Statistical Analysis of Longitudinal Network Data With Changing Composition

MARK HUISMAN
TOM A. B. SNIJDERS
University of Groningen

Markov chains can be used for the modeling of complex longitudinal network data. One class of probability models to model the evolution of social networks are stochastic actor-oriented models for network change proposed by Snijders. These models are continuous-time Markov chain models that are implemented as simulation models. The authors propose an extension of the simulation algorithm of stochastic actor-oriented models to include networks of changing composition. In empirical research, the composition of networks may change due to actors joining or leaving the network at some point in time. The composition changes are modeled as exogenous events that occur at given time points and are implemented in the simulation algorithm. The estimation of the network effects, as well as the effects of actor and dyadic attributes that influence the evolution of the network, is based on the simulation of Markov chains.

Keywords: network evolution; Markov process; stochastic actor-oriented network model; changing composition

1. INTRODUCTION

Social network analysis is concerned with the study of relationships among social actors (e.g., persons, organizations, countries, etc.). The relations refer to the collection of ties (of a specific kind) between pairs of actors from a set of actors, for instance, friendship ties between individuals or trade agreements between companies. These dyadic relations are not necessarily symmetric. Many relations are directional, in which ties are ordered from one actor to another.

A network consisting of directed relations can be represented by a directed graph (*digraph*) and an accompanying adjacency matrix (*sociomatrix*) $X = [X_{ij}]$. The element $[X_{ij}]$ is the tie variable from actor i to j. In this article, all ties within a given set of n actors are considered. The relation is supposed to be dichotomous: The elements of the adjacency matrix are either $X_{ij} = 1$, if there is a tie from i to j, or $X_{ij} = 0$ if there is not such a tie.

The elements of the adjacency matrix X hang together in a complicated dependence structure because each element X_{ij} refers to two actors i and j instead of one, and ties with and between other actors. Other actors can also have a bearing on the ties between i and j. These dependencies are called network effects, and stochastic social network models have to represent these network effects. Therefore, such models cannot be built on broad independence assumptions (as is usually done in statistical modeling), which makes the statistical analysis of social networks complicated. Some interesting network effects (i.e., dependencies) are the following (e.g., see Wasserman and Faust 1994). They are defined from the point of view of one actor, i, and will return in later sections (see also Snijders 2001).

- *Density effect*—the dependence between elements in a row of X, that is, the tendency of matrix elements X_{ij} to be 1 rather than 0.
- *Reciprocity effect*—the dependence within the dyad (X_{ij}, X_{ji}), defined by the number of reciprocated relations, indicating the preference for reciprocated relations.
- *Transitivity effect*—the dependence within triads (triplets of actors and the ties among them), defined by the number of transitive patterns, indicating the preference for transitive relations.
- *Balance*—the dependence between outgoing relations of an actor i (row X_i of X) and outgoing relations of others to whom i is related, indicating the preference for others who make the same choices as i.
- *Indirect relations effect*—the dependence between two actors i and k via one intermediary actor j, defined by the number of actors to whom i is indirectly related (at distance 2).
- *Popularity*—the dependence between an element X_{ij} and the elements in a column X_j of X, defined by the sum of the indegrees of others j to whom actor i is related, indicating the preference for popular others.
- *Activity*—the dependence between an element X_{ij} and the elements in a row X_j of X, defined by the sum of the outdegrees of others j to whom actor i is related, indicating the preference for active others.

Modeling longitudinal network data is difficult because models for the social network have to represent the dependencies of the network effects. Moreover, network effects are typically endogenous feedback effects in which, on one hand, the actions of the actors (changing their relationships with others) constitute each other's changing environment while, on the other hand, the changing network acts as a dynamic constraint on the actors' behavior (cf. Zeggelink 1994). This makes modeling network evolution even more complicated (see Doreian and Stokman 1997 and Stokman and Doreian 2001 for collections of papers about modeling longitudinal network data).

Snijders (1996, 2001) proposed a probability model for social network change in which network dynamics is treated as an endogenous dynamic process that evolves in continuous time (even though it is observed only at discrete observation moments). This model, referred to as a *stochastic actor-oriented model*, is implemented as a stochastic simulation model and is based on the idea that actors evaluate their position in the network and strive for the "best" possible configuration of relations. All actors are assumed to have full knowledge of the present state of the network, and given this state, all actors are assumed to behave independently. The evaluation of the configuration is defined as a function of the actor's position in the network and depends on parameters that can be estimated from the data by a Markov chain Monte Carlo procedure.

This study is concerned with the statistical analysis of longitudinal network data of which the composition changes over time. In a longitudinal study of the evolution of a network, some actor, or a group of actors, may join or leave the network. If analysis of the network has to be restricted to the subset of actors for whom data are available at all time points, valuable information may be lost and results may be less reliable or even biased. Moreover, statistical analysis then even could become impossible because the number of completely observed actors may be too small.

Stochastic actor-oriented models offer the opportunity to model network evolution for networks of stable composition. The models are based on simulating Markov chains of networks (adjacency matrices) between each pair of consecutive observations, which are used in a stochastic approximation procedure to estimate the model

parameters. In this article, an extension of the simulation algorithm used to generate the Markov chains is proposed to include networks of which the composition changes. These changes due to actors joining or leaving the network are treated as exogenous events that occur at fixed (known) time points. The proposed algorithm generates Markov chains of networks between two time points that represent observation times, times of composition change, or both.

Section 2 gives a brief introduction to stochastic actor-oriented network models. The basic model ingredients are discussed, as well as the model specification and estimation procedure. In Section 3, the procedure to simulate the network evolution with Markov chains is presented. The implementation of a changing set of (active) actors is described, as well as the simulation algorithm. To investigate the performance of the algorithm, we performed a small simulation study in which the simulated networks are based on a friendship network of university freshmen (Van de Bunt 1999). The design of the study and the results are presented in Section 4. Also, the algorithm is used to analyze a real-life longitudinal data set consisting of friendship relations between children in a elementary school class observed at two time points (Van Rossem et al. 2000). A discussion is given in the last section.

2. STOCHASTIC ACTOR-ORIENTED MODELS

Snijders (1996) proposed a class of dynamic network models for repeated observations of a network, represented by a directed graph. These models combine random utility modeling and continuous-time Markov processes. Holland and Leinhardt (1977a, 1977b) and Wasserman (1977) already proposed to use continuous-time Markov models for longitudinal network data, and this was further elaborated for dichotomous networks by Wasserman (1977, 1979, 1980) and Leenders (1995) in a basic Markov chain model, the reciprocity model. This model, which represents only reciprocity as a network effect, was extended by Leenders (1996) to include similarity effects. The actor-oriented model of Snijders (1996) was elaborated by Snijders and Van Duijn (1997) and Snijders (2001) and can include arbitrary network effects (see also Snijders 2003).

In the stochastic actor-oriented model, the actors try to attain a rewarding configuration of the network by individually optimizing their utility (objectives) based on the present state of the network and under the constraints of their environment. The objective functions are modeled as random utility models (cf. Maddala 1983) and are implemented as stochastic (Markov chain) simulation models. The simulations form the basis of a Markov chain Monte Carlo procedure to estimate the model parameters. This section summarizes the model and estimation procedure proposed in Snijders (2001).

2.1. MODEL SPECIFICATION

The outcome space for stochastic actor-oriented network models consists of the following basic elements.

- A continuous *time parameter* $t \in \mathcal{T}$. We consider longitudinal data, that is, a time series observed at M ordered observation times $t_1 < \cdots < t_m < \cdots < t_M$. These observation times are embedded in the set $\mathcal{T} = [t_1, t_M] = \{t \in \mathbb{R}^+ : t_1 \leq t \leq t_M\}$, and the process is modeled as evolving in continuous time, although being observed at discrete time points $t_m, m = 1, \ldots, M$.
- A *set of actors* $\mathcal{I} = \{1, \ldots, n\}$.
- An outcome space of possible *networks of relations* between the actors, $\mathcal{X}$. The networks are represented by time-dependent $n \times n$ adjacency matrices $x(t) = [x_{ij}(t)]$, where $x_{ij}(t)$ represents the existence of a tie from actors i to $j \in \mathcal{I}$ at time $t \in \mathcal{T}$. Dichotomous relations are considered, in which a tie from actor i to j is either present, $x_{ij}(t) = 1$, or absent, $x_{ij}(t) = 0$. Self-relations are not considered.
- Actor-dependent *attributes* V and pair-dependent (dyadic) attributes W. These actor attributes can be time-dependent (i.e., changeable, like attitudes): $V(t), t \in \mathcal{T}$.

The approach is methodologically individualistic; that is, the driving force behind the dynamics is constituted by the actor's actions (Snijders 1996). This means that the actors are assumed to "control" their outgoing relations, which are given in the row vectors of the adjacency matrix $x(t)$. At stochastic times, the actors are allowed to change their outgoing relations. The occurrence of these changes is governed by a *rate function*, and the actors act independently, given the current state of the network at time t, $x(t) = x$. One at a time, the actors are allowed to change one tie variable x_{ij} into the opposite $1 - x_{ij}$ by either withdrawing or initiating an outgoing tie.

The actors try to attain a positively evaluated configuration of the network. They evaluate the present network configuration by maximizing a function of the network. This function consists of three parts: an *objective function*, a *gratification function*, and a random component. The first two functions represent the two "goals" the actors are assumed to pursue: attaining a configuration of the network that is rewarding for the actors and instantaneous gratification inherent in making a specific change. The random component represents the preferences of the actors that are not explicitly modeled.

The specification of the stochastic actor-oriented network models is given by the three basic functions: the rate function, the objective function, and the gratification function. All functions depend on the current state of the network and a K-dimensional statistical parameter $\theta = (\rho, \beta, \gamma)$.

1. The *rate function*, indicating the rate at which actor i may change one of his or her outgoing relations in the time period $t_m \leq t < t_{m+1}$:

$$\lambda_i(\rho, x, m), \quad i \in \mathcal{I}, \quad x \in \mathcal{X}, \quad m = 1, \ldots, M - 1.$$

This function can either be constant or depend on actor-specific covariates or on network effects. In the remainder of this article, we restrict the exposition to constant rate functions: $\lambda_i(\rho, x, m) = \rho_m$. See Snijders (2001) for examples of nonconstant rate functions.

2. The *objective function* (or utility function), indicating the preference of actor i for a given state of the network $x(t) = x$:

$$f_i(\beta, x), \quad i \in \mathcal{I}, \quad x \in \mathcal{X}.$$

The objective function contains the substantive ingredients of the model (i.e., the network effects, the actor-dependent covariate effects, and the dyad-dependent covariate effects):

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

with the weights β_k the statistical parameters indicating the strength of the effect $s_{ik}(x)$, controlling for other effects, and L the number of included effects. The effects $s_{ik}(x)$ are relevant functions of the adjacency matrix and the attributes that are supposed to determine the evolution of the network. See Snijders (2001) or Snijders and Huisman (2003) for lists of possible effects. The effects that are included in the model have to be determined from (social network) theory and experience with modeling network dynamics.

3. The *gratification function*, indicating the instant gratification experienced by an actor i when changing the relation with an actor j, given the current state of the network $x(t) = x$:

$$g_i(\gamma, x, j), \quad i, j \in \mathcal{I}, \quad j \neq i, \quad x \in \mathcal{X}.$$

This function is motivated by the experience that a given effect may be stronger (or weaker) for initiating new relations than for breaking existing relations (Van de Bunt 1999 describes the gratification effect of initiating and breaking reciprocated relations). Like the objective function, the gratification function is a weighted sum of network and attribute effects:

$$g_i(\gamma, x, j) = \sum_{h=1}^{H} \gamma_h r_{ijh}(x), \tag{2}$$

with γ_h the statistical parameters indicating the strength of the effect $r_{ijh}(x)$ and H the number of included effects. The effects $r_{ijh}(x)$ always include a factor x_{ij} (reflecting the effect of breaking a relation) or $1 - x_{ij}$ (reflecting the effect of initiating a relation). See Snijders (2001) for examples of possible effects.

The choice made by an actor to perform an action is based on the function

$$r(\theta, i, j, x) = f_i(\beta, x(i \rightsquigarrow j)) + g_i(\gamma, x, j), \tag{3}$$

where $x(i \rightsquigarrow j)$ denotes the state of the network after actor i changed the outgoing relation to actor j: x_{ij} changed into $1 - x_{ij}$, with $i, j \in \mathcal{I}$ and $j \neq i$. The actor tries to optimize this function plus a random component; that is, actor i chooses to change his or her relation with that actor j such that

$$r(\theta, i, j, x) + U_i(t, x, j) \tag{4}$$

is maximized. The random variables $U_i(t, x, j)$ are assumed to be independent and identically distributed for all i, j, x, and t, having the Type I extreme value distribution with mean 0 and scale parameter 1. This assumption is commonly made in random utility modeling in econometrics (Maddala 1983). Given this distribution, the probability that a given actor i chooses actor j and changes the relation x_{ij} is

$$p_{ij}(\theta, x) = \frac{\exp(r(\theta, i, j, x))}{\sum_{l=1, l \neq i}^{n} \exp(r(\theta, i, l, x))}, \quad j \neq i. \tag{5}$$

This is the multinomial logit form of a random utility model (Maddala 1983:60).

2.2. ESTIMATION

In the model specification with constant rate functions ρ_m and the objective function and gratification function given by (1) and (2), respectively, the statistical parameter is $\theta = (\rho_1, \ldots, \rho_{M-1}, \beta_1, \ldots, \beta_L, \gamma_1, \ldots, \gamma_H)$, with dimension $K = M - 1 + L + H$. The data consist of repeated measurements of adjacency matrices $x(t)$, observed at time points $t = t_1, \ldots, t_M$ ($M \geq 2$), and the covariates $V(t)$ and W included in the objective and gratification functions. The parameters β_k and γ_h remain the same in the periods between the observations, only the basic change rate ρ_m is different in each period (t_m, t_{m+1}).

The K-dimensional parameter θ is estimated from the data with the method of moments (see Bowman and Shenton 1985). The method of moments is based on the intuitive idea that sample moments (statistics) are the natural estimators of population moments (expected values of the statistics). For the method of moments, a statistic Z should be chosen that captures the variability of the data accounted for by the parameters. The parameter estimates are obtained by equating the observed values (sample) and expected values (population) of Z, that is, as the solution in θ of the K-dimensional moment equation

$$E_\theta[Z] = z, \tag{6}$$

with z the observed value of the statistic Z.

A suitable definition of a vector Z_m depending on $X(t_{m-1})$ and $X(t_m)$, of which the distribution is sensitive to the parameter vector θ, is proposed in Snijders (2001). The statistic used for fitting is $Z = \sum_{m=2}^{M} Z_m$.

The system of equations (6), however, cannot be solved analytically or numerically. Therefore, a stochastic approximation procedure was proposed by Snijders (1996, 2001) based on the method of Robbins and Monro (1951; see Pflug 1996 for an introduction to the Robbins-Monro algorithm). This procedure approximates the moment estimates by simulating random adjacency matrices with the desired distributions. The solution of the moment equation (6) is obtained with

an iterative procedure in which the parameter estimates are updated by repeatedly simulating adjacency matrices.

Updating steps are generated as follows. The simulation process is carried out using the current provisional estimate of the parameter $\hat{\theta}_N$, which is used to generate a Markov chain of adjacency matrices in the time interval (t_m, t_{m+1}), given the state $x(t_m)$ of the network at observation moment t_m. For both the simulated network $X(t_{m+1})$ and the observed network $x(t_{m+1})$ at time point t_{m+1}, the statistic Z_m is computed. The difference between the simulated and observed statistics is used to update the estimate of θ until convergence, with updating formula

$$\hat{\theta}_{N+1} = \hat{\theta}_N - a_N D_0^{-1}(Z_N - z), \qquad (7)$$

where Z_N is the statistic in step N of the estimation procedure based on the simulated network and z is the observed statistic. The vector a_N slowly tends to 0, and D_0 is a matrix not depending on N (for details, see Snijders 2001; Pflug 1996).

The procedure of simulating adjacency matrices is discussed in Section 3, as well as an extension of the simulation procedure to include networks of changing composition.

An extension of the model is proposed by Snijders (2003) to include a distribution of the outdegrees. Because degrees refer directly to the individual actors, they have an important position combining structural and individual relevance. Snijders argues that the degree distribution is a primary characteristic for the structure of networks, but structural features are also important; he suggests a two-step model in which the determination of the distribution of the outdegrees is separated from the determination of the structural aspects of the network dynamics. This extension of the model offers the opportunity to assess the fit of the model by comparing the observed and fitted degree distributions. Unfortunately, other goodness-of-fit measures (e.g., "variance explained") are not (or not yet) available.

2.3. EXAMPLES

The evolution of friendship networks of university freshmen was studied by Van de Bunt (1999); Van de Bunt, Van Duijn, and Snijders (1999); and Van Duijn et al. (2003). The network data were collected

at seven time points during the academic year and are repeated measures of a friendship network of 32 university freshmen in the same discipline at a university in the Netherlands.

Van de Bunt (1999) and Van de Bunt et al. (1999) include effects that represent theoretical mechanisms that are expected to be important in friendship formation, such as the principle of diminishing returns, the tendency toward reciprocated relations, and the preference for relations with similar others. Van Duijn et al. (2003) consider a range of proximity, similarity, and network opportunity effects and use a stepwise model selection procedure with the aim of including only the important effects. They find that proximity and similarity variables (e.g., study program and gender) determine change in the network structure in the early stages of friendship development and that network opportunity (e.g., balance, popularity) is important during all stages.

Another example of stepwise model selection is given by Snijders and Baerveldt (2003), who study the effects of delinquent behavior on friendship evolution. Effects that are found to be relevant are structural effects such as reciprocity and transitivity and covariate effects of ethnicity and gender. Controlling for these effects, they found evidence for an effect of similarity in delinquent behavior on friendship evolution. Snijders (2003) gives an example of the dynamics of a network of political actors in which the degree distribution is modeled separately from structural effects (reciprocated ties, indirect relations, and gender popularity).

3. SIMULATING NETWORK EVOLUTION

The stochastic actor-oriented models are implemented as Markov chain simulation models. In this section, the generation of a continuous-time Markov chain of adjacency matrices is described, and a procedure is presented to include network evolution in which the actors are allowed to join or leave between two observation times.

3.1. MARKOV CHAINS OF ADJACENCY MATRICES

Closed social networks with n actors are represented by $n \times n$ adjacency matrices of directed and dichotomous relations. The class of all

matrices, denoted by $\mathcal{X}$, has $2^{n(n-1)}$ elements. Consider the stochastic process $\{X(t) : t \in \mathcal{T}\}$ of network evolution. This process has a finite state space $\mathcal{X}$; elements of the state space are denoted x_k and x_l.

The network evolution process is modeled as a continuous-time Markov chain. Recall that for any adjacency matrix x, the symbol $x(i \rightsquigarrow j)$ denotes the matrix obtained by changing x_{ij} into $1 - x_{ij}$ and leaving all other elements unchanged. Since the adjacency matrix $X(t)$ can change only by one element at a time, the Markov chain can be specified completely by the functions

$$q(x_k, x_l) = q_{ij}(x) = \lim_{dt \downarrow 0} \frac{\mathbb{P}[X(t + dt) = x(ij)|X(t) = x]}{dt}, \quad (8)$$

for $i \neq j$, with $x_k = x$ and $x_l = x(i \, ; \, j)$. The functions $q(x_k, x_l)$ can be interpreted as transition rates or change intensities of going from network state x_k to network state x_l, where

$$q(x_k, x_k) = -\sum_{x_l \neq x_k} q(x_k, x_l) = -\sum_i \sum_{j \neq i} q_{ij}(x). \quad (9)$$

Further define $q(x_k) = -q(x_k, x_k)$, which indicates the rate of leaving state x_k. The functions $q_{ij}(x)$ are collected in the matrix $Q = Q(x)$.

The Markov process $\{X(t) : t \in \mathcal{T}\}$ can also be defined in terms of its jump chain, the consecutive states visited by the Markov chain, and holding times, the times between the consecutive changes. Writing the Markov chain as the combination of a discrete-time jump process $\{Y_\tau : \tau \in \mathbb{Z}^+\}$ and the corresponding holding times S_τ corresponds with the idea that actors change their relations at discrete time points and that, in between these points, the network structure remains constant. The process is modeled to be right continuous, which means that for all $t \geq 0$, there exists an $\varepsilon > 0$ such that $X(t) = X(t_m)$ for $t_m \leq t < t_m + \varepsilon$.

The Markov property implies that for each $\tau \geq 1$, conditional on $Y_0, \ldots, Y_{\tau-1}$, the holding times $S_1, \ldots, S_\tau$ are independent exponential random variables with parameters $q(Y_0), \ldots, q(Y_{\tau-1})$, denoted $S_\tau \sim Exp(q(Y_{\tau-1}))$ (Norris 1997). The parameters of the exponential distributions equal

$$q(Y_{\tau-1}) = q(X(J_{\tau-1})) = \sum_i \sum_{j \neq i} q_{ij}(X(J_{\tau-1})), \quad (10)$$

where $J_\tau = \sum_{i=1}^{\tau} S_i$ is the time at which the process jumps from one state to another (i.e., the jump time). The parameters (10) are the rates of leaving state $X(J_{\tau-1})$.

Section 2 gives the rate of change for each single actor as $\lambda_i(\rho, x, m)$. The holding time S_τ of the Markov chain, however, is the minimum, over all actors, of the hypothetical holding times S_τ^i before the next change by a single actor i. A basic property of the exponential distribution implies that this S_τ is exponentially distributed with parameter

$$\lambda_+(\rho, x, m) = \sum_{i=1}^{n} \lambda_i(\rho, x, m) = \sum_i \sum_{j \neq i} q_{ij}(x). \qquad (11)$$

In the case of constant rate functions, this parameter equals $n\rho_m$.

The probability that, in a very short time interval $(t, t + dt)$, actor i changes his or her relation with actor j is equal to $q_{ij}(x)dt$ (equation (8)). This probability can be written as the total probability that an event occurs between t and $t + dt$, that the actor who may change one relation is actor i, and that this actor chooses to change his or her relation with actor j, given the current state of the network. Given that an event occurs, the probability that it is actor i who may change an outgoing relation is (see Norris 1997:72)

$$\mathbb{P}[S_\tau^i < S_\tau^j \text{ for all } j \neq i | S_\tau \leq t, X(t) = x] = \frac{\lambda_i(\rho, x, m)}{\lambda_+(\rho, x, m)}. \qquad (12)$$

Therefore, given that the probability that an event occurs is governed by S_τ, and given the probability p_{ij} that a given actor i chooses actor j to change his or her relation with.

$$\begin{aligned}
q_{ij}(x)\, dt &= \mathbb{P}[S_\tau \leq dt, S_\tau^i < S_\tau^j \text{ for all } j \neq i, \\
&\qquad X(t + dt) = x(i \rightsquigarrow j) | X(t) = x] + o(dt) \\
&= (\lambda_+(\rho, x, m)dt \\
&\qquad + o(dt)) \frac{\lambda_i(\rho, x, m)}{\lambda_+(\rho, x, m)} p_{ij}(\theta, x) + o(dt) \\
&= \lambda_i(\rho, x, m) p_{ij}(\theta, x)dt + o(dt), \qquad (13)
\end{aligned}$$

with $S_\tau = \min_i S_\tau^i$ and S_τ^i the holding time for actor i at time point τ. The symbol $o(dt)$ represents a function such that $o(dt)/dt$ tends to 0

for $dt \downarrow 0$. This implies $q_{ij}(x) = \lambda_i(\rho, x, m)\, p_{ij}(\theta, x)$. For constant rate functions ρ_m, $q_{ij}(x)$ is given by $q_{ij}(x) = \rho_m p_{ij}(\theta, x)$ for time period (t_m, t_{m+1}).

3.2. SIMULATION OF NETWORK EVOLUTION

A Markov chain of adjacency matrices for the time period (t_m, t_{m+1}) can be constructed from the initial state $X(0) = x(t_m)$, the observed configuration of the network at t_m, by generating a jump process $\{Y_\tau : \tau \in \mathbb{Z}^+\}$ and randomly drawing holding times S_τ. The simulation algorithm consists of the following steps:

  I. Begin with an initial state $x(t_m) = X(0) = Y_0$.
 II. Set $t = 0$ and complete the following steps for $\tau = 0, 1, 2, \ldots$ until $t \geq t_{m+1} - t_m$.

     1. Set $x = Y_\tau$.
     2. Conditional on $Y_\tau = x$, $S_{\tau+1} \sim Exp(\lambda_+(\rho, x, m))$. Randomly draw a holding time $S_{\tau+1}$ from this exponential distribution.
     3. Randomly draw an actor i from $\mathcal{I}$ using probabilities (12).
     4. Randomly draw an actor $j \neq i$ from $\mathcal{I}$ using probabilities $p_{ij}(\theta, x)$.
     5. Set $t = t + S_{\tau+1}$ and $Y_{\tau+1} = x(i \rightsquigarrow j)$.

Conditional on $Y_\tau = x$, $Y_{\tau+1}$ has distribution

$$\pi_{ij}(x) = \frac{\lambda_i(\rho, x, m)}{\lambda_+(\rho, x, m)}\, p_{ij}(\theta, x), \qquad (14)$$

that is, the probability that the new state equals $x(i \rightsquigarrow j)$. $S_{\tau+1}$ and $Y_{\tau+1}$ are independent, as well as independent of $Y_0, \ldots, Y_\tau$ and $S_1, \ldots, S_\tau$.

3.3. NETWORKS OF CHANGING COMPOSITION

Up to now, only closed networks were considered, with n actors from a set of actors $\mathcal{I}$ who are present at every time point $t \in \mathcal{T}$. In empirical research, however, the composition of networks may change as actors join or leave the network at some point in time. For example, the structure of a school class can change because children move to other cities, thereby leaving one network (school class) and joining another. Longitudinal data of networks of changing composition can be modeled using stochastic actor-oriented models by extending the simulation procedure. The evolutionary process is simulated

so that, at each moment, the composition (i.e., the set of actors) of the simulated network is equal to the actual composition at that moment.

Composition changes are modeled as exogenous events. This means that actors are allowed to join or leave the network at fixed time points and that neither the events nor the event times are determined by the model. To model these exogenous events, the outcome space of stochastic actor-oriented models is extended and now consists of the following basic elements.

- A continuous *time parameter* $t \in \mathcal{T} = [t_1, t_M]$. The data are a time series of networks observed at ordered observation times $t_1 < \cdots < t_m < \cdots < t_M$.
- A *total set of actors* $\mathcal{I} = \{1, \ldots, n\}$. This set is the collection of all actors that are part of the network on at least one observation time. The number of time points where actors join or leave the network is denoted by E, the number of events.
- A *time-dependent set of active actors* $\mathcal{I}_a(t) \subseteq \mathcal{I}$. This set consists of the $n_a(t)$ actors that are part of the network at some time $t \in \mathcal{T}$, that is, the time-dependent set of actors who have already joined and did not (yet) leave the network.
- A *set of times of composition change* $\mathcal{T}_e = \{t_{e,1}, \ldots, t_{e,E}\} \subset \mathcal{T}$. This is the ordered set of fixed time points at which the exogenous events occur, that is, the E time points at which actors join or leave the network: $t_1 < t_{e,1} < \cdots < t_{e,E} < t_M$. The set $\mathcal{T}_e$ can be deduced from $\mathcal{I}_a(t), t \in \mathcal{T}$.
- An outcome space of possible *networks of relations* $\mathcal{X}$. The networks are represented by $n \times n$ time-dependent adjacency matrices $x(t)$, $t \in \mathcal{T}$. The variables $x_{ij}(t)$ are defined only for $i, j \in \mathcal{I}_a(t)$. The treatment of the variables for $i, j \notin \mathcal{I}_a(t)$ is discussed later.
- Actor and pair-dependent *attributes* $V(t), t \in \mathcal{T}$, and W, respectively.

In Figure 1, an example is given of a network of which the composition changes. The total set of actors is $\mathcal{I} = \{A,B,C,D,E,F\}$, and the time-dependent sets of active actors are $\mathcal{I}_a(t_m) = \{A,B,C,D\}$ and $\mathcal{I}_a(t_{m+1}) = \{A,B,D,E\}$, where Actor C left the network in the time period (t_m, t_{m+1}) and Actor E joined the network in the time period (t_m, t_{m+1}). Actor F did leave the network before observation time t_m, will join the network after time point t_{m+1}, or both.

At each time point $t \in \mathcal{T}$, only active actors $i \in \mathcal{I}_a(t)$ are allowed to change their outgoing relations $x_{ij}(t)$ by initiating or withdrawing relations with other active actors $j \in \mathcal{I}_a(t)$. They
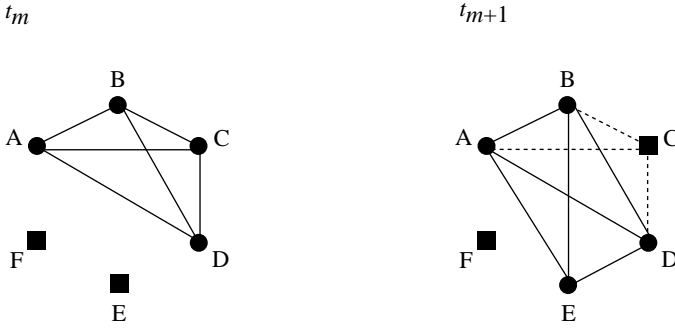
Figure 1:   Network Observed at Two Time Points t_m and t_{m+1}, Where Actor C Leaves the Network Between the Two Observation Times, Actor E Joins the Network, and Actor F Either Joins After t_{m+1} or Has Left Before t_m

cannot change relations with nonactive actors. Nonactive actors are not allowed to change their outgoing relations. Changed relations at time point t lead to a new state of the network, $x(i \rightsquigarrow j)$, with $i, j \in \mathcal{I}_a(t)$ and $j \neq i$. The utility and probability functions (4) and (5) are calculated over the set of active actors.

Exogenous events that occur in the time period (t_m, t_{m+1}) divide that period into several parts. If E' events occur in the current time period, $E' + 1$ new, consecutive periods will emerge: $(t_m, t_{e,k})$, $(t_{e,k}, t_{e,k+1})$, $\dots$, $(t_{e,k+E'}, t_{m+1})$, with $t_{e,k}$ the first time of composition change in the total period (t_m, t_{m+1}). Denote these new time periods by $(t_{m'}, t_{m'+1}) \subseteq (t_m, t_{m+1})$. The set of actors $\mathcal{I}_a(t)$ is constant for $t_{m'} \leq t < t_{m'+1}$.

In each of these time periods, a continuous-time Markov chain of adjacency matrices is constructed from the initial state $X(0) = x(t_{m'})$ by generating a jump process and holding times. Because only active actors can act, the set of active actors $\mathcal{I}_a(t_{m'})$ has to be determined at the beginning of each time period $(t_{m'}, t_{m'+1})$.

The continuous-time Markov chains of adjacency matrices in the periods $(t_{m'}, t_{m'+1})$ have a Q-matrix depending on the set $\mathcal{I}_a(t_{m'})$. Although the outcome space $\mathcal{X}$ was defined in terms of the total set of actors, $\mathcal{I}$, the Q-matrices change at each exogenous event because some states cannot be reached anymore. This means that for i or $j \notin \mathcal{I}_a(t_{m'})$, the corresponding elements $q_{ij}(x)$ are set to 0. Also, some of the nonzero elements of Q have changed because new states can be reached from the changed present state with different rates of

leaving. These changes in Q can be expressed in terms of changes in the probabilities $p_{ij}(\theta, x)$ (equation (5)). This is in agreement with the change in the group of actors j from which an actor i can choose one actor to change the relations with; a joining actor was added or a leaving actor removed. In the period $(t_{m'}, t_{m'+1}) \subseteq (t_m, t_{m+1})$, the probabilities equal

$$p_{ij}(\theta, x) = \frac{\exp(r(\theta, i, j, x))}{\sum_{l \in \mathcal{I}_a(t_{m'}), l \neq i} \exp(r(\theta, i, l, x))},$$
$$i, j, l \in \mathcal{I}_a(t_{m'}), \;\; j \neq i. \tag{15}$$

Furthermore, the parameter of the holding times S_τ of the Markov chain is different in the different time periods. The time until the next change of any actor is exponentially distributed with parameter $\lambda_+(\rho, x, m)$, defined in equation (11). In the period $(t_{m'}, t_{m'+1})$, this parameter equals

$$\lambda_+(\rho, x, m') = \sum_{i \in \mathcal{I}_a(t_{m'})} \lambda_i(\rho, x, m), \tag{16}$$

where we assume that the individual change rates of actors, $\lambda_i(\rho, x, m)$, are determined for the total period (t_m, t_{m+1}) and remain the same in the subperiods $(t_{m'}, t_{m'+1})$. In the case of constant rate functions, this parameter equals $\lambda_+(\rho, x, m') = n_a(t_{m'}) \rho_m$, and given that some actor is allowed to act, the probability that actor i may change one of his or her outgoing relations is equal to

$$\frac{\lambda_i(\rho, x, m)}{\lambda_+(\rho, x, m')} = \frac{1}{n_a(t_{m'})}. \tag{17}$$

In time period (t_m, t_{m+1}), the Markov chain starts from the initial state $X(0) = x(t_m)$ and continues until the first time of composition change, $t_{e,k}$, is reached. At this point, the network configuration changes and the Markov chain starts anew from the last simulated state of the network to which a joining actor is added or from which a leaving actor is removed. Because of the memoryless property of the exponential holding times and the Markov chains (see Norris 1997: 93), conditional on $X(t_{e,k})$, the process $\{X(t_{e,k} + t) : 0 < t < t_{e,k+1} - t_{e,k}\}$ is a Markov chain with generator matrix Q depending on $\mathcal{I}_a(t_{e,k})$ and is independent of $\{X(s) : s < t_{e,k}\}$.

3.4.1. Simulation With Composition Change

The simulation algorithm described in the previous section is used to generate Markov chains from a given value θ. Consider the ordered set of times $\{t_m, t_{e,k}, \ldots, t_{e,k+E'}, t_{m+1}\}$ with E' the number of exogenous events in the period (t_m, t_{m+1}). Denote this set by $\{t_{m'}, \ldots, t_{m'+(E'+1)}\}$, and complete the following steps for each time period $(t_{m'}, t_{m'+1}) \subseteq (t_m, t_{m+1})$:

   I. Determine the set of active actors $\mathcal{I}_a(t_{m'})$, the value of the parameter $\lambda_+(\rho, x, m')$, and the probabilities $p_{ij}(\theta, x)$.
  II. Begin with an initial state $x(t_{m'}) = X(0) = Y_0$.
 III. Set $t = 0$ and complete the following steps for $\tau = 0, 1, 2, \ldots$ until $t \geq t_{m'+1} - t_{m'}$.

   1. Set $x = Y_\tau$.
   2. Generate a holding time $S_{\tau+1} \sim Exp(\lambda_+(\rho, x, m'))$.
   3. Randomly draw an actor i from $\mathcal{I}_a(t_{m'})$ with probability (17).
   4. Randomly draw an actor $j \neq i$ from $\mathcal{I}_a(t_{m'})$ with probability $p_{ij}(\theta, x)$ (given in equation (15)).
   5. Set $t = t + S_{\tau+1}$ and $Y_{\tau+1} = x(i \rightsquigarrow j)$.

3.4.2. Relations of Joiners and Leavers

The simulation algorithm described above allows actors to join and leave the network more than once. Although this may be plausible in some specific cases, we restrict the procedure (for software programming reasons) to actors who join or leave the network only once. This means that actors can "behave" in five different ways:

   1. present in the whole time period $\mathcal{T} = [t_1, t_M]$;
   2. joining: not present in the period $[t_1, t_{e,k})$ and present in $(t_{e,k}, t_M]$;
   3. leaving: present in the period $[t_1, t_{e,k})$ and not present in $(t_{e,k}, t_M]$;
   4. joining and leaving: not present in the period $[t_1, t_{e,k})$, present in $(t_{e,k}, t_{e,l})$, and not present in $(t_{e,l}, t_M]$, $M > l > k$; and
   5. leaving and joining: present in $[t_1, t_{e,k})$, not present in $(t_{e,k}, t_{e,l})$, and present in $(t_{e,l}, t_M]$, $M > l > k$.

The networks are represented by time-dependent $n \times n$ adjacency matrices with relations $x_{ij}(t)$ between all actors in $\mathcal{I}$ (including active and nonactive actors). Not all relations between actors in $\mathcal{I}$ are

specified at the beginning and end of each time period (t_m, t_{m+1}) because the nonactive actors are not present in the network. Figure 1 shows examples of the relations that are not specified: all relations from and to Actor E (joiner) at time point t_m, all relations from and to Actor C (leaver) at time point t_{m+1}, and all relations from and to Actor F. It is assumed that these relations cannot be observed and, therefore, will not contribute to the statistic Z and the parameter estimation. This problem is dealt with by imputing values for the joining and leaving actors in the adjacency matrices.

The following is proposed. In the period before joining, the relations of joining actors are given the value 0. That is, if $i \notin \mathcal{I}_a(t)$ for $t_1 \leq t < t_{e,k}$ and $i \in \mathcal{I}_a(t)$ for $t_{e,k} \leq t \leq t_M$, then $x_{ij}(t) = x_{ji}(t) = 0$ for all $j \in \mathcal{I}$ and all $t_1 \leq t < t_{e,k}$. These imputations assume that joining actors do not have relations with members of the network before entry. However, prior information may be available on the relations between joining actors and network members. If such information is available, it can be used by giving the relations the corresponding values instead of the value 0.

In the period after leaving, the relations of leaving actors are given the value equal to their last observation. That is, if $i \in \mathcal{I}_a(t)$ for $t_1 \leq t < t_{e,k}$ and $i \notin \mathcal{I}_a(t)$, for $t_{e,k} \leq t \leq t_M$, then $x_{ij}(t) = x_{ij}(t_{obs(i)})$ and $x_{ji}(t) = x_{ji}(t_{obs(i)})$ for all $j \in \mathcal{I}$ and all $t_{e,k} \leq t \leq t_M$, with $t_{obs(i)}$ the time actor i was last observed, $t_{obs(i)} \in [t_1, t_{e,k})$.

These values are imputed to minimize the influence of leaving actors on the calculation of the statistics. By imputing the last observed values, leavers only indirectly affect the evolution of the network (i.e., the modeling of the evolutionary process) via the relations they have with other network members. These relations affect the values of the utility and probability functions (4) and (15) and, therefore, the actions of other actors. The estimation process will not be directly affected because the contribution of the leavers to the statistic Z has not changed. Imputing the value 0 for the relations after an actor left would suggest that the changes (some relations will change from 1 to 0) are due to endogenous effects influencing the evolutionary process (network effects and attribute effects). This would not be correct under the exogeneity assumption for composition changes.

If an actor first joins the network and then leaves again, the imputations described above apply. If an actor first leaves the network

and then joins again, the last observed values of that actor before leaving are imputed. That is, if $i \in \mathcal{I}_a(t)$ for $t_1 \leq t < t_{e,k}$ and for $t_{e,l} \leq t \leq t_M$, and $i \notin \mathcal{I}_a(t)$ for $t_{e,k} \leq t < t_{e,l}$, then $x_{ij}(t) = x_{ij}(t_{obs(i)})$ and $x_{ji}(t) = x_{ji}(t_{obs(i)})$ for all $j \in \mathcal{I}$ and all $t_{e,k} \leq t < t_{e,l}$, with $t_{obs(i)}$ the time actor i was last observed, $t_{obs(i)} \in [t_1, t_{e,k})$.

The exogeneity assumption of the composition changes cannot be assessed in practice. It is assumed that the researcher has substantial knowledge of the network data to assess the plausibility of the assumption. This means that the times of joining or leaving should be known, as well as knowledge on the causes of joining or leaving. The network of children in a primary school class discussed in Section 5 is an example of such a situation. Here the joining or leaving of children is due to external events (moving to another city or school, skipping classes, changes in the class composition due to school policies) that are not related to the relational variable that is studied—namely, "plays with" or "likes to play with."

4. A SMALL SIMULATION STUDY

The simulation algorithm is implemented as part of the estimation procedure for stochastic actor-oriented models for network evolution. It is available in the freeware PC program SIENA (Simulation Investigation for Empirical Network Analysis) (Snijders and Huisman 2003) to analyze repeated observations on social networks. The program runs under Windows and is included in the freeware package StOCNET for the advanced statistical analysis of social networks (Boer et al. 2003). To investigate the proposed extension of the model, we performed a small simulation study. In this section, the design and results of this study are presented.

Snijders (forthcoming) illustrates the stochastic actor-oriented models using the freshmen networks collected and studied by Van de Bunt (1999). The data were collected at seven time points during the years 1994 and 1995 and consist of 24 female and 8 male students. The students are grouped into three programs according to the length of the study program: two, three, or four years. The relation studied is defined as "at least a friendly relationship"; see Van de Bunt (1999) for a precise definition. The second (t_2) and third (t_3) observations

of this network are used in the simulation study. These consecutive observation moments are spaced three weeks apart.

4.1. DESIGN

The actors in the freshmen network who were completely observed at t_2 are used in the simulation study ($n = 29$). From the results of Snijders (forthcoming-a), it follows that a model for the evolution of this network may contain the following parameters: constant rate, density, reciprocity, indirect relations, sex-related activity, and program-related dissimilarity. This model was estimated using the networks of the $n = 29$ actors observed at t_2 and t_3. Next, the estimated parameters were treated as fixed (true) values and used to simulate the evolutionary process from t_2 up to t_3. This results in a simulated network at t_3, with known parameters of the evolutionary process. The fixed values are -0.990, 2.254, -0.524, -0.755, and -0.363 for density up to program-related dissimilarity, respectively. This means that the actors prefer reciprocated, direct relations with actors in the same program and that women are more active in creating relations than men. The observed network at t_2 and the simulated network at t_3 were used in the simulation study.

There are two independent factors in the study: percentage of joiners/leavers and modeling procedure. Three levels of the first factor were used: 10 percent (one joiner, two leavers), 21 percent (three joiners, three leavers), and 41 percent (six joiners, six leavers). Actors were randomly drawn to be a joiner or a leaver until the desired numbers were reached. The order in which the actors joined or left is determined by their position in the adjacency matrix.

As the composition changes are modeled as exogenous events, the times of change have to be specified by the researcher. Specifying different sets of times $\mathcal{T}_e$ results in different modeling procedures. The second factor in the simulations, modeling procedure, is composed of four levels: (1) the missing data procedure implemented in the SIENA software; (2) the proposed procedure for composition change in which the times of change are the same for all joiners and leavers, that is, $t_{e,k} = 0.25$ for all events k; (3) the proposed procedure for composition change in which the times of change are evenly distributed over the time period (t_2, t_3); and (4) the proposed procedure

with $t_{e,k} = 0.75$ for all k. In the first option, the joiners and leavers are considered missing and handled in a simple way to minimize their influence on the estimation results. If an element is missing in the adjacency matrix for the observation at the start and/or the end of the time period, this element is set to 0 in both observations. This means that for the calculation of the statistics used in the estimation process, the missing element is not counted. In the course of the simulations, however, the values are allowed to change from 0 to 1. This means that there may be small indirect effects of the missing values in the evolution of the network (see Snijders and Huisman 2003).

4.2. RESULTS

In each of the 3×4 cells in the design, a model was estimated containing the six parameters mentioned above. The estimation was repeated 10 times (a small number in view of the long computer time required for each run). In Table 1, the mean values of the parameter estimates and standard errors over the 10 replications are presented. Several things should be noted while interpreting the results presented in Table 1. First, the parameter estimates for the complete network (without generated joiners and leavers) are 3.546, -0.715, 2.025, -0.423, -1.154, and -0.362, respectively. When comparing these values with the estimates obtained with the missing value procedure, differences between the procedures are only due to the different number of actors. For the composition change procedures, however, differences between the estimates and the estimates obtained with the complete data reflect not only differences in the number of (active) actors but, more important, differences in the evolutionary process. The exogenous events influence the importance of the effects in the process because joiners generally start with no relations and the relation changes of the leavers are not counted.

Second, for the cells with small and moderate changes (p_1 and p_2), the differences between the parameter estimates in the 10 replications are never larger than 0.017, which is small in comparison to the standard errors. Furthermore, in all cells with large changes (p_3), the estimation algorithm did not converge, and instable estimates were found. Inspection of the estimation results showed that this was mainly caused by the parameter for sex-related activity. Removing

this parameter and reestimating the model resulted in convergence, but not all parameters in all replications were found to be significant. This means that the actual values of the parameters in the cells with large changes cannot be directly compared with those in the other cells, but the direction of the change in the parameters can.

Some general results are found in Table 1. First, the change rate increases from left to right, that is, increases when the composition changes occur later in time (at the end of the period). Treating the composition changes as missing values reduces the number of actors involved in the evolutionary process, and therefore the number of relational changes will be smaller than in the three other procedures. Also, the change rate increases when the proportion of leavers and joiners increases. The more actors join and leave the network, the more relational changes have to be made by the actors to "arrive at" the network at t_3, especially when the exogenous events occur at the end of the time period.

Second, the absolute value of the density parameter decreases when the number of joiners and leavers increases. Because the density parameter acts as a control variable when other network effects are included in the model, this decrease reflects an increase of the absolute values of other parameters. This is generally the case, except for the reciprocity and indirect relations effects when the joiner and leavers join or leave late in the period ($t_e = 0.75$). This indicates that first a certain number of relations have to exist before these effects can emerge.

Third, the standard errors increase when the number of changes increases (the only exception is the variable program, in which the values are even slightly smaller for p_2). Sometimes the increase is large, resulting in insignificant parameter values (e.g., density), especially when the proportion of joiners and leavers is large. As the failure to model the effect of sex-related activity shows, the evolution of the network is hard to model when there are many composition changes.

Comparison of the three composition change procedures shows that the differences are relatively small for covariate-related effects but can be large for network effects. Generally, the parameters for the procedure with $t_e = 0.25$ have the largest absolute values, and the parameters for $t_e = 0.75$ have the smallest absolute values, indicating that the strength of the accompanying effects decreases when actors

TABLE 1: Results of the Simulation of Network Evolution With Three Percentages of Leavers/Joiners (p_1 = 10 percent, p_2 = 21 percent, p_3 = 41 percent) and Four Methods of Handling the Composition Change: Average Parameter Values and Standard Errors (Between Parentheses) Over 10 replications

		Missing		$t_e = 0.25$		t_e Even		$t_e = 0.75$	
Rate	p_1	3.736	(0.634)	3.999	(0.681)	4.080	(0.705)	4.156	(0.705)
	p_2	4.145	(0.779)	5.121	(0.900)	5.353	(0.939)	5.513	(0.944)
	p_3	4.091	(1.163)	5.375	(1.082)	6.046	(1.221)	6.770	(1.177)
Density	p_1	−0.793	(0.312)	−0.779	(0.277)	−0.775	(0.275)	−0.754	(0.277)
	p_2	−0.711	(0.328)	−0.733	(0.298)	−0.726	(0.289)	−0.638	(0.317)
	p_3	*−0.706*	(1.543)	*−0.399*	(2.160)	*−0.386*	(1.167)	*−0.135*	(1.536)
Reciprocity	p_1	2.185	(0.387)	2.101	(0.356)	2.093	(0.356)	2.030	(0.354)
	p_2	2.004	(0.406)	2.283	(0.419)	2.216	(0.402)	2.000	(0.377)
	p_3	*2.812*	(2.608)	2.536	(0.976)	*2.345*	(1.374)	1.788	(0.663)
Indirect relations	p_1	−0.465	(0.142)	−0.506	(0.156)	−0.498	(0.153)	−0.493	(0.161)
	p_2	−0.473	(0.143)	−0.604	(0.187)	−0.587	(0.184)	−0.554	(0.198)
	p_3	*−0.539*	(0.341)	−0.784	(0.269)	−0.712	(0.297)	*−0.459*	(0.333)
Sex: activity	p_1	−1.449	(0.449)	−0.879	(0.372)	−0.869	(0.365)	−0.795	(0.359)
	p_2	−1.411	(0.468)	−0.997	(0.391)	−0.893	(0.380)	−0.791	(0.379)
	p_3	−	−	−	−	−	−	−	−
Program: dissimilarity	p_1	*−0.295*	(0.168)	−0.324	(0.158)	−0.323	(0.156)	−0.328	(0.154)
	p_2	*−0.295*	(0.190)	−0.510	(0.156)	−0.515	(0.154)	−0.539	(0.153)
	p_3	*−0.365*	(0.262)	−0.595	(0.180)	−0.654	(0.175)	−0.658	(0.169)

NOTE: For p_3, the effect of sex-related activity was excluded from the model. Italicized numbers indicate significant effects. The dashes indicate the exclusion of the corresponding effect.

join or leave at a later time point. Comparing the composition change procedures with the missing value procedure shows relatively large differences, especially for the covariate-related values.

5. EXAMPLES

In this section, two illustrations will be given of the analysis of longitudinal network data of changing composition with the SIENA program. The first example uses the freshmen data (Van de Bunt 1999) in which, artificially, joiners and leavers were created and illustrates a case in which there are few composition changes (Section 5.1). The second example is a case of real-life data from children in a primary school class, with (relatively) many changes in the composition of the network (Section 5.2).

5.1. AN EVOLVING NETWORK OF UNIVERSITY FRESHMEN

Two observations of the freshmen friendship network were used in the simulation study above. In this example, the observed networks at the second to fourth time point, t_2, t_3, and t_4 ($M = 3$), are used. See also Snijders (2001) for analyses of these data. The data are available with the SIENA program.

In the original data, the composition of the network does not change. A network of changing composition was artificially created by using actors whose outgoing relations are missing at one or more observation times due to nonresponse. These actors were treated as if they had joined or left the network. In Table 2, the number of observed and missing active actors at each observation time is presented. Also, the gender of the joining and leaving actors is given, as well as the times at which they join or leave the network. These times of composition change were randomly chosen fractions of the length of the observation periods.

In the observation period $[t_2, t_4]$, there are six joining actors and six leaving actors—three joiners and three leavers in both periods $[t_2, t_3]$ and $[t_3, t_4]$. There are nine female actors (37.5 percent of the females) and three male actors (37.5 percent of the males) who either join or leave. One actor with missing observations at t_3 was treated as regular missing data.

TABLE 2: Number of Observed and Missing Active Actors of the University Freshmen Network (n=32) at the Beginning and Ending of Each Time Period and the Created Joiners and Leavers (f = female, m = male) and Their Joining/Leaving Time

Time[a]	t_2	1	2	3	4	5	6	7	8	9	t_3	1	2	3	4	5	6	7	8	9	t_4
Observed	26										25										26
Joined		f				f			f				f		f				m		
Left			f		m	f										f			f	m	
Missing	0										1										0

a. Decimal fraction of the time between two observations.

The estimated models are the same as those described by Snijders (2001). The first model includes only network effects: the basic effects of density and reciprocity, the triadic effects of transitivity, relations at distance 2 (i.e., indirect relations), and balance. The density effect should always be included because the other effects should be controlled for density (for a more elaborate discussion of the density effect, see Snijders forthcoming). The second model includes gender effects: gender-related popularity, activity, and dissimilarity. The models are estimated in two ways: (1) the proposed procedure for composition change (models labeled Change 1 and Change 2 in Table 3) and (2) the missing data procedure implemented in SIENA (models labeled Missing 1 and Missing 2 in Table 3).

The estimated rate parameters, $\hat{\rho}_1 = 3.64$ and $\hat{\rho}_2 = 5.19$, in Change 1 show that the actors made 3.64 changes of relationships between t_2 and t_3 and 5.19 changes between t_3 and t_4. There are three significant network effects (at the 5 percent level). The significance is tested with t tests defined by the ratio of parameter estimate to standard error (e.g., for the reciprocity parameter $t = 2.24/0.28 = 8.00$) (see Snijders 1996, 2001). There is a strongly significant reciprocity effect and a significant indirect relations effect, indicating a preference for closed networks (high reciprocity, low number of indirect relations). The other two triadic effects are not significant in Change 1. When the model is reestimated without these effects, the reciprocity and indirect relations effects become larger. See Snijders (forthcoming) for a general discussion on the interpretation of the parameters.

In Change 2, three actor attribute effects were added to the significant network effects: gender-related activity, popularity, and dissimilarity. Before estimation, the means were subtracted from the values

TABLE 3:   Parameters for Models Estimated Using Observations t_2, t_3, and t_4 of the Freshmen Data

Effect	Change 1 Parameter	(SE)	Missing 1 Parameter	(SE)	Change 2 Parameter	(SE)	Missing 2 Parameter	(SE)
Rate (Period 1)	3.64		3.86		3.57		3.86	
Rate (Period 2)	5.19		3.29		5.08		3.26	
Density	−1.18	(0.31)	−1.40	(0.28)	−0.95	(0.18)	−1.08	(0.17)
Reciprocity	2.24	(0.28)	1.91	(0.30)	2.27	(0.29)	2.43	(0.35)
Transitivity	0.10	(0.072)	*0.20	(0.081)	–		–	
Balance	−0.79	(0.93)	−0.21	(0.80)	–		–	
Indirect relations	−0.41	(0.094)	−0.34	(0.079)	−0.50	(0.12)	−0.50	(0.083)
Gender popularity	–		–		0.45	(0.22)	*0.61	(0.22)
Gender activity	–		–		0.16	(0.32)	*−0.60	(0.27)
Gender dissimilarity	–		–		−0.53	(0.26)	*−0.40	(0.24)

NOTE: Parameters significant at the 5 percent level are italicized. Asterisks indicate important differences between the models Change and Missing.

of gender (1 = female, 2 = male), giving new values -0.25 and 0.75, respectively. Table 3 shows that the parameters for gender popularity and dissimilarity are significant, indicating that men receive more positive choices (positive popularity) and that men are inclined to choose men and women to choose women (negative dissimilarity).

When the joining and leaving actors are treated as missing, the parameter values of some effects change. The comparison is made for both models, and the estimated parameters are presented in Table 3 (Missing 1 and Missing 2, respectively). The transitivity effect in Missing 1 is significant, indicating a preference for transitive relations between actors ("a friend of a friend is also my friend"). Because of this larger transitivity effect, the values of the other parameters are smaller. In Missing 2, the (negative) gender activity effect is significant, making the dissimilarity effect nonsignificant. This indicates that women are more active in initiating relations then men and that there is no longer a significant liking for similar others (same sex).

The differences between the two procedures (composition change and missing) are in line with the results of the simulation study (21 percent joiners/leavers per period, evenly distributed times of change), although care should be taken when comparing the results because the models contain different effects. The estimated rate parameter for the first period, $\hat{\rho}_1$, is somewhat smaller in the change models than in the missing models, but the estimated rates for the second period, $\hat{\rho}_2$, are much larger in the change models. Allowing the actors to join or leave the network results in more changes in the second period than treating them as missing values. Also, the indirect relations effect and the effect of sex-related activity are similar to those found in the simulations.

5.2. THE DEVELOPMENT OF SOCIAL COMPETENCE IN CHILDREN

Within the framework of a longitudinal study of social development in primary school children (Utrecht Social Development Project [USDP]) (Van den Oord and Rispens 1999), the structure of the relations between children in school classes was studied to gain insight into children's peer relations and the developmental importance of these relations (Van den Oord et al. 2000; Van Rossem et al. 2000). The networks of children are the classes of a sample of elementary

TABLE 4:   Number of Observed and Missing Active Actors of the Elementary School
            Class Network (*n*=24) at the Two Observation Times and the Joiners and
            Leavers (f = female, m = male; number between brackets) and Their
            Joining/Leaving Time

Time[a]	t_2	0.25	0.50	0.75	t_3
Observed	13				13
Joined				f(2), m(3)	
Left				f(4), m(3)	
Missing	6				4

a. Fraction of the time between two observations.

schools in the province of Utrecht in the center of the Netherlands. The data were collected at three time points: t_1, during the 1996-1997 school year, 94 kindergarten classes (ages four to five); t_2, during the 1998-1999 school year, 71 first-grade classes (ages six to seven); and t_3, during the 2000-2001 school year, 65 third-grade classes (ages eight to nine).

The data used in this example come from one class of 24 children and are repeated measures of positive relations obtained at the time points t_2 and t_3. These positive relations are defined as "plays often with" or "likes to play with" (see Van den Oord et al. 2000). The two observation times ($M = 2$) are spaced two years apart.

The network of 24 children consists of 14 girls and 10 boys. The composition of the school class changed between the two observations times. In Table 4, the number of observed and missing active actors at both observation times is presented, as well as the number of joining and leaving children (female/male) and the times of change. There are 5 children who join the network (2 girls and 3 boys) in the time period [t_2, t_3], and there are 7 children who leave the class (4 girls and 3 boys). The times of composition change are the same for all leavers and joiners. It was set at 0.75, the time at which the children go from second grade to third grade (note that the data were collected during the school year, that is, t_2 is set halfway through first grade, and t_3 is halfway through third grade).

There is a considerable amount of missing data due to nonresponse in the data set: 6 out of 19 active actors at t_2 (31.6 percent) and 4 out of 17 actors at t_3 (23.5 percent) are missing. Moreover, of the 6 missing actors at t_2, 1 actor leaves the network (and is therefore not observed at t_3), and all 4 missing actors at t_3 are joiners (and therefore

not observed at t_2). This means that the actual number of actors is $n = 19$. The reason for the missing values of the 4 joiners at t_3 is a combination of the data collection method and the joining itself. Photographs of the children were used to generate the network data. These were not always available for children who joined the network at later times and, therefore, caused missing values.

Van den Oord et al. (2000) describe the networks at t_1 and Van Rossem et al. (2000) describe the networks at t_1 and t_2 in terms of structural properties (e.g., outdegree, reciprocity, transitivity) and the attributes gender and age. Here, the evolution of the network in the time period $[t_2, t_3]$ is modeled with the same structural and attribute effects. Two models were estimated—one with only the network effects density, reciprocity, and transitivity and the other with both the network and attribute effects. As in the previous example, the models are estimated in two ways: (1) modeling the network composition change and (2) treating the joiners and leavers as missing. The results are presented in Table 5 as Change 1 and Change 2, as well as Missing 1 and Missing 2, respectively.

The models Change 1 and Missing 1 are practically the same if differences in parameter estimates are evaluated in terms of the standard errors. Only the estimated rate parameter $\hat{\rho}$ shows that the second model is more restrictive: Actors make less changes because they are not included in the estimation procedure. There are some differences between the parameters of both models, but none of the parameters is significant. When attribute effects are included, the network effects change but still are not significant. The attribute effects, however, are different in the two models (only the significant attribute effects are reported in Table 5). From Change 2, which uses the composition change algorithm, it can be concluded that the children prefer relations with others of the same sex ($t = -1.72/0.46 = -3.74$) and that younger children receive more positive choices ($t = -0.18/0.066 = -2.73$). In Missing 2, these effects have the same direction but are smaller and not significant. The same was found for the attribute effects in the simulation study.

The results found in this (single) class of children show that the evolution of the network is largely determined by the gender and age of the children. Van Rossem et al. (2000) found evidence for the network effects at the separate time points, but in the evolutionary

TABLE 5:   Parameters for Models Estimated Using Observations t_1 and t_2 of the Elementary School Data

Effect	Change 1 Parameter	(SE)	Missing 1 Parameter	(SE)	Change 2 Parameter	(SE)	Missing 2 Parameter	(SE)
Rate	7.52		5.41		7.59		6.15	
Density	0.29	(1.03)	−0.10	(0.96)	1.04	(1.79)	0.82	(0.76)
Reciprocity	0.84	(0.99)	0.11	(0.84)	0.73	(0.95)	*−0.038	(0.82)
Transitivity	0.059	(0.36)	0.17	(0.26)	−0.061	(0.25)	−0.030	(0.22)
Gender dissimilarity	–		–		*−1.72*	(0.46)	*−0.68	(0.48)
Age popularity	–		–		*−0.18*	(0.066)	*−0.12	(0.10)

NOTE: Parameters significant at the 5 percent level are italicized. Asterisks indicate important differences between the models Change and Missing.

process, these effects were found to be not significant for the single class analyzed. It should be noted that the time period between the observations is considerable (two years), which may cause the measured network effects to be weak. Moreover, the amount of missingness is large. Also, the composition of the network changes considerably, and this change occurs at one time point. Together, this can make the results less stable.

6. DISCUSSION

In this article, a simulation algorithm was proposed to generate Markov chains of adjacency matrices for networks of which the composition may change due to actors joining or leaving the network. The algorithm was implemented as part of the estimation procedure for stochastic actor-oriented models for network evolution (Snijders 1996, 2001). This class of models provides a method for the statistical analysis of longitudinal social network data, in which actors are allowed to join or leave the network (or even do both).

The changes in the composition are treated as exogenous events that occur at given time points between two observation times. In the simulation procedure, Markov chains with stationary transition distributions are generated between the times of change (or between an observation time and a time of change) based on the actors who are part of the network at that time. Only these active actors are allowed to change their outgoing relations with other active actors, leading to a new state of the network. The choices of the actors are governed by their objective and gratification function, of which the parameters are estimated on the basis of the observed and simulated networks.

The actor-oriented model for longitudinal network data with changing composition is implemented in the freeware program SIENA (Snijders and Huisman 2003), which is available within the StOC-NET package (Boer et al. 2003). A small simulation study was performed using observations on friendship networks of university freshmen to investigate the performance of the procedure. Also, examples were presented of analyses of the freshmen network data and the network of children in a primary school class. The simulations and the examples show that including the joining and leaving actors in the

estimation procedure results in different parameter estimates than treating joiners and leavers as missing.

Modeling the composition change as suggested in this article means that extra information on the relations of newly joined actors and of actors who did not yet leave is included in the estimation procedure. This results in higher change rates, indicating that more relational changes were made by the actors. Also, differences in the structural and covariate effects were found between the used procedures. Comparison of the composition change procedure and the missing value procedure (simulations and examples) shows that the differences are largest for covariate effects. The direction of the differences is, however, hard to predict because it varies between data sets (and percentage joiners/leavers) and estimated models (included effects). For instance, the reciprocity effect is generally strongest for the missing data procedure in the simulations, weakest for the same procedure in the children example, and changes between the two procedures in the freshmen example.

Knowledge of the times of composition change is very important and has a large influence on the parameter estimates, especially in case of few actors and/or many events. The simulations with the three composition change procedures—that is, three different sets of times of change—showed (sometimes large) differences between the estimates. The differences are especially large for the two extreme cases in which all actors join or leave the network at the same time, early or late in the period between the observations. The estimated values of the procedure in which the times of change are evenly distributed always are between those obtained by the two other procedures. Apart from the rate, which is always larger for procedures in which the changes occur late in the period, the direction of the difference again varies with number of changes and included effects.

The composition change procedure allows for more information to be included in the estimation process in comparison to treating the joining and leaving actors as missing. Given the assumptions, this provides a better description of the evolutionary process, although extra information is needed on the times of change.

It should be noted that the nature of the stochastic approximation procedure of stochastic actor-oriented models causes the parameter estimates to be slightly different each time the model is (re)estimated.

If the convergence of the model is good, these differences will be small, but they make comparisons of the composition change models and the missing value models difficult. Large differences are certainly due to differences between the two treatments, but the cause of small differences cannot be attributed to the treatments with certainty. Also, substantive changes in the composition of networks can lead to instable solutions and problems with the convergence of the estimation algorithm. This was found to be the case in the simulations when the number of joiners and leavers was extremely large (41 percent) and in the children data, in which only a small model (containing few effects) could be estimated.

Missing data due to composition changes or nonresponse are one aspect of data quality, which plays an important role in modeling network evolution. Another important aspect is the time between the consecutive observations. In the freshmen data, the time between two observations varied between three weeks and three months. The periods between the first observations were shorter because in the beginning, more changes in the network were anticipated, and later observations were spaced further apart as the friendship network will change less (Van de Bunt 1999). It is important that there is enough time between the observations for relational changes to occur, but observations should not be spaced too far apart because too many relational changes will lead to instable solutions and problems with the convergence of the estimation algorithm. This problem was encountered in the children data, in which the observations were spaced two years apart. More research is needed to gain insight in the stability of the solutions.

Further extensions of the simulation algorithm are possible. The composition changes were modeled as exogenous events, but endogenous reasons for leaving the network may be found in the structural properties or attributes of the network or in the attributes of the actors (e.g., isolates leaving a friendship network or children leaving a school class because of bullying, modeled as a dyadic covariate). This would require extra information on the reasons for composition change, which then could be used to improve the analysis.

REFERENCES

Boer, Peter, Mark Huisman, Tom A. B. Snijders, and Evelien P. H. Zeggelink. 2003. StOCNET*: An Open Software System for the Advanced Statistical Analysis of Social Networks. Version 1.4*. Groningen: *Pro*GAMMA/ ICS. Available: http://stat.gamma.rug.nl/stocnet/

Bowman, Kamiko O. and L. R. Shenton. 1985. "Method of Moments." Pp. 467-73, in *Encyclopedia of Statistical Sciences*, vol. 5, edited by S. Kotz, N. L. Johnson, and C. B. Read. New York: John Wiley.

Doreian, Patrick and Frans N. Stokman, eds. 1997. *Evolution of Social Networks*. Amsterdam: Gordon and Breach.

Holland, Paul W. and Samuel Leinhardt. 1977a. "A Dynamic Model for Social Networks." *Journal of Mathematical Sociology* 5:5-20.

———. 1977b. "Social Structure as a Network Process." *Zeitschrift für Soziologie* 6:386-402.

Leenders, Roger Th. A. J. 1995. "Models for Network Dynamics: A Markovian Framework." *Journal of Mathematical Sociology* 20:1-21.

———. 1996. "Evolution of Friendship and Best Friendship Choices." *Journal of Mathematical Sociology* 21:133-48.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.

Norris, James R. 1997. *Markov Chains*. New York: Cambridge University Press.

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

Robbins, H. and S. Monro. 1951. "A Stochastic Approximation Method." *Annals of Mathematical Statistics* 22:400-7.

Snijders, Tom A. B. 1996. "Stochastic Actor-Oriented Models for Network Change." *Journal of Mathematical Sociology* 21:149-72.

———. 2001. "The Statistical Evaluation of Social Network Dynamics." Pp. 361-95 in *Sociological Methodology*, edited by M. E. Sobel and M. P. Becker. London: Basil Blackwell.

———. Forthcoming. "Models for Longitudinal Network Data." In *Models and Methods in Social Network Analysis*, edited by P. J. Carrington, J. Scott, and S. Wasserman. New York: Cambridge University Press.

Snijders, Tom A. B. 2003. "Accounting for Degree Distributions in Empirical Analysis of Network Dynamics." Pp. 146-161 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by R. Breiger, K. Carley, and P. Pattison. National Research Council of the National Academies, 2003. Washington, DC: The National Academies Press.

Snijders, Tom A. B. and Chris Baerveldt. 2003. "A Multilevel Network Study of the Effects of Delinquent Behavior on Friendship Evolution." *Journal of Mathematical Sociology* 27:123-151.

Snijders, Tom A. B. and Mark Huisman. 2003. *Manual for* SIENA. *Version 1.98*. Groningen: ICS, University of Groningen. Available: http://stat.gamma.rug.nl/snijders/socnet.htm

Snijders, Tom A. B. and Marijtje A. J. Van Duijn. 1997. "Simulation for Statistical Inference in Dynamic Network Models." Pp. 493-512 in *Simulating Social Phenomena*, edited by R. Conte, R. Hegselmann, and P. Terna. Berlin: Springer.

Stokman, Frans N. and Patrick Doreian, eds. 2001. "Evolution of Social Networks, Part II" [Special issue]. *Journal of Mathematical Sociology* 25.

Van de Bunt, Gerhard G. 1999. *Friends by Choice: An Actor-Oriented Statistical Network Model for Friendship Networks Through Time*. Amsterdam: Thesis.

Van de Bunt, Gerhard G., Marijtje A. J. Van Duijn, and Tom A. B. Snijders. 1999. "Friendship Networks Through Time: An Actor-Oriented Statistical Network Model." *Computational and Mathematical Organization Theory* 5:167-92.

Van den Oord, Edwin J. C. G. and Jan Rispens. 1999. "Differences Between School Classes in Preschoolers' Psychosocial Adjustment: Evidence for the Importance of Children's Interpersonal Relations." *Journal of Child Psychology and Psychiatry* 40:417-30.

Van den Oord, Edwin J. C. G., Jan Rispens, Paul P. Goudena, and Marjolijn M. Vermande. 2000. "Some Developmental Implications of Structural Aspects of Preschoolers' Relations With Classmates." *Journal of Applied Developmental Psychology* 21:619-39.

Van Duijn, Marijtje A. J., Evelien P. H. Zeggelink, Mark Huisman, Frans N. Stokman, and Frans W. Wasseur. 2003. "Evolution of Sociology Freshmen into a Friendship Network." *Journal of Mathematical Sociology* 27:153-191.

Van Rossem, Ronan, Edwin J. C. G. Van den Oord, Marleen H. Gerrits, and Jan Rispens. 2000. "The Development of Social Competence in Children: Structural Changes in the Social Relationships of Children From Kindergarten to First Grade." Unpublished manuscript.

Wasserman, Stanley. 1977. "Stochastic Models for Directed Graphs." Doctoral dissertation, University of Harvard, Department of Statistics.

———. 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.

———. 1980. "Analyzing Social Networks as Stochastic Processes." *Journal of the American Statistical Association* 75:280-94.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Zeggelink, Evelien P. H. 1994. "Dynamics of Structure: An Individual Oriented Approach." *Social Networks* 16:295-333.

Mark Huisman is an assistant professor in the Department of Statistics and Data Analysis at the University of Groningen, the Netherlands. His current research interests are statistical models for social networks and network evolution and missing data methods. One of his recent publications is "Imputation of Missing Scale Data With Item Response Models" with Ivo Molenaar (2001) in Essays on Item Response Theory.

Tom A. B. Snijders is a professor of statistics and methodology in the Department of Statistics at the University of Groningen, the Netherlands. He is scientific director of the Interuniversity Center for Social Science Theory and Methodology (ICS). He is interested especially in the interplay between methodology and substantive social science. His main current research interests include social network analysis and multilevel analysis. One of his recent publications is "The Statistical Evaluation of Social Network Dynamics" in Sociological Methodology (2001).