be higher for all m. The statistics S_{mk} are combined over the $M-1$ time intervals by an unweighted sum.

Combining all these proposals, the moment estimator for θ is defined as the solution of the system of equations

$$\mathcal{E}_\theta\{C_m \,|\, X(t_m) = x(t_m)\} \ = \ c_m \ \ (m = 1, ..., M-1) \tag{16}$$

$$\sum_{m=1}^{M-1} \mathcal{E}_\theta\{S_{mk} \,|\, X(t_m) = x(t_m)\} \ = \ \sum_{m=1}^{M-1} s_{mk} \ \ (k = 1, ..., L), \tag{17}$$

where c_m and s_{mk} are the observed outcomes of the statistics C_m and S_{mk}.

Although in our experience these equations mostly seem to have exactly one solution, they do not always have a solution. This can be seen as follows. For a fixed value of β, the left-hand side of (16) is an increasing function of ρ_m, tending to an asymptote which is lower than the maximum possible value of c_m, this maximum being $n(n-1)$. This implies that the method proposed here is not suitable for observations $x(t_m)$ and $x(t_{m+1})$ which are too far apart in the sense of the metric (14). For such observations the dependence of $x(t_{m+1})$ on the initial situation $x(t_m)$ is practically extinguished, and it

may be more relevant to estimate the parameters of the process generating $x(t_{m+1})$ without taking this initial situation into account.

For the trivial submodel where all $X_{ij}(t)$ are independent, the existence of maximum likelihood and moment estimators is discussed in Snijders and Van Duijn (1997).

4.1. Covariance matrix of the estimator

The delta method (see, e.g., Bishop, Fienberg, and Holland, 1973, section 14.6) can be used to derive an approximate covariance matrix for the moment estimator $\hat{\theta}$. (This holds generally for moment estimators, see Bowman and Shenton, 1985, formula (5).) For a homogeneous notation for the parameters ρ_m and β, denote $C_{mm} = C_m$ and formally define $C_{mk} = 0$ for $k \neq m$, and denote

$$Z_m = (C_{m1}, ..., C_{m,M-1}, S_{m1}, ..., S_{mL}) \,.$$

Then the moment equations (16, 17) can be written as

$$\sum_{m=1}^{M-1} \mathcal{E}_\theta\{Z_m \,|\, X(t_m) = x(t_m)\} = \sum_{m=1}^{M-1} z_m \,. \tag{18}$$

Further denote

$$\Sigma_\theta \;=\; \sum_{m=1}^{M-1} \mathrm{cov}\{Z_m \,|\, X(t_m) = x(t_m)\} \tag{19}$$

$$D_\theta \;=\; \frac{\partial}{\partial \theta} \sum_{m=1}^{M-1} \mathcal{E}\{Z_m \,|\, X(t_m) = x(t_m)\} \,. \tag{20}$$

Then it follows from the delta method, combined with the implicit function theorem and the Markov property for the $X(t)$ process, that the approximate covariance matrix of $\hat{\theta}$ is

$$\mathrm{cov}(\hat{\theta}) \approx D_\theta^{-1} \Sigma_\theta {D'_\theta}^{-1} \;. \tag{21}$$

It is plausible that these estimators have approximately normal distributions, although a proof is not yet available. Based on the assumption of normally distributed estimates, the parameters can be tested using the t-ratios defined as the parameter estimate divided by its standard error, referred to a standard normal null distribution. (In other words, the test is carried out as a t-test with infinite degrees of freedom; this test should be regarded as a rough approximation, since no definite results are yet available on the distribution of this test statistic.)

4.2. Conditional moment estimation

The method of moments can be modified by conditioning on the outcomes c_m of C_m ($m = 1, ..., M-1$) rather than using moment equations involving these statistics. This provides a more stable and efficient algorithm and reduces the parameter estimated by the method of moments to the L-dimensional β. This can be helpful especially for larger values of M.

The modified method is based on the property that the distribution of a continuous-time Markov chain $X(t)$ remains invariant when the time parameter is divided by some constant value while the rate parameter is multiplied by the same value. Specifically, when the rate parameter ρ_m obtains for all $t \geq t_m$, then the distribution of $X(t_m + t)$, conditional on $X(t_m)$ and for $t > 0$, depends on ρ_m and t only through their product, $t\rho_m$. The modified method can be loosely described as follows. For each period m independently, the Markov chain is started at time $t = 0$ with the initial value $x^{[m]} = x(t_m)$ and a rate parameter equal to 1. The process is stopped at the first moment t when $\sum_{ij} \mid X_{ij}(t) - x_{ij}^{[m]} \mid = c_m$. This value of t is expected to be close to the product $\rho_m(t_{m+1} - t_m)$ and the statistics observed at this moment are compared with the statistics calculated from observation $x(t_{m+1})$.

To explain this more formally, denote by $X^{(1)}(t)$ a Markov chain evolving according to our model with a fixed and constant rate parameter $\rho = 1$ and a given value of β, and denote by $S_k^{(1)}(t)$ the corresponding statistics (15). Independent replications of this stochastic process, starting at $t = 0$ with $X^{(1)}(0) = x(t_m)$, are used as models for the $M - 1$ periods. Define the statistic

$$C^{(1)}(t) = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} |X_{ij}^{(1)}(t) - X_{ij}^{(1)}(0)| \tag{22}$$

and the stopping time

$$T_m^{\text{fin}} = \min\{t \geq 0 \mid C^{(1)}(t) \geq c_m\} \ . \tag{23}$$

The conditional moment estimator for β is defined as the solution of

$$\sum_{m=1}^{M-1} \mathcal{E}_\beta\{S_k^{(1)}(T_m^{\text{fin}}) \mid X^{(1)}(0) = x(t_m)\} = \sum_{m=1}^{M-1} s_{mk} \quad (k = 1, ..., L) \tag{24}$$

and, given the resulting estimate $\hat{\beta}$, ρ_m is estimated by

$$\hat{\rho}_m = (t_{m+1} - t_m)^{-1} \, \mathcal{E}_{\hat{\beta}}\{T_m^{\text{fin}} \mid X^{(1)}(0) = x(t_m)\} \ . \tag{25}$$

It follows from the general theory of Markov chains that for all possible values of c_m the stopping time T_m^{fin} is finite with probability 1, and even has

a finite expected value. Therefore the difficulties with the definition of the estimator for large values of c_m , as discussed for the unconditional moment estimator, do not arise here. However, this consolation is only theoretical, because in practice, for large t the value of $C^{(1)}(t)$ fluctuates randomly about an asymptote lower than the maximum possible value of $n(n-1)$, and the stopping time T_m^{fin} is indeed finite but horribly large. The simulation-based algorithm, explained below, is not practically feasible for values of c_m larger than this asymptote.

5. STOCHASTIC APPROXIMATION

The moment equations for the two estimation methods are defined by (18) and (24), but the conditional expectations which are central in these equations cannot be calculated explicitly (except for some special and rather trivial cases, as discussed in Snijders and Van Duijn, 1997). However, it is rather straightforward to simulate random digraphs with the desired distributions. Therefore, stochastic approximation methods, in particular, versions of the Robbins-Monro (1951) procedure, can be used to approximate the moment estimates. Introductions to stochastic approximation and the Robbins-Monro algorithm are given, e.g., by Ruppert (1991) and Pflug (1996).

The algorithm to solve the equation (12) is based on a sequence $\hat{\theta}_N$ generated according to the iteration step

$$\hat{\theta}_{N+1} = \hat{\theta}_N \; - \; a_N \, D_0^{-1} \, (Z_N - z) \; , \tag{26}$$

where Z_N is generated according to the probability distribution defined by the parameter value $\hat{\theta}_N$. For a_N, a sequence is used that converges slowly to 0. D_0 is a positive diagonal matrix. In principle, the optimal choice of D_0 might be non-diagonal. However, Polyak (1990), Ruppert (1988), and Yin (1991) (as discussed also by Pflug, 1996, Section 5.1.3, and Kushner and Yin, 1997) showed that if all eigenvalues of the matrix of partial derivatives, $(\partial \mathcal{E}_\theta Z / \partial \theta)$, have positive real parts and certain regularity conditions are satisfied, then convergence at an optimal rate can be achieved when D_0 is the identity matrix, with a_N a sequence of positive numbers converging to 0 at the rate N^{-c}, where $0.5 < c < 1$. To obtain this optimal convergence rate, the solution of (12) must be estimated not by the last value $\hat{\theta}_N$ itself, but by the average of the consecutively generated $\hat{\theta}_N$ values. This algorithm is a Markov chain Monte Carlo algorithm because the iteration rule (26) indeed defines a Markov chain.

The convergence properties of this algorithm hold asymptotically for $N \to \infty$. To have good properties already for relatively low values of N

it is important to specify the algorithm in such a way that it quickly comes close to the target value. This can be achieved by applying a result due to Pflug (1990), who showed that the limiting first order autocorrelation of the sequence $(Z_N - z)$ generated by (26) is negative. This means that as long as the partial sums of successive values of the product $(Z_N - z)'(Z_{N-1} - z)$ are positive, it must be assumed that the sequence $\hat{\theta}_N$ still is drifting toward the limit point rather than wandering around the limit point, so that it is not desirable to decrease the step sizes a_N. Therefore a_N remains constant as long as there still seems to be such a drift going on, except that when N gets too large a_N is decreased anyway, in order to retain the convergence rate N^{-c} for the sequence a_N.

These ideas are combined in the specification of the algorithm as given in the appendix. The algorithm provides an arbitrarily accurate approximation to the solution of (12) as well as an estimate of the covariance matrix (21). It is available in the freeware PC program SIENA (see the discussion section).

6. AN EVOLVING NETWORK OF UNIVERSITY FRESHMEN

As an illustration, data are used of a study by Van De Bunt (1999) which were analyzed also by Van De Bunt, Van Duijn, and Snijders (1999). For a more extensive description of this data set we refer to these publications. In the present paper, this data set is used only as an illustration without paying much attention to the theoretical interpretations.

The actors in this network are a group of 32 university freshmen who were following a common study program in a Dutch university. This group comprised 24 female and 8 male students. The number of observations used here is $M = 3$. The data used here are those for the time points labeled t_2, t_3, and t_4 in Van De Bunt, Van Duijn, and Snijders (1999). There are 3 weeks between time points t_2 and t_3, and also between t_3 and t_4. For the purpose of this illustration, the time points are relabeled t_1, t_2, and t_3. The relation studied is defined as 'at least a friendly relationship', referred to here as a positive relation ($x_{ij} = 1$). The absence of a positive relation is referred to as a null relation ($x_{ij} = 0$).

There is missing data due to non-response, increasing from 9% at t_1 to 19% at t_3. This incompleteness of data is treated in the estimation procedure in the following ad hoc fashion. (It will be important to conduct further studies to evaluate this way of dealing with incomplete data, and compare it with potential alternatives.)

Missing data are treated in a simple way, trying to minimize their influence on the estimation results. The simulations are carried out over all

$n = 32$ actors. In the initial observation $x(t_m)$ for each period, missing entries $x_{ij}(t_m)$ are set to 0. In the course of the simulations, however, these values are allowed to become 1 like any other values $x_{ij}(t)$. For the calculation of the statistics S_{mk} and C_m, the values of $x_{ij}(t_m)$ as well as of $X_{ij}(t_{m+1})$ are set to 0 whenever at least one of the two observations $x_{ij}(t_m)$ and $x_{ij}(t_{m+1})$ is missing.

To get a basic impression of the data, it may be noted that densities (calculated over the available data) at the three observation moments increase from 0.15 via 0.18 to 0.22. The number of observed changes between the observations at t_1 and t_2 was 60 (out of 744 directed pairs (i, j) for which the value of x_{ij} was observed at observations t_1 and t_2); between t_2 and t_3 this was 51 (out of 679 observations).

The first model estimated includes the basic effects of density and reciprocity, together with the three basic triadic effects: transitivity, indirect relations, and balance. The purpose of this stage in the analysis is to investigate which of these triadic effects are empirically supported by these network evolution data. The number b_0 in (10) is defined by (11). The conditional moment estimator was used and the algorithm was specified as described in the appendix, except that to increase precision 5 subphases were carried out in phase 2 and $n_3 = 1000$ steps were made in phase 3. The results are displayed as Model 1 in Table 1.

Table 1: Parameters for models estimated using observations at t_1, t_2, t_3.

	Model 1		Model 2		Model 3	
Effect	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
Rate (period 1)	3.87		3.78		3.91	
Rate (period 2)	3.10		3.14		3.07	
Density	-1.48	(0.30)	-1.05	(0.19)	-1.13	(0.22)
Reciprocity	1.98	(0.31)	2.44	(0.40)	2.52	(0.37)
Transitivity	0.21	(0.11)	–		–	
Balance	-0.33	(0.66)	–		–	
Indirect relations	-0.347	(0.074)	-0.557	(0.083)	-0.502	(0.084)
Gender activity	–		–		-0.60	(0.28)
Gender popularity	–		–		0.64	(0.24)
Gender dissimilarity	–		–		-0.42	(0.24)

The estimated rate parameters, $\hat{\rho}_1 = 3.87$ and $\hat{\rho}_2 = 3.10$, indicate that on average the actors made 3.87 changes of relationships between the first two observations, and 3.10 changes between the last two observations. (This

includes two-way changes between two observations which remained unobserved because they canceled each other.)

As suggested in Section 4.1, the effects are tested by t-statistics defined by the ratio of parameter estimate to standard error, referred to a standard normal distribution. There is a strongly significant reciprocity effect ($t = 1.98/0.31 = 6.39$). Of the three triadic effects, the indirect relations effect is significant ($t = -0.347/0.074 = -4.69$), but the other two are not significant at the 5% level, although the transitivity effect comes close. When the balance effect was deleted from the model, the t-value for the transitivity effect became 1.94 (results not shown here), just short of significance at the 5% level. The results obtained when deleting the two non-significant effects from the model are shown as Model 2 in Table 1. The indirect relations effect becomes larger, and the density and reciprocity effects change, because these effects now also must represent the effects represented by transitivity and balance in Model 1. It can be concluded that there is evidence of a tendency to have closed networks in the sense of a relatively low number of indirect relations; controlling for this effect and for reciprocity, there is no significant tendency toward a high number of transitive triplets or toward balanced relationships. No significant evidence was found for other structural network effects (estimation results not shown here).

As a next step, the three basic effects of gender were included in the model. In the original dataset gender was represented by a dummy variable equal to 0 for women and 1 for men. The means were subtracted from this variable as well as from the dissimilarity variable $| v_i - v_j |$. Given that the proportion of women was 75%, this leads to the variable v_i being -0.25 for women and +0.75 for men, and the dissimilarity variable being -0.387 for equal-gender pairs and 0.613 for unequal-gender pairs. The results for the model including the structural effects of reciprocity and indirect relations as well as the three covariate effects of gender are presented in Table 1 as Model 3. It can be concluded that women are more active in creating positive relations than men ($t = -0.60/0.28 = -2.14$), while men receive more positive choices ($t = 0.64/0.24 = 2.67$), but there are no significant (dis)similarity effects associated with gender. The control for gender does not have an important influence on the reciprocity or indirect relations effects.

The results based on the observations at these three moments can be compared to results based on only two of these observations. This can be used to check the model assumption that the parameter values β_k are in the time interval between t_1 and t_2 the same as between t_2 and t_3. Further, for the analysis of the evolution of the network from t_1 to t_3 this illustrates the greater precision obtainable by including the information about the network at t_2. The comparison is made only for Model 3, and reported in Table 2.

Table 2: Parameter estimates for Model 3, estimated from two observations.

Observations	$t_1,\ t_2$		$t_2,\ t_3$		$t_1,\ t_3$	
Effect	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
Rate	3.64		3.21		5.29	
Density	-0.99	(0.32)	-1.30	(0.28)	-0.78	(0.31)
Reciprocity	2.36	(0.52)	2.89	(0.67)	2.40	(0.48)
Indirect relations	-0.432	(0.113)	-0.653	(0.140)	-0.536	(0.146)
Gender activity	-0.75	(0.40)	-0.39	(0.42)	-0.77	(0.36)
Gender popularity	0.40	(0.31)	1.03	(0.44)	0.36	(0.26)
Gender dissimilarity	-0.35	(0.35)	-0.58	(0.43)	-0.22	(0.31)

None of the estimates are significantly different between the periods t_1–t_2 and t_2–t_3. This supports the use of a common model for the entire period t_1–t_3.

To compare the 'Model 3' column of Table 1 with the 't_1, t_3' column of Table 2, the estimates in the former column are called 'three-observation' and those in the latter column 'two-observation' estimates. It appears that the corresponding estimates differ at most by about one 'two-observation' standard error; for all parameters but one, the 'three-observation' estimates are closer than the 'two-observation' estimates to the mean of the separate estimates for the t_1–t_2 and t_2–t_3 periods. The 'three-observation' standard errors all are clearly smaller than the 'two-observation' standard errors. This provides some support for the expected greater reliability of the 'three-observation' as compared to the 'two-observation' estimates.

7. EXTENDED MODEL SPECIFICATION

The general model specification contains, in addition to the objective function, two other elements: the rate function, representing that actors may differ in the rate at which they change their relations; and the gratification function, representing that various effects may operate differently for the creation of a relation (where x_{ij} goes from 0 to 1) than for its dissolution (x_{ij} changing from 1 to 0).

7.1. Rate function

The rate function for actor i is denoted

$$\lambda_i(\rho, \alpha, x, m) \quad \text{for } x \in \mathcal{X}\,, \tag{27}$$

and indicates the rate at which actor i is allowed to change something in his outgoing relations in the time period $t_m \leq t < t_{m+1}$. In the simple specification given above, this rate function depended only on m and not on i or x, and was defined as $\lambda_i(\rho, \alpha, x, m) = \rho_m$. The roles of the statistical parameters ρ and α are discussed below.

These rate functions and the conditional independence of the actors imply that at each time point t, the time until the next change by *any* actor has the negative exponential distribution with parameter

$$\lambda_+(\rho, \alpha, x, m) = \sum_{i=1}^{n} \lambda_i(\rho, \alpha, x, m)\ ,\ \text{for } x = x(t),\ t_m \leq t < t_{m+1} \tag{28}$$

(provided that this next change still is before time t_{m+1}). The parameter of the negative exponential distribution is taken here as the reciprocal of the expectation, so the expected waiting time until the next change after time t is $1/\lambda_+(\rho, \alpha, x(t), m)$ (where a possible change to the following time interval is not taken into account). Given that a change occurs, the probability that it is actor i who may change his out-relations is

$$\frac{\lambda_i(\rho, \alpha, x, m)}{\lambda_+(\rho, \alpha, x, m)}\ . \tag{29}$$

Non-constant rate functions can depend, e.g., on actor-specific covariates or on network statistics expressing the degree to which the actor is satisfied with the present network structure. Of course the rate function must be restricted to positive values. In order not to burden the specification with too many complications, it is proposed to define the rate function as a product

$$\lambda_i(\rho, \alpha, x, m) = \lambda_{i1}\lambda_{i2}\lambda_{i3}$$

of factors depending, respectively, on period m, actor covariates, and the actor's personal network. The corresponding factors in the rate function are the following:

1. The dependence on the period can be represented by a simple factor

   $$\lambda_{i1} = \rho_m$$

   for $m = 1, ..., M-1$.

2. The effect of actor covariates with values v_{hi} can be represented by the factor

   $$\lambda_{i2} = \exp(\sum_h \alpha_h \, v_{hi}) \, . \tag{30}$$

3. The dependence on the network can be modeled, e.g., as a function of the actor's out-degree, in-degree, and number of reciprocated relations. Define these by

   $$x_{i+} = \sum_j x_{ij}, \;\; x_{+i} = \sum_j x_{ji}, \;\; x_{i(r)} = \sum_j x_{ij} x_{ji}$$

   (recalling that $x_{ii} = 0$ for all i).
   Snijders and Van Duijn (1997) investigated how the rate function should be specified in order to obtain Wasserman's (1979) reciprocity model as a special case. Denoting the corresponding parameter by α_1, for the dependence on the out-degree this led to the factor

   $$\lambda_{i3} = \frac{x_{i+}}{n-1} \exp(\alpha_1) \; + \; (1 - \frac{x_{i+}}{n-1}) \exp(-\alpha_1). \tag{31}$$

   This defines a linear function of the out-degree, parametrized in such a way that it is necessarily positive.
   For a general dependence on the out-degree, in-degree, and number of reciprocated relations, one can use an average of such terms, the second and third one depending on x_{+i} and $x_{i(r)}$, respectively.

It would be interesting to explore other specifications of the rate function, expressing in a theoretically more satisfactory way the circumstances and characteristics upon which it depends how quickly actors change their relations.

7.2. Gratification function

The basic motivation for the third model ingredient, the *gratification function*, is that a given effect may operate more strongly, or less strongly, for the creation than for the dissolution of relations. E.g., it is conceivable that although actors prefer to establish reciprocated relations, they are quite willing to initiate as yet unreciprocated relations; but that, once they have a reciprocated relationship, they are very reluctant to let it go, e.g., because of the investments accumulated in this relation, cf. Van De Bunt (1999). This would mean that the reciprocity effect is greater for dissolution than for creation of ties. Such a difference cannot be represented by the objective function alone. Therefore the model includes also a gratification function

$$g_i(\gamma, x, j)\,, \quad \text{defined for } i, j = 1, ..., n,\ i \neq j,\ x \in \mathcal{X}\,, \tag{32}$$

which indicates the instantaneous gratification experienced by actor i when, from the given network configuration x, element x_{ij} is changed into its opposite, $1 - x_{ij}$.

When a gratification function is included in the model, expression (4) for the momentary objective function maximized by i is replaced by the sum of the actor's preference for the new state, the gratification experienced as a result of the change, and a random element:

$$f_i(\beta, x(i \rightsquigarrow j)) + g_i(\gamma, x, j) + U_i(t, x, j)\,. \tag{33}$$

Using the same assumptions for the random term $U_i(t, x, j)$ as above, the probabilities of the various possible new states $x(i \rightsquigarrow j)$ now are given by

$$p_{ij}(\theta, x) = \frac{\exp(r(\theta, i, j, x))}{\sum_{h=1, h\neq i}^{n} \exp(r(\theta, i, h, x))} \qquad (j \neq i). \tag{34}$$

where

$$r(\theta, i, j, x) = f_i(\beta, x(i \rightsquigarrow j)) + g_i(\gamma, x, j)\ .$$

These probabilities do not change when to $r(\theta, i, j, x)$ a term is added that does not depend on j. It is often more convenient to work with

$$r(\theta, i, j, x) = f_i(\beta, x(i \rightsquigarrow j)) - f_i(\beta, x) + g_i(\gamma, x, j)\ . \tag{35}$$

The instantaneous effect g_i is a more general model component than the objective function f_i , because the objective function depends only on the new state $x(i \rightsquigarrow j)$, whereas the gratification function depends arbitrarily on the new state as well as the old state x. The reason for not working with just the gratification function is that the objective function, attaching a value to

each network configuration, often is conceptually more attractive and better interpretable than the instantaneous gratification effect.

The gratification function can be specified by a weighted sum,

$$g_i(\gamma, x, j) = \sum_{h=1}^{H} \gamma_h \, r_{ijh}(x) \tag{36}$$

for certain statistics $r_{ijh}(x)$, each containing either a factor x_{ij} (if it reflects the gratification involved in withdrawing a relation, i.e., changing x_{ij} from 1 to 0) or a factor $(1 - x_{ij})$ (if the effect is about the gratification involved in creating a relation). Some examples of such terms are the following.

1. $\gamma_1 \, x_{ij} \, x_{ji}$: indicator of a reciprocated relation; a negative value of γ_1 reflects the costs associated with breaking off a reciprocated relation.

2. $\gamma_2 \, (1 - x_{ij}) \sum_h x_{ih} x_{hj}$: the number of actors through whom i is indirectly related to j; a positive value of γ_2 reflects that it is easier to establish a new relation to another actor j if i has many indirect relations to j via others who can serve as an introduction;

3. $\gamma_3 \, x_{ij} \, w_{ij}$: the value w_{ij} for another actor to whom i has a relation; e.g., a negative value of γ_3 reflects the costs for i associated with breaking off an existing relation to other actors j with a high value for w_{ij} .

7.3. Intensity matrix and simulation

The model that includes an arbitrary rate function $\lambda_i(\rho, \alpha, x, m)$, an objective function, and a gratification function, still is a continuous time Markov chain. The intensity matrix $q(x, y)$ still is given by (7), now with

$$q_{ij}(x) = \lambda_i(\rho, \alpha, x, m) \, p_{ij}(\theta, x) \, , \tag{37}$$

where p_{ij} now is given by (34).

Note that it is straightforward to define an algorithm that simulates this stochastic process. Schematically, this can be done as follows. Suppose that the present time point is $t \in [t_m, t_{m+1})$. The time until the next change by any actor is generated by a negative exponential distribution with parameter (28), provided that the moment so determined is before time t_{m+1}. The actor who is to change a relation (i.e., the row of the adjacency matrix in which a change will occur) is actor i with probability (29). The other actor with whom actor i will change the relation (column of the adjacency matrix) is j with probability (34). When j is chosen, element x_{ij} is changed into its opposite, $1 - x_{ij}$.

7.4. Choice of statistics for estimation

The use of the method of moments requires also the selection of statistics that are relevant for the parameters included in the rate and gratification functions.

A tentative choice for statistics to estimate the parameters α_h in (30) is provided by the total amounts of change weighted by v_{hi} ,

$$C_{M+h-1} = \sum_{m=1}^{M} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mid X_{ij}(t_{m+1}) - x_{ij}(t_m) \mid v_{hi} \,. \tag{38}$$

To estimate the parameter α_1 in (31) for the effect of out-degree on rate of change, the statistic

$$C_{M+H} = \sum_{m=1}^{M} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mid X_{ij}(t_{m+1}) - x_{ij}(t_m) \mid x_{i+}(t_m) \tag{39}$$

can be used (where H is the total number of covariates used for modeling the rate function), and similarly for the effects of the in-degree and the number of reciprocated relations. These choices are intuitively plausible and have led to reasonable estimates in some trial data sets, but more research is required.

For the parameters γ_h included in the gratification function (36), a relevant statistic is

$$R_h = \sum_{m=1}^{M-1} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mid X_{ij}(t_{m+1}) - x_{ij}(t_m) \mid r_{ijh}(x(t_m)) \;, \tag{40}$$

which is the sum of the r_{ijh} values of newly formed relations if r_{ijh} contains a factor $(1 - x_{ij})$, and the sum of r_{ijh} values of disappeared relations if r_{ijh} contains a factor x_{ij}.

These statistics C_{M+h} and R_h are used in the method of moments in the same way as $\sum_m S_{mk}$ in (17) and (25).

8. CONTINUATION OF THE EXAMPLE

Continuing the example of the network of university freshmen, the effect (31) of the out-degrees on the rate of change is included, and the gratification function is defined as the sum of the effect of breaking reciprocated relations and the effect of gender difference on breaking a relation,

$$g_i(\gamma, x, j) = \gamma_1 \, x_{ij} \, x_{ji} \; + \; \gamma_2 \, x_{ij} \mid v_i - v_j \mid$$

where v_i indicates the gender of actor i.

Table 3: Parameter estimates for model with rate and gratification effects

	Model 4	
Effect	par.	(s.e.)
Rate (period 1)	5.05	
Rate (period 2)	3.95	
Out-degree effect on rate	0.90	(0.47)
Density	-0.99	(0.20)
Reciprocity	2.82	(0.56)
Indirect relations	-0.508	(0.091)
Gender activity	-0.52	(0.31)
Gender popularity	0.55	(0.30)
Gender dissimilarity	0.08	(0.37)
Breaking reciprocated relation	-0.58	(1.06)
Breaking relation with different-gender other	1.64	(0.62)

The results are given as Model 4 in Table 3. It can be concluded that the tendency of actors with higher out-degrees to change their relations more often is close to significance at the 5% level ($t = 0.90/0.47 = 1.91$), and that relations with other actors of the other sex are terminated more quickly than those with others of the same sex ($t = 1.64/0.62 = 2.65$). The effect of reciprocity on breaking a relation is not different from what may be expected from the main reciprocity effect ($t = -0.58/1.06 = -0.55$). Comparing these results to those for Model 3 in Table 1, it can be concluded that the activity and popularity effect for gender now are somewhat weaker (having lost their significance at the 5% level), and the main gender dissimilarity effect has vanished due to the inclusion of the effect of gender dissimilarity on breaking a relation. Thus Model 4 suggests that friendly relations with others of the other sex are less stable, and that there is no evidence (as one might erroneously conclude from Model 3) that friendly relations are initiated less with others of the different than with those of the same sex.

9. ASYMPTOTIC DISTRIBUTION AND RELATION WITH THE p^* MODEL

If it is possible to reach every state from every given initial state in a finite number of steps (as is the case here), the distribution of a Markov chain with stationary intensity matrix on a finite outcome space tends to a unique limiting distribution as $t \to \infty$, independent of the initial distribution. For a certain specification of our model, this limiting distribution is the p^* model for social networks proposed by Wasserman and Pattison (1996), generalizing the Markov graph distribution proposed by Frank and Strauss (1986). The p^* model is a family of probability distributions for a single observation x on a stochastic directed graph X. The probability distribution for the p^* model is defined by

$$\mathrm{P}\{X = x\} = \frac{\exp(\beta' z(x))}{\kappa(\beta)} \tag{41}$$

where $z(x)$ is a vector of statistics of the digraph and $\kappa(\beta)$ is a normalization factor. The following proposition indicates a specification for the actor-oriented model that yields the p^* distribution as the limiting distribution.

Proposition 1. Define for all i the objective function by

$$f_i(\beta, x) = \beta' z(x) \tag{42}$$

and the gratification function by $g_i = 0$. Furthermore, define the rate function by

$$\lambda_i(x) = \sum_{\substack{h=1 \\ h \neq i}}^{n} \exp(\beta' z(x(i \rightsquigarrow h))) \ . \tag{43}$$

Then the limiting probability distribution of $X(t)$ for $t \to \infty$ is the p^* distribution with probability function (41).

Proof. It follows from (34), (37), and (43) that

$$q_{ij}(x) = \exp(\beta' z(x(i \rightsquigarrow j))) \ .$$

Note that the symbol $x(i \rightsquigarrow j)$ can be understood as the result of taking matrix x and applying the operation of changing x_{ij} into $1 - x_{ij}$. Applying this operation twice returns the original matrix x, which can be represented as $(x(i \rightsquigarrow j))(i \rightsquigarrow j) = x$. Therefore,

$$q_{ij}(x(i \rightsquigarrow j)) = \exp(\beta' z(x))$$

which implies

$$\exp(\beta' z(x))q_{ij}(x) = \exp(\beta' z(x(i \rightsquigarrow j)))q_{ij}(x(i \rightsquigarrow j))$$

and, for Q defined by (7), that

$$\exp(\beta' z(x))q(x, y) = \exp(\beta' z(y))q(y, x)$$

for all x, y. In terms of the theory of Markov chains (e.g., Norris, 1997, p. 124–125), this means that the intensity matrix Q and the distribution (41) are in detailed balance, which implies that (41) is the stationary distribution for Q. Since all states communicate with one another, the stationary distribution is unique and (41) also is the limiting distribution. Q.E.D.

An interpretation of the rate function (43) is that actors for whom changed relations have a higher value, will indeed change their relations more quickly.

10. DISCUSSION

The procedure proposed in this paper provides a method for the analysis of two or more repeated observations on a social network, in which network as well as covariate effects are taken into account. In view of processes in the real-life evolution of social networks, in which endogenous network effects cumulate continuously over time, the continuous-time nature of this model will be attractive in many applications. The procedure is available in the PC program SIENA ("Simulation Investigation for Empirical Network Analysis", available free of charge from http://stat.gamma.rug.nl/snijders/siena.html), which runs under Windows, and is comprised in the StOCNET package (http://stat.gamma.rug.nl/stocnet).

The present article provides the basic procedure, but this methodology could benefit from further elaborations and improvements, e.g., along the following lines. The algorithm has proven to work well in various applications, but it is rather time-consuming and improvements may be possible. A proof of the sufficient condition for its convergence (see the appendix: the eigenvalues of $(D_0^{-1}\, \partial \mathcal{E}_\theta Z/\partial\theta)$ should have positive real parts) still is lacking. The frequency properties of the standard errors and the hypothesis tests are based on large sample approximations and should be investigated. The robustness of the proposed estimates and tests to deviations from the model assumptions is an interesting point for further study. The method of moments was chosen because of its feasibility, but it may be possible to develop other estimation methods for this model. As additions to the toolbox, it would be useful to have measures for goodness of fit and some kind of standardized effect sizes.

The present implementation contains an ad hoc way of dealing with missing data which merits further investigation.

Although the model is presented as an actor-oriented model, it uses an extremely simple and myopic behavioral model for the actors. This simplicity is a strength because more complicated models for the behavior of actors in a relational network would be more restrictive and less general in their domain of applicability. On the other hand, for specific applications it could be interesting to develop statistical network evolution models incorporating a sociologically more interesting behavioral model.

Further extensions are possible. An extension to relations with ordered outcome categories would increase the scope of the model. One could also think of extending the model to include unobserved heterogeneity by means of random effects, but this would lead the model outside of the realm of complete observations of the state of a Markov process, and therefore require more complex estimation methods.

APPENDIX: STOCHASTIC APPROXIMATION ALGORITHM

The purpose of the algorithm is to approximate the solution of the moment equation (12). In this appendix, the solution is denoted by θ_0. As mentioned in the text above, the algorithm uses the idea of Polyak (1990) and Ruppert (1988) to employ a diagonal matrix D_0 in the iteration step (26) and estimate the solution by partial averages of $\hat{\theta}_N$ rather than the last value; and it uses the idea of Pflug (1990) to let the values of a_N remain constant if the average products of successive values $(Z_N - z)(Z_{N-1} - z)$ are positive, since this suggests that the process still is drifting toward its limit value. However, the specification used here deviates from Pflug's proposal by requiring, for the premature decrease of a_N , that for *each* coordinate the partial sum of the product of successive values be negative, rather than requiring this only for the sum over the coordinates. Further, the number of steps for which a_N is constant is bounded between a lower and an upper limit to ensure that a_N is of order N^{-c}.

A crucial condition for the Polyak-Ruppert result about the optimal convergence rate of the partial sums of $\hat{\theta}_N$ to the solution of (12), is the assumption that all eigenvalues of the matrix of partial derivatives, $(D_0^{-1}\, \partial \mathcal{E}_\theta Z / \partial \theta)$, have positive real parts; see Yin (1991), Pflug (1996), or Kushner and Yin (1997). This condition is implied by condition (13) if D_0 is the identity matrix. For our model and the proposed statistics used in the moment estimators we conjecture that this condition is satisfied, but the proof still is a matter of further research. Whether the algorithm yields an estimate that

indeed solves the moment equation (12) to a satisfactory degree of precision is checked in the 'third phase' of the algorithm below. The practical experience with the convergence of the algorithm is, for most models applied to most data sets, quite favorable.

The reason for incorporating the matrix D_0 is to achieve better compatibility between the scales of Z and of θ. The diagonal elements of D_0 are defined as the estimated values of the derivatives $\partial\mathcal{E}_\theta(Z_k)/\partial\theta_k$ where θ is at its initial value. To see that this leads to compatibility of the scales of Z and θ note that in the extreme case where $\mathrm{var}(Z_k) = 0$ and the diagonal elements of D_0 are equal to $\partial\mathcal{E}_\theta(Z_k)/\partial\theta_k$, (26) for $a_N = 1$ is just the iteration step of the Newton-Raphson algorithm applied to each coordinate of Z separately. Thus, beginning the algorithm with a_N in the order of magnitude of 1 will imply that the initial steps have an approximately right order of magnitude.

The algorithm consists of three phases, which can be sketched as follows. The number of dimensions of θ and of Z is denoted by p and the initial value is denoted θ_1.

Phase 1. In this phase a small number n_1 of steps are made to estimate $D(\theta_1) = (\partial\mathcal{E}_\theta(Z)/\partial\theta)\mid_{\theta=\theta_1}$, using common random numbers; the diagonal elements of this estimate are used to define D_0.
This is described formally as follows. Denote by e_j the j'th unit vector in p dimensions. In step N, generate $Z_{N0} \sim \theta_1$ and $Z_{Nj} \sim \theta_1 + \epsilon_j e_j$, where all the $p+1$ random vectors use a common random number stream to make them strongly positively dependent and where ϵ_j are suitable constants. For different N, the random vectors are generated independently. Compute the difference quotients

$$d_{Nj} = \epsilon_j^{-1}(Z_{Nj} - Z_{N0}) \ ;$$

for small values of ϵ_j the expected value of the matrix $d_N = (d_{N1}, ..., d_{Np})$ approximates $D(\theta_1)$. However, ϵ_j must be chosen not too small because otherwise the variances of the d_{Nj} become too large.
At the end of this phase, estimate $E_{\theta_1}Z$ and $D(\theta_1)$ by

$$\bar{z} = \frac{1}{n_1}\sum_{N=1}^{n_1} Z_{N0} \text{ and } \hat{D} = \frac{1}{n_1}\sum_{N=1}^{n_1} d_N \, ,$$

respectively, make one estimated Newton-Raphson step,

$$\hat{\theta}_{n_1} = \theta_1 - \hat{D}^{-1}\,(\bar{z} - z) \ ,$$

and use the diagonal matrix $\tilde{D} = \mathrm{diag}(\hat{D})$ in phase 2.

Phase 2. This is the main phase. It consists of several subphases. The number of iteration steps per subphase is determined by a stopping rule, but bounded for subphase k by a minimum value n_{2k}^- and a maximum value n_{2k}^+. In each subphase, a_N is constant. The only difference between the subphases is the value of a_N. The subphase is ended after less than n_{2k}^+ steps as soon as the number of steps in this subphase exceeds n_{2k}^- while, for each coordinate Z_k, the sum within this subphase of successive products $(Z_{Nk} - z_k)(Z_{N-1,k} - z_k)$ is negative. If the upper bound n_{2k}^+ is reached, then the subphase is terminated anyway.
In each iteration step within each subphase, Z_N is generated according to the the current parameter value $\hat{\theta}_N$ and after each step this value is updated according to the formula

$$\hat{\theta}_{N+1} = \hat{\theta}_N \; - \; a_N \, \tilde{D}^{-1} \, (Z_N - z) \, . \tag{44}$$

At the end of each subphase, the average of $\hat{\theta}_N$ over this subphase is used as the new value for $\hat{\theta}_N$.
The value of a_N is divided by 2 when a new subphase is entered. The bounds n_{2k}^- and n_{2k}^+ are determined so that $N^{3/4} a_N$ tends to a finite positive limit.
The average of $\hat{\theta}_N$ over the last subphase is the eventual estimate $\hat{\theta}$.

Phase 3. Phase 3 is used only for the estimation of $D(\theta)$ and $\Sigma(\theta)$, using common random numbers for the estimation of the derivatives; and as a check for the (approximate) validity of (12). Therefore the value of $\hat{\theta}_N$ is left unchanged in this phase and is equal to the value obtained after last subphase of phase 2. The procedure further is as in phase 1. The covariance matrix of Z, required for the calculation of (21), is estimated in the usual way.

This algorithm contains various constants that can be adapted so as to achieve favorable convergence properties. Experience with various data sets led to the following values. The number of steps in phase 1 is $n_1 = 7+3p$. The values of ϵ_i are chosen at least 0.1, in most cases 1.0, because the variability obtained by the use of small values of ϵ_i is more serious than the bias obtained by the use of this large value. The minimum number of steps in subphase $2.k$ is $n_{2k}^- = 2^{4(k-1)/3}(7+p)$ and the maximum number is $n_{2k}^+ = n_{2k}^- + 200$. The initial value of a_N in phase 2 is 0.2. The default number of subphases is 4; more or less subphases can be used to obtain smaller or larger precision. The default number of steps in phase 3 is $n_3 = 500$. Phase 3 takes much time because each step requires $p+1$ simulations; but the variance estimate is rather unstable if the number of steps is much smaller.

REFERENCES

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975), *Discrete multivariate analysis*, Cambridge, MA: MIT Press.

Bowman, K.O., and L.R. Shenton. 1985. Method of moments. Pp. 467 – 473 in *Encyclopedia of Statistical Sciences, vol. 5*, edited by S. Kotz, N.L. Johnson, and C.B. Read. New York: Wiley.

Doreian, P., and F.N. Stokman (eds.). 1997. *Evolution of Social Networks*. Amsterdam etc.: Gordon and Breach.

Fararo, T.J., and N.P. Hummon. 1994. Discrete event simulation and theoretical models in sociology. *Advances in Group Processes, 11*: 25 – 66.

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

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

Holland, P., and S. Leinhardt. 1977b. Social structure as a network process. *Zeitschrift für Soziologie* 6: 386 – 402.

Karlin, S., and H.M. Taylor. 1975. *A first course in stochastic processes*. New York: Academic Press.

Katz, L., and C.H. Proctor. 1959. The configuration of interpersonal relations in a group as a time-dependent stochastic process. *Psychometrika* 24: 317 – 327.

Kushner, H.J. and G.G. Yin. 1997. *Stochastic Approximation: Algorithms and Applications*. New York: Springer.

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

Leenders, R.Th.A.J. 1995b. *Structure and influence. Statistical models for the dynamics of actor attributes, network structure and their interdependence.* Amsterdam: Thesis Publishers.

Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

Mayer, T.F. 1984. Parties and networks: stochastic models for relationship networks. *Journal of Mathematical Sociology* 10: 51 – 103.

Norris, J.R. 1997. *Markov Chains*. Cambridge: Cambridge University Press.

Pflug, G.Ch. 1990. Non-asymptotic confidence bounds for stochastic approximation algorithms with constant step size. *Monatshefte für Mathematik* 110: 297 – 314.

Pflug, G.Ch. 1996. *Optimization of Stochastic Models*. Boston: Kluwer.

Polyak, B.T. 1990. New method of stochastic approximation type. *Automation and Remote Control* 51: 937 – 946.

Robbins, H., and S. Monro. 1951. A stochastic approximation method. *Annals of Mathematical Statistics* 22: 400 – 407.

Ruppert, D. 1988. *Efficient estimation from a slowly convergent Robbins-Monro process.* Technical Report no. 781, School of Operations Research and Industrial Engineering, Cornell University.

Ruppert, D. 1991. Stochastic approximation. In *Handbook of Sequential Analysis* edited by Gosh, B.K., and P.K. Sen. New York: Marcel Dekker.

Snijders, T.A.B. 1996. Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology, 21*: 149 – 172. Also published in Doreian and Stokman (1997).

Snijders, T.A.B. 1999. The transition probabilities of the reciprocity model. *Journal of Mathematical Sociology* 23: 241 – 253.

Snijders, T.A.B., and M.A.J. Van Duijn. 1997. Simulation for statistical inference in dynamic network models. Pp. 493 – 512 in *Simulating Social Phenomena*, edited by Conte, R., R. Hegselmann, and P. Terna. Berlin: Springer.

Van de Bunt, G.G. 1999. *Friends by choice. An actor-oriented statistical network model for friendship networks through time.* Amsterdam: Thesis Publishers, 1999.

Van de Bunt, G.G., M.A.J. Van Duijn, and T.A.B. Snijders. 1999. Friendship networks through time: An actor-oriented statistical network model. *Computational and Mathematical Organization Theory* 5: 167 – 192.

Wasserman, S. 1977. *Stochastic models for directed graphs.* Ph.D. dissertation. University of Harvard, Dept. of Statistics.

Wasserman, S. 1979. A stochastic model for directed graphs with transition rates determined by reciprocity. Pp. 392 – 412 in *Sociological Methodology 1980*, edited by K.F. Schuessler. 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 D. Iacobucci. 1988. Sequential social network data. *Psychometrika*, 53, 261–282.

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

Yin, G. 1991. On extensions of Polyak's averaging approach to stochastic approximation. *Stochastics* 36: 245 – 264.

Zeggelink, E.P.H. 1994. Dynamics of structure: an individual oriented approach. *Social Networks* 16: 295 – 333.