Models for Longitudinal Network Data

Tom A.B. Snijders
ICS, Department of Sociology
University of Groningen

Abstract

This chapter treats statistical methods for network evolution. It is argued that it is most fruitful to consider models where network evolution is represented as the result of many (usually non-observed) small changes occurring between the consecutively observed networks. Accordingly, the focus is on models where a continuous-time network evolution is assumed although the observations are made at discrete time points (two or more).

Three models are considered in detail, all based on the assumption that the observed networks are outcomes of a Markov process evolving in continuous time. The independent arcs model is a trivial baseline model. The reciprocity model expresses effects of reciprocity, but lacks other structural effects. The actor-oriented model is based on a model of actors changing their outgoing ties as a consequence of myopic stochastic optimization of an objective function. This framework offers the flexibility to represent a variety of network effects. An estimation algorithm is treated, based on a Markov chain Monte Carlo implementation of the method of moments.

Preprint of:
Snijders, Tom A.B. (2005). Models for Longitudinal Network Data.
Chapter 11 (pp. 215–247) in
P. Carrington, J. Scott, and S. Wasserman (Eds.),
Models and methods in social network analysis.
New York: Cambridge University Press.

1 Some basic ideas about longitudinal social network data

The statistical modeling of social networks is difficult because of the complicated dependence structures of the processes underlying their genesis and development. One might think that the statistical modeling of longitudinal data on social networks is more difficult than modeling single observations of social networks. It is plausible, however, that in many cases, the rules defining the dynamics of network evolution are simpler than the rules required to describe a single network, because a network usually is the result of a complex and untraceable history. This chapter on the statistical modeling of network dynamics focuses on models assuming that the network is observed at a number of discrete time points, but there is an unobserved network evolution going on between these time points. The first observation of the network is not modeled but regarded as given, so that the history leading to this network is disregarded in the model construction. This will give, hopefully, a better insight in the rules of network evolution than also modeling the very first network observation. Further, it is not assumed that the network process is in a steady state. Equilibrium assumptions mostly are unwarranted for observations on network processes, and making such assumptions could lead to biased conclusions.

The treatment of methods for analyzing longitudinal network data presupposes that such data are available. It is evident that the collection of such data requires even more effort than the collection of network data on a single moment because, in most types of network data collection, the researcher will have to retain the collaboration of the network members.

As data, we suppose that we have M repeated observations on a network with the same set of g actors. The observed networks are represented as digraphs with adjacency matrices $\mathbf{X}(t_m) = \big(X_{ij}(t_m)\big)$ for $m = 1, \ldots, M$, where i and j range from 1 to g. The variable $X_{ij}(t)$ indicates whether at time t there is a tie from i to j (value 1) or not (value 0). The diagonal of the adjacency matrix is defined to be 0, $X_{ii}(t) = 0$ for all i. The number M of repeated observations must be at least 2.

Various models have been proposed for the statistical analysis of longitudinal social network data. Earlier reviews were given by Wasserman (1978), Frank (1991), and Snijders (1995). This chapter does not give a general review of this literature but focuses on models based on the assumption of continuous-time network evolution. The motivation for this choice is the following.

When thinking of how to construct a statistical model for the network dynamics that lead to the change from $\mathbf{X}(t_1)$ to $\mathbf{X}(t_2)$, then on to $\mathbf{X}(t_3)$, etc., a first question is whether these changes are represented by one 'jump', or as the result of a series of small changes. It is a natural idea to conceive of network dynamics as not being bound in a special way to the observation moments, but as a more or less continuous process which feeds back upon itself because at each moment, the current network structure is an important determinant of the likelihood of the changes that might occur next. The idea of regarding the dynamics of social phenomena as being the result of a continuous-time process, even though observations are made at discrete time points, was proposed already by Coleman (1964). Several methods have been proposed for analyzing repeated observations on social networks using models where changes are made in discrete steps from one observation moment to the next (Katz and Proctor, 1959, Wasserman, 1987, Wasserman and Iacobucci, 1988, Sanil, Banks, and Carley, 1994, Banks and Carley, 1996, and Robins and Pattison, 2001). This chapter does not treat these models, but focuses on models which assume that the network $\mathbf{X}(t)$ is evolving in continuous time, although being observed only at the discrete moments $t_m,\ m = 1, \ldots, M$.

In this class of models, the ones most directly amenable to statistical analysis are those postulating that the network $\mathbf{X}(t)$ is a continuous-time Markov chain. For categorical non-network data, such models were proposed by Coleman (1964) and the statistical treatment was elaborated by Kalbfleisch and Lawless (1985). Modeling the evolution of network data using continuous-time Markov chains was proposed by Holland and Leinhardt (1977a, 1977b) and Wasserman (1977). The first authors proposed the principle but did not work it out in practical detail. Wasserman (1977, 1979, 1980), followed by Leenders (1995a), elaborated the so-called reciprocity model, which is a continuous-time Markov model which represents only reciprocity as a network effect. Leenders (1995a, 1996) also included similarity effects (as a function of covariates) in this model. Snijders and Van Duijn (1997) and Snijders (2001) elaborated the so-called stochastic actor-oriented model which is a model for network dynamics that can include arbitrary network effects. This chapter treats some earlier models such as the reciprocity model, and focuses on the actor-oriented model.

2 Descriptive statistics

Any empirical analysis of longitudinal network data should start by making a basic data description in the form of making graphs of the networks or plotting some basic network statistics over time. These can include the density or average degree, degree variance, number of isolates, number of components of given sizes, parameters for reciprocity, transitivity, segmentation, etc.

Next to sequences of statistics for the M observed networks, it is instructive to give a description of the number and types of changes that occurred. This can be done in increasing stages of structural complexity. The simplest stage is given by the change counts indicating how many tie variables changed from h to k from observation moment t_m to t_{m+1},

$$N_{hk}(m) = \sharp\{(i,j) \mid X_{ij}(t_m) = h,\, X_{ij}(t_{m+1}) = k\,\} \tag{1}$$

for $h, k = 0, 1$, where $\sharp A$ denotes the number of elements of the set A; and the corresponding change rates

$$r_h(m) \;=\; \frac{N_{h1}(m)}{N_{h0}(m) \;+\; N_{h1}(m)}\;. \tag{2}$$

This idea can also be applied at the dyadic level; see Wasserman (1980), Table 5. The added complication here is that there are two ways in which a dyad can be asymmetric at two consecutive observation moments: it can have remained the same, or the two tie variables can have interchanged their values. Triadic extensions are also possible.

3 Example

As an example, the network of 32 freshmen students is used that was studied by van de Bunt (1999) and also by van de Bunt, van Duijn, and Snijders (1999). These references give more detailed background information on this data set. It was collected in 1994/95. The network consists of 32 freshmen students in the same discipline at a university in The Netherlands, who answered a questionnaire with sociometric (and other) questions at seven times points during the academic year, coded t_0 to t_6. Times t_0 to t_4 are spaced three weeks apart, t_4 to t_6 six weeks. This data set is distributed with the SIENA program (Snijders and Huisman, 2003). The set of all students majoring in this discipline started with 56 persons. A number of them stopped with the university studies during the freshmen year, and were deleted from this data set. Of the remaining persons, there were 32 who responded to most of the questionnaires; they form the network analysed here. The relation studied here is defined as a 'friendly relation'; the precise definition can be found in van de Bunt (1999).

Figures of the changing network are not presented, because these are not very illuminating due to the large numbers of arcs. Table 1 presents some descriptive statistics. Each statistic is calculated on the basis of all available data required for calculating this statistic.

Time	t_0	t_1	t_2	t_3	t_4	t_5	t_6
Average degree	0.19	3.78	4.63	5.60	6.95	7.73	6.96
Mutuality index	0.67	0.66	0.67	0.64	0.66	0.74	0.71
Transitivity index	–	0.44	0.51	0.44	0.45	0.56	0.46
Fraction missing	0.00	0.06	0.09	0.16	0.19	0.04	0.22

Table 1: Basic descriptives.

The average degree, starting at virtually nil, rises rapidly to a value about 7. The mutuality index (defined as the fraction of ties reciprocated) is remarkably constant at almost 0.7. The transitivity index (defined as the number of transitive triplets, divided by the number of potentially transitive triplets) also is rather constant at almost 0.5.

The change counts (1) are indicated in Table 2. The total number of changes between consecutive observation moments is 104 in the first period, and 51 to 80 in all further periods.

	m					
h, k	0	1	2	3	4	5
0 to 0	820	716	590	530	546	546
0 to 1	104	43	47	31	50	35
1 to 0	0	22	13	20	30	30
1 to 1	6	87	94	98	140	130

Table 2: Change frequencies $N_{hk}(m)$ for the periods $t_m - t_{m+1}$. (Only for arc variables available at t_m and t_{m+1}.)

4 Continuous-time Markov chains

This section introduces the basics of continous-time Markov chains. These stochastic processes are treated extensively in textbooks such as Taylor and Karlin (1998) and Norris (1997). Introductions aiming specifically at social networks are given by Leenders (1995b) and Wasserman (1979, 1980).

The present section is phrased in terms of an arbitrary finite outcome space $\mathcal{X}$, which in the case of network dynamics is the set of all directed graphs – equivalently, all adjacency matrices. The observation times t_1 to t_M are embedded in an interval of time points $\mathcal{T} = [t_1, t_M] = \{t \in \mathbb{R} \mid t_1 \leq t \leq t_M\}$. It is assumed that changes can take place unobserved between the observation moments.

Consider a stochastic process $\{X(t) \mid t \in \mathcal{T}\}$ with a finite outcome space $\mathcal{X}$, where 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,

$\{X(t) \,|\, t > t_a\}$ given the present and the past, $\{X(t) \,|\, t \le t_a\}$, is a function only of the present, $X(t_a)$. This implies that for any possible outcome $\tilde{x} \in \mathcal{X}$, and for any pair of time points $t_a < t_b$,

$$\begin{aligned} &\mathrm{P}\{X(t_b) = \tilde{x} \mid X(t) = x(t) \text{ for all } t \le t_a\} \\ &\qquad = \mathrm{P}\{X(t_b) = \tilde{x} \mid X(t_a) = x(t_a)\} \;. \end{aligned} \tag{3}$$

The Markov chain is said to have a stationary transition distribution if the probability (3) depends on the time points t_a and t_b only as a function of the time elapsed in between, $t_b - t_a$. It can be proven that if $\{X(t) \,|\, t \in \mathcal{T}\}$ is a continuous-time Markov chain with stationary transition distribution, then there exists a function $q : \mathcal{X}^2 \to \mathbb{R}$ such that

$$\begin{aligned} q(x, \tilde{x}) &= \lim_{dt \downarrow 0} \frac{\mathrm{P}\{X(t + dt) = \tilde{x} \mid X(t) = x\}}{dt} \quad \text{for } \tilde{x} \ne x \\ q(x, x) &= \lim_{dt \downarrow 0} \frac{1 - \mathrm{P}\{X(t + dt) = x \mid X(t) = x\}}{dt} \;. \end{aligned} \tag{4}$$

This function q is called the intensity matrix or the infinitesimal generator. The interpretation is that for any given value x, if $X(t) = x$ at some moment t, then the probability that the process changes to the new value $\tilde{x}$ in the short time interval from t to $t + dt$ is approximately $q(x, \tilde{x})\, dt$. The element $q(x, \tilde{x})$ is referred to as the *rate* at which x tends to change into $\tilde{x}$ (for $x \ne \tilde{x}$). 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 $r\, dt$. Note that the diagonal elements $q(x, x)$ are negative and are defined such that the rows sums of the matrix Q are 0.

Some more understanding of what the intensity matrix means for the distribution of $X(t)$ can be obtained by considering how the distribution could be simulated. A process $X(t)$ for $t \ge t_0$ with this distribution can be simulated as follows, given the current value $X(t_0) = x$:

1. Generate a random variable D with the exponential distribution with parameter $-q(x, x)$ (it may be noted that the expected value of this distribution is $-1/q(x, x)$).
2. Choose a random value $Y \in \mathcal{X}$, with probabilities

$$\mathrm{P}\{Y = \tilde{x}\} = \frac{q(x, \tilde{x})}{-q(x, x)} \quad \text{for } \tilde{x} \ne x \,;\; \mathrm{P}\{Y = x\} = 0 \;.$$

3. Define $X(t) = x$ for $t_0 < t < t_0 + D$ and $X(t_0 + D) = Y$.
4. Set $t_0 := t_0 + D$ and $x := Y$ and continue with step 1.

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

$$P(t_b - t_a) \;=\; \Big(\, \mathrm{P}\{X(t_b) = \tilde{x} \mid X(t_a) = x\} \,\Big)_{x, \tilde{x} \in \mathcal{X}} \tag{5}$$

must satisfy

$$\frac{d}{dt} P(t) \;=\; Q\, P(t)\;. \tag{6}$$

The solution to this system of differential equations is given by

$$P(t) \;=\; e^{t\,Q}\;, \tag{7}$$

where Q is the matrix with elements $q(x, \tilde{x})$ and the matrix exponential is defined by

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

If the Markov chain has a stationary transition distribution, and starting from each state x it is possible (with a positive probability) to reach each other state $\tilde{x}$, then the random process $X(t)$ has a unique limiting distribution. Representing this distribution by the probability vector π with elements $\pi_x = \mathrm{P}\{X = x\}$, this means that

$$\lim_{t \to \infty} \mathrm{P}\{X(t) = \tilde{x} \mid X(0) = x\} \;=\; \pi_{\tilde{x}} \text{ for all } \tilde{x},\, x \in \mathcal{X}\;.$$

This is also the stationary distribution in the sense that

$$\pi'\, P(t) \;=\; \pi' \quad \text{for all } t,$$

i.e., if the initial probability distribution is π, then this is the distribution of $X(t)$ for all t. It can be shown that the stationary distribution also satisfies

$$\pi'\, Q \;=\; 0\;.$$

It can be hard to find this limiting distribution for a given intensity matrix. Sometimes, it can be found by checking a convenient sufficient condition for stationarity, the so-called *detailed balance condition*. The probability vector π and the intensity matrix Q are said to be in detailed balance if

$$\pi_x \, q(x, \tilde{x}) \;=\; \pi_{\tilde{x}} \, q(\tilde{x}, x) \;\text{ for all } \tilde{x} \neq x \; . \tag{8}$$

This can be understood as follows: assume a mass distribution over the vertex set $\mathcal{X}$ and a flow of this mass between the vertices; if there is a mass $\pi(x)$ at vertex x, then the rate of flow is $\pi(x) \, q(x, \tilde{x})$ from x to any $\tilde{x} \neq x$. Then (8) indicates that as much mass flows directly from $\tilde{x}$ to x as directly from x to $\tilde{x}$, so the flow keeps the mass distribution unchanged. If each state $\tilde{x}$ is (directly or indirectly) reachable from each other state x and the detailed balance equation holds, then indeed π is the unique stationary distribution.

In the present chapter, this theory is applied to stochastic processes where $\mathcal{X}$ is the set of all digraphs, or adjacency matrices, with elements denoted by $\mathbf{x}$. The models discussed here have the property that at most one tie changes at any time point (a model where several ties can change simultaneously is the party model of Mayer, 1984). All transition rates $q(\mathbf{x}, \tilde{\mathbf{x}})$ for adjacency matrices $\mathbf{x}$ and $\tilde{\mathbf{x}}$ differing in two or more elements, then are 0. A more convenient notation for such models is obtained by working with the rate at which $X_{ij}(t)$ changes to its opposite (0 to 1, or 1 to 0), defined by

$$q_{ij}(\mathbf{x}) \;=\; q(\mathbf{x}, \tilde{\mathbf{x}}) \tag{9}$$

where

$$\tilde{x}_{hk} \;=\; \begin{cases} x_{hk} & \text{if } (h,k) \neq (i,j) \\ 1 - x_{ij} & \text{if } (h,k) = (i,j) \end{cases} \; .$$

The value $q_{ij}(\mathbf{x})$ can be interpreted as the propensity for the arc variable X_{ij} to change into its opposite $(1 - X_{ij})$, given that the current state of the network is $\mathbf{X} = \mathbf{x}$.

5 A simple model: independent arcs

The simplest network model of this kind is the total independence model, in which all arc variables $X_{ij}(t)$ follow independent Markov processes. This may be an uninteresting model for practical purposes, but it sometimes provides a useful baseline because it allows explicit calculations. It also is a nice and simple illustration of the theory of the preceding section. For each arc variable separately, the model applies with $\mathcal{X} = \{0, 1\}$ and the rates at which the two states change into each other are denoted λ_0 and λ_1 .

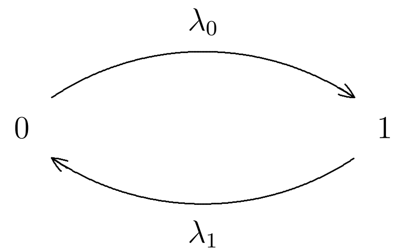

Figure 1: Transition rates in the independent arcs model.

The value $X_{ij} = 0$ changes into 1 at a rate λ_0, while the value 1 changes into 0 at a rate λ_1. The intensity matrix for the tie variables is equal to

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix} .$$

This means that the intensity matrix (9) for the entire adjacency matrix is given by

$$q_{ij}(\mathbf{x}) = \lambda_{x_{ij}} . \tag{10}$$

The transition probabilities can be derived from (6) as follows. (These results also are given in Taylor and Karlin (1998), p. 362-364.)

Denote $\xi_h(t) = \mathrm{P}\{X_{ij}(t) = 1 \mid X_{ij}(0) = h\}$ for $h = 0, 1$. The transition matrix (5) then is equal to

$$P(t) = \begin{pmatrix} 1 - \xi_0(t) & \xi_0(t) \\ 1 - \xi_1(t) & \xi_1(t) \end{pmatrix} .$$

This implies that (6) can be written as

$$\xi'_h(t) \;=\; \lambda_0 \;-\; (\lambda_0 + \lambda_1)\,\xi_h(t) \qquad\quad (h = 0, 1)\ .$$

This differential equation has the solution

$$\xi_h(t) \;=\; \frac{1}{\lambda_0 + \lambda_1}\,\{\lambda_0 \;-\; \exp(-\lambda_0 + \lambda_1(t + c))\}\ \ ,$$

where c depends on the initial condition $X_{ij}(0)$. With the initial conditions $\xi_h(0) = h$ we obtain the solutions

$$\begin{aligned} \xi_0(t) &= \frac{\lambda_0}{\lambda_+}\,\{1 - \exp(-\lambda_+ t)\}\ \ , \\ \xi_1(t) &= \frac{1}{\lambda_+}\,\{\lambda_0 + \lambda_1 \exp(-\lambda_+ t)\}\ \ , \end{aligned}$$

where $\lambda_+ = \lambda_0 + \lambda_1$. Note that this implies $0 < \xi_0(t) < \lambda_0/\lambda_+ < \xi_1(t) < 1$. These equations imply that, for all t,

$$\frac{\mathrm{P}\{X_{ij}(t) = 1 \mid X_{ij}(0) = 0\}}{\mathrm{P}\{X_{ij}(t) = 0 \mid X_{ij}(0) = 1\}} \;=\; \frac{\xi_0(t)}{1 - \xi_1(t)} \;=\; \frac{\lambda_0}{\lambda_1}\ . \tag{11}$$

For $t \to \infty$, the probability that $X_{ij}(t) = 1$ approaches the limit λ_0/λ_+ irrespective of the initial condition. The stationary probability vector $\pi = (\lambda_1/\lambda_+,\ \lambda_0/\lambda_+)$ satisfies the detailed balance equations (8), given here by

$$\pi_0\,\lambda_0 \;=\; \pi_1\,\lambda_1\ .$$

Maximum likelihood estimators for the parameters in this model are discussed by Snijders and van Duijn (1997).

6 The reciprocity model

The reciprocity model (Wasserman, 1977, 1979, 1980) is a continuous-time Markov chain model for directed graphs where all dyads $(X_{ij}(t),\ X_{ji}(t))$ are independent and have the same transition distribution, but the arc variables within the dyad are dependent. This model can be regarded as a Markov chain for the dyads, with outcome space $\mathcal{X} = \{00, 01, 10, 11\}$. The transition rates can be expressed by

$$q_{ij}(\mathbf{x}) \;=\; \lambda_h \;+\; \mu_h\, x_{ji} \;\;\text{ for } h = x_{ij}\ . \tag{12}$$

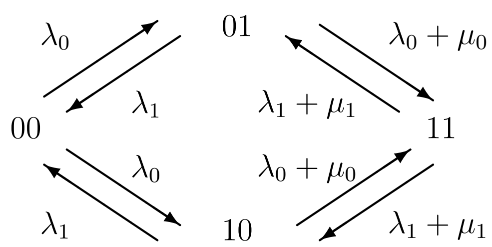

Figure 2: Transition rates between dyads.

These transition rates are summarized in Figure 2.

The stationary distribution for the dyads can be derived by solving the detailed balance equations. It is given by

$$\begin{aligned}
\pi_{00} &= \frac{\lambda_1(\lambda_1 + \mu_1)}{\lambda_0(\lambda_0 + \mu_0) + (\lambda_1 + \mu_1)(2\lambda_0 + \lambda_1)} \,, \\
\pi_{11} &= \frac{\lambda_0(\lambda_0 + \mu_0)}{\lambda_0(\lambda_0 + \mu_0) + (\lambda_1 + \mu_1)(2\lambda_0 + \lambda_1)} \,, \qquad (13) \\
\pi_{01} &= \pi_{10} = \tfrac{1}{2}(1 - \pi_{00} - \pi_{11}) \qquad (14) \\
&= \frac{\lambda_0(\lambda_1 + \mu_1)}{\lambda_0(\lambda_0 + \mu_0) + (\lambda_1 + \mu_1)(2\lambda_0 + \lambda_1)} \,,
\end{aligned}$$

cf. Wasserman (1979), Snijders (1999).

The transition matrix $P(t)$ has a rather complicated form. It was derived by Wasserman (1977) (whose result contains a minor error) and Leenders (1995a) from (7) by an eigenvalue decomposition of Q, and by Snijders (1999) by solving the differential equation system (6). The reader is referred to the latter two publications for the precise expressions.

This model can be extended by making the change rates (12) dependent on covariates. This was done by Leenders (1995a, 1996) who combined the effects of reciprocity and covariate-dependent similarity. However, such extensions are limited by the fact that the reciprocity model postulates that all dyads are independent, which is a severe restriction that runs counter to many basic ideas of social network analysis.

7 The popularity model

A model in which transition rates depend on in-degrees was proposed by Wasserman (1977, 1980). He called this the *popularity model* because it expresses that the popularity of actors, as measured by their in-degrees, is determined endogenously by the network evolution. The transition rates of this popularity model are given by

$$q_{ij}(\mathbf{x}) \;=\; \lambda_h \;+\; \pi_h \, x_{+j} \;\; \text{for } h = x_{ij} \; . \tag{15}$$

A mathematical equivalent model is the expansiveness model, in which the transition rates depend on the out-degrees, see Wasserman (1997). Under the popularity model, the columns of the adjacency matrix follow independent stochastic processes. The in-degrees $X_{+j}(t)$ themselves follow so-called birth-and-death processes, which property was exploited by Wasserman (1980) to derive the stationary distribution.

8 Actor-oriented models

In models for network dynamics that represent the effects of current network structure on the ongoing changes in the network, it must be allowed that the probabilities of relational changes depend on the entire network structure. This generalizes the models presented in the preceding two sections, where only one effect (reciprocity and popularity, respectively) is considered, isolated from other effects. This more encompassing approach may be regarded as a kind of macro-to-micro modeling, where the entire network is the macro level and the single tie, or the collection of ties of a single actor, is the micro level. The model will be a stochastic process on the set of all digraphs, which from now on shall be the set denoted by $\mathcal{X}$.

An actor-oriented approach to this type of modeling was proposed by Snijders (1995, 1996), Snijders and van Duijn (1997), and Snijders (2001). The elements of the actor-oriented approach are listed in Snijders (1996, Section 2). Some applications were presented by Van de Bunt et al. (1999), de Nooy (2002), and van Duijn, Zeggelink, Huisman, Stokman, and Wasseur (2003). This actor-orientation means that, for each change in the network, the perspective is taken of the actor 'whose tie' is changing. It is assumed that actor i controls the set of outgoing tie variables $(X_{i1}, \ldots, X_{ig})$, collected in the i'th row of the adjacency matrix. The network changes only by one tie at a time. Such a change is called a *ministep*. The moment when actor i changes one of his ties, and the particular change that he makes, can depend on the network structure and on attributes represented by observed covariates. The 'moment when' is stochastically determined in the model by the *rate function*, 'the particular change to make' by the *objective function* and the *gratification function*. First we discuss the roles of these three ingredients of the model, later we discuss how they can be specified.

8.1 Rate function

The rate function indicates how frequently the actors make ministeps:

> The ***Rate Function*** $\lambda_i(\mathbf{x})$ for actor i is the rate
> at which there occur changes in this actor's outgoing ties.

The rate function can be formally defined by

$$\lambda_i(\mathbf{x}) = \lim_{dt \downarrow 0} \frac{1}{dt} \mathrm{P}\big\{X_{ij}(t+dt) \neq X_{ij}(t) \text{ for some } j \in \{1, \ldots, g\} \mid \mathbf{X}(t) = \mathbf{x}\big\} . \tag{16}$$

The simplest specification of the rate of change of the network is that all actors have the same rate of change ρ of their ties. This means that for each actor, the probability that this actor makes a ministep in the short time interval $(t,\, t+dt)$ is approximately $\rho\, dt$, and in a short time interval there is independence between the actors in whether they take a ministep. Then $\lambda_i(\mathbf{x}) = \rho$ for all i. The waiting times D between successive ministeps of each given actor then have the exponential distribution with probability density function $\rho e^{-\rho d}$ for $d > 0$, and the expected total number of ministeps made by all actors between time points t_a and t_b is $g\rho(t_b - t_a)$: as is intuitively clear, this expected number is proportional to the total number of actors g, proportional to the rate of change ρ, and proportional to the time length $t_b - t_a$.

Sometimes it can be theoretically or empirically plausible to let these change rates differ between actors as a function of covariates, or to let them depend dynamically on network structure. This is elaborated in Section 9.2.

8.2 Objective function

The basic idea of the actor-oriented model is that, when actor i has the occasion to make a change in his outgoing tie variables $(X_{i1}, \ldots, X_{ig})$, this actor selects the change which gives the greatest increase in the so-called objective function plus a random term.

The ***Objective Function*** $f_i(\mathbf{x})$ of actor i is the value attached by this actor to the network configuration $\mathbf{x}$.

Thus, the objective function represents the preference distribution of the actor over the set $\mathcal{X}$ of all possible networks. It will be assumed that if there are differences between actors in their objective functions, these can be identified on the basis of covariates; in other words, the objective function does not contain unknown actor-specific parameters, but it can contain known actor-specific covariates.

When actor i makes a change in $(X_{i1}, \ldots, X_{ig})$ (i.e., makes a ministep), he changes how he is tied to exactly one of the $g-1$ other actors. From one of the $X_{i+} = \sum_j X_{ij}$ other actors to whom i is tied, he could withdraw the tie; or to one of the $g-1-X_{i+}$ others to whom he is not tied, he could extend a tie. Given that the present network is denoted by $\mathbf{x} = \mathbf{X}(t)$, the new network that would result by changing the single tie variable x_{ij} into its opposite $1-x_{ij}$ is denoted $\mathbf{x}(i \rightsquigarrow j)$ (to be interpreted as "the digraph obtained from $\mathbf{x}$ when i changes the tie variable to j"). The choice is modeled as follows. Denote by $U(j)$ a random variable which indicates the unexplained, or residual, part of the attraction for i to j. These U_j are assumed to be random variables distributed symmetrically about 0 and independently generated for each new ministep (this is left implicit in the notation). The actor chooses to change his tie variable with that other actor j ($j \neq i$) for whom the value of

$$f_i\big(\mathbf{x}(i \rightsquigarrow j)\big) \; + \; U(j)$$

is highest. This can be regarded as a myopic stochastic optimization rule: myopic because only the situation obtained immediately after the ministep is considered, stochastic because the unexplained part is modeled by means of a random variable.

A convenient and traditional choice for the distribution of $U(j)$ is the type 1 extreme value distribution or Gumbel distribution with mean 0 and scale parameter 1 (Maddala, 1983). Under this assumption, the probability that i chooses to change x_{ij} for any particular j, given that i makes some change, is given by

$$p_{ij}(\mathbf{x}) \;=\; \frac{\exp\big(f_i(\mathbf{x}(i \leadsto j))\big)}{\sum_{h=1,h\neq i}^{g} \exp\big(f_i(\mathbf{x}(i \leadsto h))\big)} \quad (j \neq i)\,, \tag{17}$$

which can also be written as

$$p_{ij}(\mathbf{x}) \;=\; \frac{\exp\big(f_i(\mathbf{x}(i \leadsto j)) - f_i(\mathbf{x})\big)}{\sum_{h=1,h\neq i}^{g} \exp\big(f_i(\mathbf{x}(i \leadsto h)) - f_i(\mathbf{x})\big)}\ . \tag{18}$$

This probability is also used in multinomial logistic regression, cf. Maddala (1983, p. 60).

8.3 Gratification function

Sometimes the order in which changes could occur, makes a difference for the desirability of the states of the network. For example, if reciprocated ties are generally preferred over non-reciprocated ties, it is possible that the difference in attractiveness between a reciprocated and a non-reciprocated tie is greater for canceling an existing tie than for extending a new tie: i.e., for actor i the existence of the tie from j to i will make it more attractive to extend the reciprocating tie from i to j if it did not already exist; but if the latter tie does exist, the reciprocation will have an even stronger effect making it very unattractive to withdraw the reciprocated tie from i to j. Such a difference between creating and canceling ties cannot be represented by the objective function. For this purpose the gratification function can be used as another model ingredient.

The ***Gratification Function*** $g_i(\mathbf{x}, j)$ of actor i
is the value attached by this actor
(in addition to what follows from the objective function)
to the act of changing the tie variable x_{ij} from i to j,
given the current network configuration $\mathbf{x}$.

Thus, the gratification function represents the gratification to i obtained – in addition to the change in objective function – when changing the current network $\mathbf{x}$ into $\mathbf{x}(i \rightsquigarrow j)$.

When a gratification function is included in the model, actor i chooses to change x_{ij} for that other actor j for whom

$$f_i\big(\mathbf{x}(i \rightsquigarrow j)\big) \; + \; g_i(\mathbf{x},\, j) \; + \; U(j)$$

is largest. Under the assumption of the Gumbel distribution for the residuals $U(j)$, this leads to the conditional choice probabilities

$$p_{ij}(\mathbf{x}) \; = \; \frac{\exp\big(f_i(\mathbf{x}(i \rightsquigarrow j)) + g_i(\mathbf{x}, j)\big)}{\sum_{h=1, h\neq i}^{g} \exp\big(f_i(\mathbf{x}(i \rightsquigarrow h)) + g_i(\mathbf{x}, h)\big)} \quad (j \neq i). \tag{19}$$

Again it can be convenient to subtract $f_i(\mathbf{x})$ within the exponential function, cf. the difference between (18) and (17).

The dissolution and creation of ties work in precisely opposite ways if

$$g_i(\mathbf{x}(i \rightsquigarrow j), j) \; = \; -g_i(\mathbf{x}, j) \; ;$$

note that $g_i(\mathbf{x}(i \rightsquigarrow j), j)$ is the gratification obtained for changing $\mathbf{x}(i \rightsquigarrow j)$ back into $\mathbf{x}$. If this condition holds there is no need for a gratification function, because its effects could be represented equally well by the objective function. The gratification function will usually be a sum of terms some of which contain the factor $(1 - x_{ij})$ while the others contain the factor x_{ij} ; the first-mentioned terms are active for creating a tie (where initially $x_{ij} = 0$), while the other are active for dissolution of a tie (where initially $x_{ij} = 1$). Such effects cannot be represented by the objective function. The further specification is discussed in Section 9.3.

8.4 Intensity matrix

The ingredients of the actor-oriented model, described above, define a continuous-time Markov chain on the space $\mathcal{X}$ of all digraphs on this set of g actors.

The intensity matrix in the representation (9) is given by

$$\begin{aligned} q_{ij}(\mathbf{x}) &= \lim_{dt \downarrow 0} \frac{1}{dt} \mathrm{P}\{\mathbf{X}(t+dt) = \mathbf{x}(i \leadsto j) \mid \mathbf{X}(t) = \mathbf{x}\} \\ &= \lambda_i(\mathbf{x})\, p_{ij}(\mathbf{x}) \end{aligned} \tag{20}$$

where $p_{ij}(\mathbf{x})$ is given by (19), or by (17) if there is no gratification function. Expression (20) is the rate at which actor i makes ministeps, multiplied by the probability that, *if* he makes a ministep, he changes the arc variable X_{ij}.

This Markov chain can be simulated by repeating the following procedure. Start at time t with digraph $\mathbf{x}$.

1. Define

   $$\lambda_+(\mathbf{x}) = \sum_{i=1}^{g} \lambda_i(\mathbf{x})$$

   and let Δt be a random variable with the exponential distribution with parameter $\lambda_+(\mathbf{x})$.
2. The actor i who makes the ministep is chosen randomly with probabilities $\lambda_i(\mathbf{x})/\lambda_+(\mathbf{x})$.
3. Given this i, choose actor j randomly with probabilities (19).
4. Now change t to $t + \Delta t$ and change x_{ij} to $(1 - x_{ij})$.

9 Specification of the actor-oriented model

The principles explained above have to be filled in with a specific model for the objective, rate, and gratification functions. These functions will depend on unknown parameters like in any statistical model, which are to be estimated from the data. When modeling longitudinal network data by actor-oriented models it will often be useful first to fit models with only an objective function (i.e., where the rate function is constant and the gratification function is nil). In a later stage, non-constant rate and gratification functions may be brought into play. At the end of section 9.2 some instances are discussed where it may be advisable to specify a non-constant rate function right from the start of modeling. A wide range of specifications could be given for the three functions. Below, specifications are given most of which were proposed in Snijders (2001) and which are implemented in the *SIENA* software (Snijders and Huisman, 2003).

9.1 Objective function

The objective function is represented as a weighted sum dependent on a parameter $\beta = (\beta_1, \ldots, \beta_L)$,

$$f_i(\beta, \mathbf{x}) = \sum_{k=1}^{L} \beta_k \, s_{ik}(\mathbf{x}) \ . \tag{21}$$

The functions $s_{ik}(\mathbf{x})$ represent meaningful aspects of the network, as seen from the viewpoint of actor i. Some potential functions $s_{ik}(\mathbf{x})$ are the following.

1. *Density effect*, defined by the out-degree
   $s_{i1}(\mathbf{x}) = x_{i+} = \sum_j x_{ij}$;
2. *reciprocity effect*, defined by the number of reciprocated ties
   $s_{i2}(\mathbf{x}) = x_{i(r)} = \sum_j x_{ij} \, x_{ji}$;

3. *transitivity effect*, defined by the number of transitive patterns in i's ties as indicated in Figure 3. A transitive triplet for actor i is an ordered

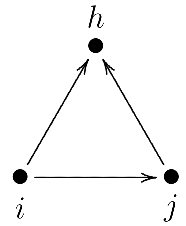

Figure 3: Transitive triplet

pairs of actors (j, h) to both of whom i is tied, while also j is tied to h. The transitivity effect is given by
$s_{i3}(\mathbf{x}) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{jh}$;

4. *balance*, defined by the similarity between the outgoing ties of actor i and the outgoing ties of the other actors j to whom i is tied,

$$s_{i4}(\mathbf{x}) = \sum_{j=1}^{g} x_{ij} \sum_{\substack{h=1 \\ h\neq i,j}}^{g} (b_0 - \mid x_{ih} - x_{jh} \mid) \ , \tag{22}$$

where b_0 is a constant included to reduce the correlation between this effect and the density effect. Given that the density effect is included in the model, the value of b_0 only amounts to a reparametrization of the model (viz., a different value for the parameter of the density effect). The proposed value is such that it yields a zero average for (22) over the first $M-1$ observed networks $\mathbf{x}(t_m)$ $(m = 1, \ldots, M-1)$ and over all actors, and is given by

$$b_0 = \frac{1}{(M-1)g(g-1)(g-2)} \sum_{m=1}^{M-1} \sum_{i,j=1}^{g} \sum_{\substack{h=1 \\ h\neq i,j}}^{g} \mid x_{ih}(t_m) - x_{jh}(t_m) \mid ;$$

5. *number of geodesic distances two effect*, or indirect relations effect, defined by the number of actors to whom i is indirectly tied (through one intermediary, i.e., at geodesic distance 2),
$s_{i5}(\mathbf{x}) = \sharp\{ j \mid x_{ij} = 0,\ \max_h (x_{ih}\, x_{hj}) > 0 \}$;

6. *popularity effect*, defined by the sum of the in-degrees of the others to whom i is tied,
   $s_{i6}(\mathbf{x}) = \sum_j x_{ij}\, x_{+j} = \sum_j x_{ij} \sum_h x_{hj}$;

7. *activity effect*, defined by the sum of the out-degrees of the others to whom i is tied, which is equal to the number of actors h who can be reached from i by a path $i \rightarrow j \rightarrow h$ of length two,
   $s_{i7}(\mathbf{x}) = \sum_j x_{ij}\, x_{j+} = \sum_j x_{ij} \sum_h x_{jh}$.

The conceptual interpretations of the effects 3–5 are closely related, and some further discussion may be helpful for their explanation. The formula for balance is motivated by writing it as the sum of centered similarities between i and those to whom he his tied. The similarity between the ties of actors i and j to the same third actor h can be expressed as $(1- \mid x_{ih} - x_{jh} \mid)$,which is 1 if $x_{ih} = x_{jh}$ and 0 otherwise. Formula (22) can be written as

$$\sum_{j=1}^{g} x_{ij}\,(r_{ij}(\mathbf{x}) - \bar{r})$$

where r_{ij} is the number of equal outgoing tie variables of i and j,

$$\begin{aligned} r_{ij}(\mathbf{x}) &= \sum_{\substack{h=1 \\ h\neq i,j}}^{g} (1 - \mid x_{ih} - x_{jh} \mid) \qquad (23) \\ \bar{r} &= \frac{1}{(M-1)g(g-1)(g-2)} \sum_{m=1}^{M-1} \sum_{i,j=1}^{g} \sum_{\substack{h=1 \\ h\neq i,j}}^{g} r_{ij}(\mathbf{x}(t_m)) \ . \end{aligned}$$

(Note that the average $\bar{r}$ is calculated not for the current network $\mathbf{x}$ but over all $M-1$ networks that figure as initial observations for time periods $(t_m,\, t_{m+1})$.)

It is more customary in network analysis to base balance on a similarity measure defined by the correlation or Euclidean distance between rows and columns of the adjacency matrix, cf. Wasserman and Faust (1994). This would be possible here, too, but the number of matches is used here because correlations or Euclidean distances are not very appropriate measures for vectors with only 0 and 1 entries.

Positive transitivity and balance effects, and negative number-of-distances-two effects, all represent some form of network closure. This can be seen from the fact that local maxima for these effects are achieved by networks consisting of several disconnected complete subgraphs, which are maximally closed networks; where a local maximum is defined as a digraph for which the said function decreases whenever one arc is shifted to another location (which keeps the density constant). These three effects differ in the precise representation of network closure. To get some more insight in their differences, it may be instructive to write them in ways that exhibit their similarities. The number of transitive triplets can be written as

$$s_{i3}(\mathbf{x}) \;=\; \sum_j x_{ij} \sum_h x_{ih}\, x_{hj} \ ,$$

and the number of distances two as

$$s_{i5}(\mathbf{x}) \;=\; \sum_j (1 - x_{ij}) \max_h (x_{ih}\, x_{hj}) \ .$$

The structure of these two functions is similar, a sum over other actors j of a variable involving third actors h, with the following differences. First, the factor x_{ij} in the definition of $s_{i3}(\mathbf{x})$ implies that the summation over other actors j is made only over those to whom i has a tie, whereas the factor $(1 - x_{ij})$ in the definition of $s_{i5}(\mathbf{x})$ means that values are summed over those j to whom i does *not* have a tie – this accounts for the fact that $s_{i3}(\mathbf{x})$ indicates a positive and $s_{i5}(\mathbf{x})$ a negative network closure effect. Second, for the third actors h in $s_{i3}(\mathbf{x})$ the *number* of actors h is counted through whom there is a two-path $\{i \to h,\, h \to j\}$ whereas in $s_{i5}(\mathbf{x})$ only the existence of *at least one* such two-path counts.

The basic component of the balance function is

$$\sum_j x_{ij}\, r_{ij}(\mathbf{x}) \;=\; \sum_{\substack{j,h=1 \\ j \neq h}}^{g} x_{ij}\, (1 + 2x_{ih}\, x_{jh} - x_{ih} - x_{jh}) \ ;$$

some calculations show that this is equal to

$$2s_{i3}(\mathbf{x}) + s_{i1}(\mathbf{x})\Big(g - 1 - s_{i1}(\mathbf{x})\Big) - s_{i7}(\mathbf{x}) \ .$$

This demonstrates that the balance effect includes the number of transitive triplets and, in addition, a quadratic function of the out-degree $s_{i1}(\mathbf{x})$ which is maximal if the out-degree is equal to $(g-1)/2$, and the negative activity effect.

Non-linear functions of the effects $s_{ik}(\mathbf{x})$ could also be included. For example, in order to represent more complicated effects of the out-degrees, one or more of the following could be used in addition to the density effect.

8. *Out-degree truncated at* c, where c is some constant, defined by
   $s_{i8}(\mathbf{x}) = \max(x_{i+}\,,\, c)$;

9. *square root out-degree* $- c \times$ *out-degree*, defined by
   $s_{i9}(\mathbf{x}) = \sqrt{x_{i+}} - c x_{i+}$,
   where c is a constant chosen by convenience to diminish the collinearity between this and the density effect;

10. *squared* (*out-degree* $- c$), defined by
    $s_{i10}(\mathbf{x}) = (x_{i+} - c)^2$,
    where again c is a constant chosen to diminish the collinearity between this and the density effect.

The squared out-degree has a graph-theoretic interpretation, which can be seen as follows. The number of two-stars outgoing from vertex i is

$$\tfrac{1}{2} \sum_{\substack{j,h=1 \\ j \neq h}}^{g} x_{ij}\, x_{ih} \;=\; \binom{x_{i+}}{2} \,,$$

a quadratic function of the out-degree x_{i+}. Therefore, including as effects the out-degree and the squared out-degree of actor i is equivalent to including as effects the out-degree and the number of outgoing two-stars of this actor.

When covariates are available, the functions $s_{ik}(\mathbf{x})$ can be dependent on them. For network data, a distinction should be made between actor-bound covariates v_i and dyadic covariates w_{ij}. The main effect for a dyadic covariate w_{ij} is defined as follows.

11. *Main effect of W (centered)*, defined by the sum of the values of w_{ij} for all others to whom i is tied,
$s_{i11}(\mathbf{x}) = \sum_j x_{ij} \, (w_{ij} - \bar{w})$
where $\bar{w}$ is the mean value of w_{ij} .

For each actor-dependent covariate V the following three effects can be considered:

12. *V-related popularity*, defined by the sum of the covariate over all actors to whom i is tied,
$s_{i12}(\mathbf{x}) = \sum_j x_{ij} \, v_j$;

13. *V-related activity*, defined by i's out-degree weighted by his covariate value,
$s_{i13}(\mathbf{x}) = v_i \, x_{i+}$;

14. *V-related dissimilarity*, defined by the sum of absolute covariate differences between i and the others to whom he is tied,
$s_{i14}(\mathbf{x}) = \sum_j x_{ij} \mid v_i - v_j \mid$.

Of course actor-dependent covariates can be represented by dyadic covariates, e.g., the three preceding effects can be represented, respectively, by main effects of the dyadic covariates $w_{ij} = v_i$, $w_{ij} = v_j$, and $w_{ij} = \mid v_i - v_j \mid$.

9.2 Rate function

The time scale at which networks change may well be quite different from the physical time scale of clocks. Therefore physical time elapsed between observations will usually have a tenuous relation with the amount of change between observed networks. If there are more than two observation moments, a natural first specification is to treat the rate of change within each period (t_m, t_{m+1}) as a free parameter ρ_m , without an a priori relation to the time difference $(t_{m+1} - t_m)$.

When actor-bound covariates are available, they could have an effect on the rate of change. An important class of examples is the following. In some cases there are size differences between actors that are associated with differences in change rate of their networks. For example, in studies of relations between companies, big companies may have more ties but also change ties more quickly than small companies. Another example is that individuals who are socially very active may have many outgoing ties and may also change these more quickly than those who are less active. Therefore, if some measure of size or activity is available, this could be used as an explanatory variable both in the objective function (as an activity effect) and in the rate function.

Since the rate of change is necessarily positive, a covariate must be related to the rate function in such a way that the rate function will always stay positive. Often it will be suitable for this purpose to use an exponential link function (where this term is used as in generalized non-linear modeling, cf. McCullagh and Nelder, 1989). The rate function then can be defined as

$$\rho_i(\alpha, \mathbf{x}) = \rho_m \exp\Big(\sum_h \alpha_h v_{hi}\Big) ,$$

where the sum extends over one or more covariates V_h .

The rate of change can also depend on positional characteristics of the actors. A primary positional characteristic is the degree, which can be distinguished in the out-degree, the in-degree, and the number of reciprocated ties

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

The latter statistic is called the *reciprocated degree* of actor i. The dependence of the rate function on the degrees can be defined in such a way that the reciprocity model is obtained as a special case of the actor-oriented model.

As the simplest case, consider the independent arcs model, where the intensity matrix is defined by

$$q_{ij}(\mathbf{x}) = \lambda_{x_{ij}} \ .$$

It will be shown now that this model can be obtained as an actor-oriented model with the objective function defined by only the density effect,

$$f_i(\beta, \mathbf{x}) = \beta_1\, x_{i+}$$

for which

$$f_i(\beta,\, \mathbf{x}(i \leadsto j)) \;-\; f_i(\beta,\, \mathbf{x}) \;=\; \beta_1\, (1 - 2x_{ij}) \;.$$

When the rate function is defined by

$$\rho\, \left\{ (g - 1 - x_{i+})\, e^{\beta_1} \;+\; x_{i+}\, e^{-\beta_1} \right\} \;, \tag{24}$$

formulae (18) and (20) show that the intensity matrix is given by

$$q_{ij}(\mathbf{x}) \;=\; \rho\, e^{\beta_1\, (1 - 2x_{ij})} \;,$$

which can be reformulated to expression (10) by defining $\lambda_0 = \rho e^{\beta_1}$, $\lambda_1 = \rho e^{-\beta_1}$. This shows that this simple actor-oriented model is the same as the independent arcs model.

More generally, Snijders and van Duijn (1997) demonstrated that the reciprocity model is obtained as a special case of the actor-oriented model when the rate function is a linear combination of the in-degree, out-degree, and reciprocated degree. This is a motivation for letting the rate function depend on the degrees by a function of the form (24) if only one of the three degree types is implicated, and by averages of such functions in the case of dependence on two or three of the degree types. An alternative would be, of course, to use the exponential link function also for the degrees.

Summarizing, it is proposed to define the rate function as a product of three factors

$$\lambda_i(\rho, \alpha, \mathbf{x}, m) \;=\; \rho_m \left\{ \exp\Big(\sum_h \alpha_h\, v_{hi} \Big) \right\} \lambda_{i3} \tag{25}$$

where the first factor represents the effect of the period, the second the effect of actor-bound covariates, and the third the effect of actor position, this effect having the form

$$\lambda_{i3} \;=\; \left\{ \frac{x_{i+}}{g-1}\, e^{\alpha_1} \;+\; \frac{g - 1 - x_{i+}}{g-1}\, e^{-\alpha_1} \right\} \tag{26}$$

if the rate depends on the out-degrees, which can be replaced by the same function of the in-degrees or reciprocated degrees. If the rate function depends on two or all three types of degree, λ_{i3} is defined as an average of such functions (cf. Snijders and van Duijn, 1997).

The discussion motivating formula (24) implies that the actor-oriented model specified by the rate function (26) – a reparametrization of (24) – and an objective function (21) including the density effect $\beta_1 \, x_{i+}$, subsumes as a special case the independent arcs model (viz., for $\alpha_1 = -\beta_1$, and $\beta_k = 0$ for all $k \geq 2$). Since the independent arcs model is suitable as an 'empty' reference model, this gives a special theoretical role to the rate function (26).

A model with a constant rate function (i.e., a rate function not depending on covariates or positional characteristics) is usually easier to explain and can be simulated in a simpler and therefore quicker way. The latter is an advantage given the time-consuming algorithm for estimation (see below). Therefore, in many cases it is advisable to start modeling using a constant rate function, and add the complexity of a non-constant rate function only at a later stage. However, exceptions can occur, e.g., if there are important size differences between the actors in the network – which can be reflected by exogenously given covariates but also by, e.g., the out-degrees as an endogenous network characteristic. The effect of such a size measure on the rate of change can be so predominant that modeling can be biased, and even the convergence of the estimation algorithm can be jeopardized, if such an indicator of size is not included as an effect on the rate function.

9.3 Gratification function

The gratification function can also be defined conveniently as a weighted sum

$$g_i(\gamma, \mathbf{x}, j) = \sum_{h=1}^{H} \gamma_h \, r_{ijh}(\mathbf{x}) \ . \tag{27}$$

Some possible functions $r_{ijh}(\mathbf{x})$ are the following. Recall that when $r_{ijh}(\mathbf{x})$ includes a factor x_{ij} it refers to the gratification experienced for breaking a tie, whereas the inclusion of a factor $(1 - x_{ij})$ refers to gratification for creating a tie.

1. *Breaking off a reciprocated tie*:
   $r_{ij1}(\mathbf{x}) = x_{ij} \, x_{ji}$;

2. *number of indirect links for creating a new tie*, representing the fact that indirect links (at geodesic distance 2) to another actor may facilitate the creation of a new tie:
   $r_{ij2}(\mathbf{x}) = (1 - x_{ij}) \sum_h x_{ih} x_{hj}$;

3. *effect of dyadic covariate W on breaking off a tie:*
   $r_{ij3}(\mathbf{x}) = x_{ij}\, w_{ij}$.

10 MCMC Estimation

The network evolution model is too complicated for explicit calculation of probabilities or expected values, but it can be simulated in a rather straighforward way. This is exploited in the method for parameter estimation which was first proposed in Snijders (1996) and elaborated for the present model in Snijders (2001). Here we sketch only the estimation method for the actor-oriented model with a constant rate function ρ_m between t_m and t_{m+1} , and without a gratification function. This sketch is restricted to the so-called conditional estimation method. A more precise and general treatment, background references, and a motivation of the estimation method are presented in Snijders (2001).

10.1 Method of moments

The observed networks are denoted $\mathbf{x}^{\text{obs}}(t_m)$, $m = 1, \ldots, M$. Suppose that the objective function is given by (21),

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

Then greater values of β_k are expected to lead for all actors i to higher values of the statistics $s_{ik}(\mathbf{X}(t_{m+1}))$, when starting from a given preceding network $\mathbf{x}^{\text{obs}}(t_m)$. The principle of estimation now is to determine the parameters β_k in such a way that, summed over i and m, the expected values of these statistics are equal to the observed values. These observed target values are denoted

$$s_k^{\text{obs}} = \sum_{m=1}^{M-1} \sum_{i=1}^{g} s_{ik}(\mathbf{x}^{\text{obs}}(t_{m+1})) \quad (k = 1, \ldots, L) \tag{28}$$

and collected in the vector s^{obs}. For historical reasons this approach to estimation by fitting 'observed' to 'expected' has in statistical theory the name of *method of moments* (Bowman and Shenton, 1985). Since in our case the expected values cannot be calculated explicitly, they are estimated from simulations.

The simulations in the conditional estimation method run as follows.

1. For two digraphs $\mathbf{x}$ and $\mathbf{y}$ define their distance by

$$\|\mathbf{x} - \mathbf{y}\| = \sum_{i,j} |x_{ij} - y_{ij}| , \tag{29}$$

   and for $m = 1, \ldots, M-1$ let c_m be the observed distances

$$c_m = \|\mathbf{x}^{\text{obs}}(t_{m+1}) - \mathbf{x}^{\text{obs}}(t_m)\| . \tag{30}$$

   This method of estimation is called 'conditional' because it conditions on these values c_m .

2. Use the given parameter vector $\beta = (\beta_1, \ldots, \beta_L)$ and the fixed rate of change $\lambda_i(\mathbf{x}) = 1$.

3. Make the following steps independently for $m = 1, \ldots, M-1$.

   (a) Define the time (arbitrarily) as 0 and start with the initial network

$$\mathbf{X}_m(0) = \mathbf{x}^{\text{obs}}(t_m) . \tag{31}$$

   (b) Simulate, as described in Section 8.4, the actor-oriented model $\mathbf{X}_m(t)$ until the first time point, denoted R_m , where $\|\mathbf{X}_m(R_m) - \mathbf{x}^{\text{obs}}(t_m)\| = c_m$.

4. Calculate for $k = 1, \ldots, L$ the generated statistics

$$S_k = \sum_{m=1}^{M-1} \sum_{i=1}^{g} s_{ik}(\mathbf{X}_m(R_m)) . \tag{32}$$

This simulation yields, for the input parameter vector β, as output the random variables $(S, R) = (S_1, \ldots, S_L, R_1, \ldots, R_{M-1})$. Note that the time parameter within the m'th simulation runs from 0 to R_m.

For the estimation procedure, it is desired to find the vector $\hat{\beta}$ for which the expected and observed vectors are the same,

$$\mathcal{E}_{\hat{\beta}} S = s^{\text{obs}} . \tag{33}$$

This is called the *moment equation*.

10.2 Robbins-Monro procedure

The procedure of Snijders (2001) for approximating the solution to the moment equation is a variation of the Robbins-Monro (1951) algorithm. Textbooks on stochastic approximation contain further explanations and particulars about such algorithms, e.g., Pflug (1996) and Chen (2002). It is a stochastic iteration method. Denote the initial value by $\beta^{(0)}$. (This could be a value obtained from fitting an earlier, possibly simpler, model, or the initial estimate mentioned in Section 13.2.) This procedure consists of three phases. The first phase is of a preliminary nature, with the purpose of roughly estimating the sensitivity of the expected value of S_k to variations in β_k; in the second phase the estimate is determined; and the third phase is for checking the resulting estimate and calculating the standard errors.

1. From a relatively small number (we use $n_1 = 7 + 3L$) of simulations estimate the derivatives

   $$\frac{\partial}{\partial \beta_k} \mathcal{E}_\beta S_k$$

   in $\beta = \beta^{(0)}$ by the averages of the corresponding difference quotients, using common random numbers. Denote by D_0 the diagonal matrix with these estimates as diagonal elements.

2. Set $\beta^{(1)} = \beta^{(0)}, \; a = 0.5, \; n_2 = L + 207$.
   Repeat a few times (advice: 4 times) the following procedure.

(a) Repeat for $n = 1, \ldots, n_2$:
for the current $\beta^{(n)}$ simulate the model in the way indicated above, and denote the resulting value of S by $S^{(n)}$. Update β by

$$\beta^{(n+1)} = \beta^{(n)} - a\, D_0^{-1} \left(S^{(n)} - s^{\text{obs}} \right) .$$

(b) Update β by

$$\beta^{(1)} = \frac{1}{n_2} \sum_{n=1}^{n_2} \beta^{(n)} .$$

(c) Redefine $a = a/2,\ n_2 = 2^{4/3}(n_2 - 200) + 200$.

3. Define the estimate $\hat{\beta}$ as the last calculated value $\beta^{(1)}$. From a rather large (e.g., $n_3 = 500$ or 1000) number of simulations with $\beta = \hat{\beta}$ estimate the covariance matrix $\hat{\Sigma}$ of S and, using common random numbers, the partial derivative matrix D with elements

$$d_{hk} = \frac{\partial}{\partial \beta_k} \mathcal{E}_\beta\, S_h \ .$$

Finally calculate the estimation covariance matrix by

$$\text{cov}(\hat{\beta}) = \hat{D}^{-1}\, \hat{\Sigma}\, (\hat{D}^{-1})' \ . \tag{34}$$

Step 2(a) is called a subphase of Phase 2. Note that from one subphase to the next the initial value $\beta^{(1)}$ changes, the updating factor a decreases, and the number of simulations n_2 increases.

The standard errors of the elements of $\hat{\beta}$ are the square roots of the diagonal elements of $\text{cov}(\hat{\beta})$ in (34). The simulations of Phase 3 can also be used to check if, for this value $\hat{\beta}$, the moment equation (33) is indeed approximately satisfied. The procedure is an instance of MCMC (Markov chain Monte Carlo) estimation because it is based on Monte Carlo simulations and the provisional estimates $\beta^{(n)}$ in each subphase are a Markov chain.

The parameter a is called the gain parameter and can initially have any value between 0 and 1. Values closer to 0 will lead to a less mobile value for $\beta^{(n)}$ and consequently may require more steps for going from the starting value to a good final estimate, but will lead to a more stable procedure.

When the algorithm has come close to the solution of the moment equation (which often happens rather quickly), the provisional values $\beta^{(n)}$ during the steps in 2(a) carry out a random dance about this solution. The reason for taking the average in step 2(b) is that the average of such a collection of random positions is a better estimate than the last value.

The parameters ρ_m are usually of minor substantive importance. They can be estimated by

$$\hat{\rho}_m = \frac{\bar{R}_m}{t_{m+1} - t_m} \tag{35}$$

where $\bar{R}_m$ is the average of the simulated time lengths for period m during Phase 3.

10.3 Missing data

It is hard to collect complete network data at multiple repeated occasions, and therefore it is of practical importance to have a reasonable procedure for dealing with missing data. There can be several reasons why data are missing.

If the composition of the set of actors in the network has changed during the observation period, with some actors joining and/or some actors leaving the group, this can be dealt with by reflecting this changing composition in an appropriate specification of the network evolution model, where only the actors present at the given moment can be involved in tie changes. This is elaborated by Huisman and Snijders (2003).

For other cases, when the composition of the network is constant and it is reasonable to assume that the missing data is due to random non-response, the following procedure is proposed. The procedure is designed to be simple and to minimize the influence of the missing data on the results.
(1.) For the initial networks (31) used in the simulations, missing arc variables $x_{ij}^{\mathrm{obs}}(t_m)$ are replaced by the value 0.
(2.) For the observed statistics s^{obs} in (28) as well as the simulated statistics S in (32) used in the estimation algorithm, an arc variable x_{ij} is replaced by 0 if it is missing for at least one of the observations $\mathbf{x}^{\mathrm{obs}}(t_m)$ or $\mathbf{x}^{\mathrm{obs}}(t_{m+1})$.

This procedure is implemented in SIENA (Snijders and Huisman, 2003) and used in the example of Section 12.

11 Testing

Standard statistical theory about estimation by the method of moments (e.g., Bowman and Shenton, 1985) yields the expression given in (34) for the estimation covariance matrix,

$$\mathrm{cov}(\hat{\beta}) \;=\; \hat{D}^{-1}\,\hat{\Sigma}\,(\hat{D}^{-1})' \;.$$

If the parameter estimates $\hat{\beta}_k$ are approximately normally distributed, the null hypothesis that a single element of the parameter vector is zero,

$$H_0: \;\beta_k \;=\; 0 \;,$$

can be tested by the t-statistic

$$t_k \;=\; \frac{\hat{\beta}_k}{\text{s.e. }(\hat{\beta}_k)} \tag{36}$$

in the standard normal distribution. The same procedure can be followed for the parameters α_k of the rate function and γ_k of the gratification function.

It is plausible that the parameter estimates are indeed approximately normally distributed, but at this moment a proof is not available. It would be useful to conduct simulation studies supporting the validity of this t-test.

12 Actor-oriented model results for the example

The example introduced in Section 3 was analysed using SIENA version 1.92 (Snijders and Huisman, 2003).

In addition to the structural effects, effects of three covariates were considered: gender, programme, and smoking. Gender and smoking are dummy variables coded 1 for female and 2 for male and, respectively, 1 for smoking and 2 for non-smoking. Programme is a numerical variable coded 2, 3, 4 for the length in years of the programme followed by the students. Greater similarity on this variable indicates a greater opportunity for interaction. All covariates are centered by SIENA (i.e., the mean is subtracted), including the dissimilarity variables defined as $(\mid v_i - v_j \mid - c)$, where c is the average of all $\mid v_i - v_j \mid$ values.

First several models were fitted provisionally to explore which are the most important effects. Next to the reciprocity effect, the distance-two effect appeared to be the main structural effect. Of the covariate effects, all three similarity effects as well as the gender activity effect seemed important. In order to avoid misspecifying the gender effect in the objective function, the gender popularity effect also was retained. The rate function seemed dependent on the out-degrees. There seemed to be no strong gratification function effects. Therefore, Table 3 presents the results for a model including these effects; for the sake of simplicity, this model further assumes that – except for the constant factors in the rate function – all parameters are constant throughout the period from t_0 to t_6. For the definition of the rate parameters, the numerical values of the time lengths $t_{m+1} - t_m$ are arbitrarily set equal to 1.0.

As a check on the assumption of constant parameters, Figure 4 gives the parameter estimates obtained for each period separately, with approximate confidence intervals extending two standard errors to either side of the parameter estimate. For the period t_0–t_1–t_2 a common vector of parameters was estimated because the period t_0–t_1, due to the very sparse network at t_0 (average degree 0.2), led to unstable results. In view of the widths of the error bars, the graphs in this figure show that there is no strong evidence for parameter differences. Adding to the model of Table 3 the other two network closure effects, transitivity and balance, led to non-significant t-tests for these parameters, while this did not make disappear the number-of-distances-two effect. Also the other effects mentioned in Section 9 were not significant. It can be concluded that Table 3 may be regarded as a reasonable representation of the network evolution in the whole observation period.

The table shows, judging by the t-ratios of parameter estimate divided by standard error, that there is strong evidence for the reciprocity effect and the network closure effect expressed by a relatively low number of distances two.

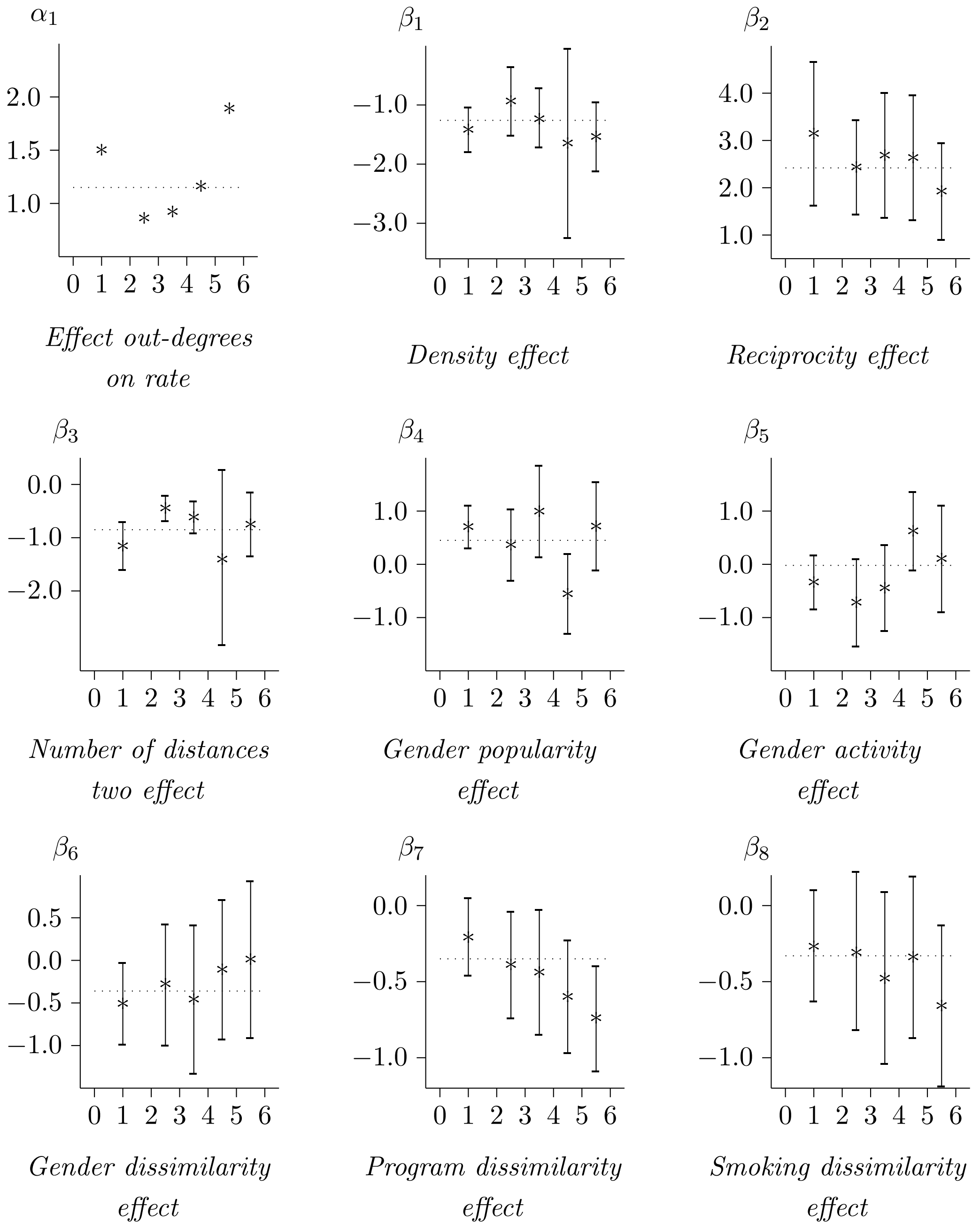

Figure 4: Parameter estimates (with bars extending two standard errors to either side) separately for periods t_0–t_2, t_2–t_3, t_3–t_4, t_4–t_5, t_5–t_6. The dotted lines indicate the corresponding parameter estimates from Table 3. The upper left figure does not show bars because these would all extend outside the figure.

	Effect	Estimate	Standard error
Rate function			
ρ_0	Rate parameter t_0–t_1	24.84	4.57
ρ_1	Rate parameter t_1–t_2	5.43	0.93
ρ_2	Rate parameter t_2–t_3	5.82	0.99
ρ_3	Rate parameter t_3–t_4	4.01	0.67
ρ_4	Rate parameter t_4–t_5	4.62	0.59
ρ_5	Rate parameter t_5–t_6	3.77	0.53
α_1	Out-degree effect on rate	1.15	0.44
Objective function			
β_1	Density	−1.26	0.09
β_2	Reciprocity	2.42	0.25
β_3	Number of distances 2	−0.85	0.08
β_4	Gender popularity	0.45	0.13
β_5	Gender activity	−0.02	0.15
β_6	Gender dissimilarity	−0.36	0.14
β_7	Program dissimilarity	−0.35	0.07
β_8	Smoking dissimilarity	−0.33	0.09

Table 3: Parameter estimates for model with (except rate parameters) constant parameters throughout period $t_1 - t_6$.

The fact that the latter effect is significant and not the transitive triplets effect (see Figure 3) indicates that what drives the network closure is not an extra attraction for individual i to other individuals j based on the *number* of indirect connections $i \to h \to j$, but rather the attraction to others j to whom i has *at least one* such indirect tie. The covariate effects show that male students tend to attract more choices than females, and similarity on gender, program, and smoking behavior leads to a higher likelihood of a tie; male and female students do not differ in the propensity to make choices. An interpretation of the numerical values of the parameter estimates is given in Section 13.3.

13 Parameter interpretation in the actor-oriented model

The interpretation of the quantitative values of the parameters in the actor-oriented model is given here with the help of some rough approximations. This section only treats the model where the change rates are constant.

13.1 Rate of change parameter

The expected number of changes per time unit during the period (t_m, t_{m+1}) is ρ_m for each of the actors. However, two subsequent changes in the same arc variable X_{ij} will cancel each other. In the unobserved interval between t_m and t_{m+1} , some of the changes therefore will be reversals to the situation observed at t_m . This implies that for each actor, the expected number of observed tie differences between the two observations will be a bit less than $\rho_m(t_{m+1} - t_m)$. The extent to which it falls below the latter value will be considerable when ρ_m is so large that the stochastic process is getting near to the equilibrium distribution. Therefore if $\rho_m(t_{m+1} - t_m)$ is small compared to $g - 1$, the expected value of the average number of changes observed per actor per unit of time,

$$\frac{1}{g\,(t_{m+1} - t_m)} \, \|\mathbf{X}(t_{m+1}) - \mathbf{X}(t_m)\| \ , \tag{37}$$

where $\|.\|$ is defined in (29), will be close to ρ_m . As $\rho_m(t_{m+1} - t_m)$ increases, this expected value will increase less than proportionately. The consequence is that $\hat{\rho}_m$ will be close to (37) if this results in a small value of $\rho_m(t_{m+1} - t_m)$, and the ratio of $\hat{\rho}_m$ to (37) will increase as a function of the observed number of changes.

13.2 Density parameter

As a prologue to the interpretation of the other parameters, note that if all parameters of the objective and gratification functions are zero and the rate of change is ρ, then the variables $X_{ij}(t)$ follow independent arc processes and (20) implies that the parameters are $\lambda_0 = \lambda_1 = \rho/(g - 1)$. The limiting digraph distribution of this process is the random graph with density 0.5. This is the 'null process' of the actor-oriented model.

For the interpretation of the parameter β_1 for the density effect $s_{i1}(\mathbf{x}) = \sum_j x_{ij}$, consider the actor-oriented model that contains just this effect, with constant change rate ρ and without a gratification function. In this model the rows $(X_{i1}(t), \ldots, X_{ig}(t))$ follow independent stochastic processes. The intensity matrix (20) is given by

$$q_{ij}(\mathbf{x}) \;=\; \frac{\rho\, e^{\beta_1\,(1-2x_{ij})}}{(g-1-x_{i+})\, e^{\beta_1} \;+\; x_{i+}\, e^{-\beta_1}} \;.$$

If the number g of actors is large and the out-degrees are small relative to the number of actors, this can be roughly approximated by

$$q_{ij}(\mathbf{x}) \;\approx\; \frac{\rho\, e^{-2\,\beta_1\, x_{ij}}}{g-1} \;,$$

which is the intensity matrix of the independent arcs model for

$$\lambda_0 \;=\; \frac{\rho}{g-1} \;,\; \lambda_1 \;=\; \frac{\rho\, e^{-2\,\beta_1}}{g-1} \;.$$

Using the results of Section 5, this implies that, for each actor i, the log-odds will tend to $2\beta_1$ and the out-degree will for $t \to \infty$ fluctuate about the asymptotic value

$$\frac{(g-1)\,\lambda_0}{\lambda_0 + \lambda_1} \;=\; \frac{(g-1)\, e^{2\beta_1}}{1 + e^{2\beta_1}} \;. \tag{38}$$

For example, if $\beta_1 = 0$, the out-degrees will tend to be $(g-1)/2$ on average. (Symmetry considerations imply that the latter result is true even though in this case the out-degrees are not small relative to g.) An exact analysis (not further discussed here) shows that for $t \to \infty$ and fixed g, the asymptotic expected value of X_{i+} is

$$\frac{(2g-3)\, e^{2\beta_1} \;+\; 1}{2 \;+\; 2e^{2\beta_1}} \;, \tag{39}$$

which is quite close to (38). All this suggests that, for the usual cases where network densities are much lower than 0.5, a negative density parameter is expected.

Now suppose that this process is observed at times t_1 and t_2. Then (2) and (11) imply that

$$\frac{r_1}{1 - r_0}\ , \tag{40}$$

the fraction of $X_{ij} = 0$ which turned into 1, divided by the fraction of $X_{ij} = 1$ which turned into 0, is expected to be $\lambda_0/\lambda_1 = e^{2\beta_1}$. Therefore an estimate for β_1 is half the corresponding log odds,

$$\tfrac{1}{2}\log\left(\frac{N_{01}}{(N_{01} + N_{00})}\ \frac{(N_{10} + N_{11})}{N_{10}}\right)\ ,$$

where N_{hk} is defined as in (1).

This can be used for an initial estimate for the estimation method of Section 10.2 in the case where M observations are available, even when more effects than just the density are included. This initial estimate is given by

$$\hat{\beta}_1\ =\ \tfrac{1}{2}\log\left(\frac{\sum_{m=1}^{M-1} N_{01}(m)}{\sum_{m=1}^{M-1}(N_{01}(m) + N_{00}(m))}\ \frac{\sum_{m=1}^{M-1}(N_{10}(m) + N_{11}(m))}{\sum_{m=1}^{M-1} N_{10}(m)}\right) \tag{41}$$

for the density effect and $\hat{\beta}_k = 0$ for all $k \geq 2$ (the other effects).

The interpretation of β_1 as approximately half the log-odds for the set of arc variables X_{ij} in an equilibrium situation, and the interpretation based on (40), do not hold any more for models that include other effects in addition to the density effect. The difference in interpretation will depend on the extent to which the parameters for the other included effects lead to lower or higher overall densities of the network. But in many practical applications we still observe negative estimates for β_1 as a reflection of the fact that the network density in a hypothetical equilibrium situation would be clearly less than 0.5.

13.2.1 Network boundary effects

What happens with these models if they are applied to networks for which the network boundary has been defined in a rather generous way – so that the number g of actors is large and only a small fraction of the network members would be candidate relational partners for any actor? Such a situation can be modeled by letting g tend to infinity while keeping the out-degrees X_{i+} finite. This is just the assumption made above for the approximation of the actor-oriented model by the independent arcs model. In the approximating limiting distribution the log-odds was found to tend to $2\beta_1$, which corresponds for the out-degrees to a binomial distribution with a mean of

$$\frac{(g-1)\, e^{2\beta_1}}{1+e^{2\beta_1}} \ ,$$

which tends to infinity with g. This is at odds with the assumption that the out-degrees X_{i+} remain finite. However, if we let

$$\beta_1 = \eta - \frac{1}{2} \log(g-1) \ ,$$

for some fixed number η, the limiting distribution tends to the Poisson distribution with mean $e^{2\eta}$, which does remain finite and is independent of g.

This suggests that if we first consider a certain network with g_0 actors, and then add further actors most of which are not relevant to the actors present earlier, so the number of ties from the earlier present actors to the new actors is quite small, we should expect the density parameter slowly to decrease, by a term slightly less than $\frac{1}{2} \log \big((g-1)/(g_0-1)\big)$.

13.3 Other parameters

Section 13.2 shows that, already for an objective function consisting only of the density effect, quite crude approximations are required to make descriptive statements about the probability distributions corresponding to certain parameter values, and these descriptions do not take us very far.

Another way to obtain insight into the parameter values is to consider the implied objective function, which indicates the preferences of the actors. For the example as presented in Table 3, this function is

$$\begin{aligned} f_i(\mathbf{x}) \ = \ & \sum_j \Big\{ -1.26 + 2.42\, x_{ji} + 0.45\, v_{1j} - 0.36 \mid v_{1i} - v_{1j} \mid \\ & \qquad - 0.35 \mid v_{2i} - v_{2j} \mid - 0.33 \mid v_{3i} - v_{3j} \mid \Big\}\, x_{ij} \\ & - 0.85 \sum_j (1 - x_{ij}) \max_h (x_{ih}\, x_{hj}) \end{aligned}$$

where (due to the centering applied) $v_{1i} = -0.25$ for female and 0.75 for male students; the program variable v_{2i} has values -1.3, -0.3, and 0.7; and $v_{3i} = -0.6$ for smokers and 0.4 for non-smokers. The contribution of the gender activity effect was set to 0.

This expression can be brought into clearer shape by some recoding. Denote $z_{1i} = 1$ for male and 0 for female students, $s_{1ij} = 1$ if students i and j have the same gender and 0 otherwise, the program similarity variable $s_{2ij} = 2 - \mid v_{2i} - v_{2j} \mid$, and $s_{3ij} = 1$ if students i and j have the same smoking behavior and 0 otherwise. Then the s_{hij} are similarity variables, equal to 0 in the case of the greatest dissimilarity. The objective function then is

$$
\begin{aligned}
f_i(\mathbf{x}) \;=\; & \sum_j \Big\{ -2.78 \,+\, 2.42\, x_{ji} \,+\, 0.85 \max_h(x_{ih}\, x_{hj}) \,+\, 0.45\, z_{1j} \\
& \qquad + 0.36\, s_{1ij} \,+\, 0.35\, s_{2ij} \,+\, 0.33\, s_{3ij} \Big\}\, x_{ij} \\
& - 0.85 \sum_j \max_h(x_{ih}\, x_{hj}) \ .
\end{aligned}
$$

The first two lines show that, e.g., for a male actor i in program $v_{2i} = 2$, creating a new tie to a female student who did not already choose i as a friend, of different smoking behavior and in program $v_{2j} = 4$, to whom no length-two path exists, leads to an objective function loss of 2.78. The third line implies that for each student h chosen by j who was not already chosen by any of i's present friends, creating the new tie from i to j leads to an additional loss for i of 0.85. Such students h would point to a lack of embeddedness of j in i's current network. On the other hand, if j already chose i as a friend, while the other characteristics are as mentioned, the first two lines imply a loss of only $2.78 - 2.42 = 0.36$. This loss is approximately nullified if the potential friend j has the same smoking behavior or is in program $v_{2j} = 3$. A very crude summary of the preceding is that a tie to another student is worthwhile only if the tie is reciprocated and there also is similarity at least on one variable.

The total contribution of gender for male students is nil for choosing a female friend, and $0.45 + 0.36 = 0.81$ for choosing a male friend; for female students it is 0.36 for choosing a female and 0.45 for choosing a male friend. Thus for female students the value of a friendship with a male or a female other student is about the same, while male students have a clear preference for friendships to other males.

The value of already being chosen by the other (equal to 2.42) is about thrice as large as the value of already having at least one indirect tie to the other (0.85); the latter value is about the same as the advantage, for males, that males have over females (0.81) and slightly larger than the advantage of following the same program compared to following the most different programs ($2 \times 0.35 = 0.70$); and about two and a half times the value of having the same smoking behavior (0.33).

14 Discussion

Longitudinal network data can yield important insights into social processes, but these insights can be obtained only when using adequate models for data analysis. There exist many models for network evolution that are not accompanied by methods for statistical data analysis, and recently there has been quite a surge in publications about such models stimulated, e.g., by applications to the growth of the world wide web. However, in order to know how strong and how uncertain are the conclusions we may draw from empirical data, and in order to know the extent to which our models are, or are not, supported by the empirical data – which will steer the development of extended or new models in directions that are empirically fruitful, it is desirable to have indeed a statistical component in models for network evolution. The requirement of statistical evaluation leads to parsimony and modesty in model building. The complexity of network dynamics, in which everything seems to depend on everything else, implies that even modest models are mathematically quite complex, as is demonstrated by the models of this chapter. These models are (as far as I know) the first statistical models for network evolution that allow a variety of endogenous network effects, of which the various types of network closure effects (transitive triplets, number of pairs at a geodesic distance equal to 2, balance, as presented in Section 9.1) are primary examples. I hope that the availability of these models and of the software to analyse data according to these models (the SIENA program that is included in the StOCNET system which can be downloaded from `http://stat.gamma.rug.nl/stocnet/`, see Snijders and Huisman, 2003, and see also Chapter 13 in this volume) will be a stimulus for the collection and statistical evaluation of longitudinal network data.

One of the assumptions in the actor-oriented model is that actors optimize myopically, considering only the situation to be obtained immediately after the next change they are going to make. It would be theoretically interesting to elaborate models with more farsighted actors, but the risk is that such models would be less robust and more limited to specific applications than the simpler myopic models. The interpretation of the myopic models is that the effects in the objective and gratification functions represent what the actors try to achieve in the short run, and do not directly reflect their goals in the long run.

The further application of these models should also indicate the points where they must be further extended and modified to provide a better fit to empirical data and to be better aligned with the theoretical questions that researchers may have. The actor-oriented approach explained here, and its implementation using the rate, objective, and gratification functions, is quite flexible and open for extension by a variety of effects in addition to those mentioned here; but other models can also be proposed. One example is the alternative actor-oriented model of Snijders (2003) in which the focus is on giving a good fit to the observed out-degrees. Another example would be a tie-oriented or dyad-oriented model, driven not by changes made by optimizing actors but by changes in tie variables, which would be closely compatible with the exponential random graph models proposed by Frank and Strauss (1986) and Wasserman and Pattison (1996) and treated in Chapters 8–10 of this volume; these tie changes could be according to Gibbs or Metropolis-Hastings steps as described in Snijders (2002). Testing goodness of fit of network evolution models, which will give empirical indications for model modifications and extensions, is the topic of Schweinberger (2003).

The approach presented here can be extended also by considering more complex data sets. A multilevel approach to network evolution, in which the data is composed of multiple parallel networks that evolve according to a similar model but with different parameters, was initiated by Snijders and Baerveldt (2003), and may be further extended. As the mutual influence between networks and behavior is theoretically as well as practically important, research also is under way about modeling the simultaneous evolution of networks and individual behavior. The models presented in this chapter have a rich potential for applications but perhaps an even richer potential for further extensions.

References

Banks, D.L., and K.M. Carley. 1996. Models for network evolution. *Journal of Mathematical Sociology, 21*: 173 – 196. Also published in Doreian and Stokman (1997).

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.

Chen, H.-F. 2002. *Stochastic Approximation and its Applications*. Dordrecht etc.: Kluwer.

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

de Nooy, W. 2002. The dynamics of artistic prestige. *Poetics* 30: 147 – 167.

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

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

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.

Huisman, M., and T.A.B. Snijders. 2003. Statistical analysis of longitudinal network data with changing composition. In press, *Sociological Methods and Research.*

Kalbfleisch, J.D., and Lawless, J.F. 1985. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* 80:

863-871.

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.

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.

Leenders, R.Th.A.J. 1996. Evolution of friendship and best friendship choices. *Journal of Mathematical Sociology* 21: 133 – 148. Also published in Doreian and Stokman (1997).

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.

McCullagh, P., and Nelder, J.A. 1989. *Generalized Linear Models*, 2^{nd} ed. London: Chapman & Hall.

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

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

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

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

Sanil, A., D.L. Banks, and K.M. Carley. 1994. Models for evolving fixed-node networks: Model fitting and model testing. *Social Networks* 17, 1–26.

Schweinberger, M. 2003. Testing goodness of fit of network evolution models. *Submitted.*

Snijders, T.A.B. 1995. Methods for longitudinal social network data. Pp. 211 – 227 in *New Trends in Probability and Statistics, Vol. 3: Multivariate Statistics and Matrices in Statistics*, edited by E.-M. Tiit, T. Kollo, and H. Niemi. Vilnius, Lithuania: TEV and Utrecht, The Netherlands: VSP.

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. 2001. The Statistical Evaluation of Social Network Dynamics. Pp. 361-395 in *Sociological Methodology – 2001*, edited by M.E. Sobel and M.P. Becker. Boston and London: Basil Blackwell.

Snijders, T.A.B. 2002. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure* vol. 3, number 2.

Snijders, T.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. Washington, DC: The National Academies Press.

Snijders, T.A.B, and C. Baerveldt. 2003. A Multilevel Network Study of the Effects of Delinquent Behavior on Friendship Evolution. *Journal of Mathematical Sociology* 27: 123 – 151.

Snijders, T.A.B., and J.M. Huisman. 2003. *Manual for SIENA version 1.98.* Groningen: ICS, University of Groningen.
Obtainable from `http://stat.gamma.rug.nl/stocnet/` .

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.

Snijders, T.A.B., and M.A.J. van Duijn. (2002). Conditional Maximum Likelihood Estimation under Various Specifications of Exponential Random Graph Models. Pp. 117 – 134 in *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank*, edited by J. Hagberg. Stockholm: Department of Statistics, University of Stockholm.

Taylor, H.M., and Karlin, S. 1998. *An introduction to stochastic modeling*, 3^d ed. New York: Academic Press.

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

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.

van Duijn, M.A.J., E.P.H. Zeggelink, J.M. Huisman, F.N. Stokman, and F.W. Wasseur. 2003. Evolution of Sociology Freshmen into a Friendship Network. *Journal of Mathematical Sociology* 27: 153–191.

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

Wasserman, S. 1978. Models for binary directed graphs and their applications. *Advances in Applied Probability* 10: 803 – 818.

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. 1987. The conformity of two sociometric relations. *Psychometrika* 53: 261–282.

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