Stochastic Actor-Oriented Models for Network Dynamics

Tom A.B. Snijders

University of Groningen, University of Oxford

Accepted for publication as a chapter in

Annual Review of Statistics and Its Application, Volume 4 (2017).

CONTENTS

1 INTRODUCTION

Social phenomena are increasingly viewed from a network perspective, and social networks accordingly receive more and more attention in various domains of science. Graphs and their dynamics have been studied in probability theory since more than half a century (Erdős and Rényi, 1960; Bollobaš, 1985). Network analysis has since long been an interdisciplinary empirical field rooted in sociology, anthropology, discrete mathematics, and visualization for which one date in its early development may be given as the creation of its own journal in 1978 (Freeman, 2004). From about the turn of the millennium there has been a growing interest in networks also from physicists and computer scientists (Watts and Strogatz, 1998). This widened interdisciplinary field since 2013 was further institutionalized with the journal *Network Science* (Brandes et al., 2013).

Statistical modeling of networks has played an important role in these developments, and has grown in scope as networks received more and more scientific attention (Goldenberg et al., 2009; Kolaczyk, 2009). This chapter presents a statistical methodology for the analysis of longitudinal network data collected in a panel design, in which repeated observations are made of a network on a constant node set. Development of this model was inspired by applications in sociology and other social sciences (e.g., Snijders and Doreian, 2010, 2012), and therefore the terminology will have a social science flavor. E.g., the nodes of the network will be referred to as social actors – which could be humans, but also organisations, etc.

2 MODEL DEFINITION

We assume that repeated observations $x(t_1), \ldots, x(t_M)$ on a directed graph (digraph), referred to as a network, are available for some $M \geq 2$. The node set $\mathcal{N} = \{1, \ldots, n\}$ is constant, while the ties are variable. The network x is identified with its $n \times n$ adjacency matrix $x = (x_{ij})$ of which the elements denote whether there is a tie from node i to node j ($x_{ij} = 1$) or not ($x_{ij} = 0$). Self-ties are not allowed, so that the diagonal is structurally zero. Random variables are denoted by capitals.

Various models have been proposed for network dynamics, most of them being Markov processes of some kind. The basic heuristic idea of actor-oriented models (Snijders, 2001) is that the nodes of the graph are social actors having the potential to change their outgoing ties, and the observed network dynamics is the result of the sequences of choices by these actors. The digraph develops as a continuous-time Markov process $X(t)$ even though being observed only at M discrete time points. The stochastic process $X(t)$ is modeled as being right-continuous. The constraints are that ties may change only one by one, and actors do not coordinate their changes of ties. Thus, at any given moment, at most one tie variable X_{ij} can be changed. This constraint was proposed by Holland and Leinhardt (1977), and it has the virtue of splitting the change process in its smallest possible constituents, reducing the model definition to the specification of rates of creating and terminating single ties.

The actor-oriented model is specified as follows; further discussion and motivation is given in Snijders (2001).

A. Opportunities for change. Each actor i gets at stochastically determined moments the opportunity to change one of the outgoing tie variables $X_{ij}(t)$ ($j \in \mathcal{N}$, $j \neq i$). Since the process is assumed to be Markovian, waiting times between opportunities have exponential distributions. Each of the actors i has a *rate function* $\lambda_i(\alpha, x)$ which defines how quickly this actor gets an opportunity to change a tie variable, when the current value of the digraph is x, where α is a parameter. At any time point t with $X(t) = x$, the waiting time until the next opportunity for change by any actor is exponentially distributed with parameter

$$\lambda(\alpha, x) = \sum_i \lambda_i(\alpha, x) \ . \tag{1}$$

Given that an opportunity for change occurs, the probability that it is actor i who gets the opportunity is given by

$$\pi_i(\alpha, x) = \frac{\lambda_i(\alpha, x)}{\lambda(\alpha, x)} \ . \tag{2}$$

The rate functions can be constant between observation moments, or they can depend on functions $u_{ik}(x)$ which may be covariates or positional characteristics of the actors, such as outdegrees $\sum_j x_{ij}$. The time moments of observation are arbitrary, not necessarily equidistant, and observation is assumed not to affect the random process. Arbitrary time spans between observations can be represented by multiplicative parameters ρ_m. Actor-dependent rate functions can be modeled, e.g., by an exponential link function,

$$\lambda_i(\rho, \alpha, x) = \rho_m \ \exp\Big(\sum_k \alpha_k u_{ik}(x)\Big) \ , \tag{3}$$

valid between the moments of observing $x(t_m)$ and $x(t_{m+1})$.

B. Options for change. When actor i gets the opportunity to make a change, this actor has a permitted set $\mathcal{A}_i(x^0)$ of values to which the digraph may be changed, where x^0 is the current value of the digraph. The idea that actors may change their outgoing ties, but only one tie variable at the time, is reflected by

$$\mathcal{A}_i(x^0) \subset \{x^0\} \cup \mathcal{A}_i^{\mathrm{r}}(x^0) \tag{4a}$$

where $\mathcal{A}_i^{\mathrm{r}}(x^0)$ is the set of adjacency matrices differing from x in exactly one element,

$$\mathcal{A}_i^{\mathrm{r}}(x^0) = \{x \mid x_{ij} = 1 - x_{ij}^0 \text{ for one } j \neq i, \text{ and } x_{j'k} = x_{j'k}^0 \text{ for all other } (j', k)\}. \tag{4b}$$

Including x^0 in $\mathcal{A}_i(x^0)$ can be important for expressing the property that actors who are satisfied with the current network will prefer to keep it unchanged. Therefore, the usual model is $\mathcal{A}_i(x^0) = \{x^0\} \cup \mathcal{A}_i^{\mathrm{r}}(x^0)$.

It is assumed that the network dynamics is driven by a so-called *objective function* $f_i(\beta, x^0, x)$ that can be interpreted as the relative attractiveness for actor i of moving from the network represented by x^0 to the network x, and where β is a parameter. The conditional probability that the next digraph is x, given that the current digraph is x^0 and actor i gets the opportunity to make a change, is assumed to be given by the multinomial model

$$p_i(\beta, x^0, x) = \begin{cases} \dfrac{\exp\big(f_i(\beta, x^0, x)\big)}{\sum_{\tilde{x}} \exp\big(f_i(\beta, x^0, \tilde{x})\big)} & x \in \mathcal{A}_i(x^0) \\ 0 & x \notin \mathcal{A}_i(x^0) \end{cases} \tag{5}$$

where the summation extends over $\tilde{x} \in \mathcal{A}_i(x^0)$. This formula can be motivated by a random utility argument as used in econometrics (e.g., Maddala, 1983), where it is assumed that the actor maximizes $f_i(\beta, x^0, x)$ plus a random disturbance having a standard Gumbel distribution. Assumption (4) implies that instead of $p_i(\beta, x^0, x)$ we can also write $p_{ij}(\beta, x^0)$ where the correspondence between x and j is defined as follows: if $x \neq x^0$, j is the unique element of $\mathcal{N}$ for which $x_{ij} \neq x_{ij}^0$;

the arbitrary definition is made that $j = i$ if $x = x^0$. This less redundant notation will be used in the sequel. Thus, for $j \neq i$, $p_{ij}(\beta, x^0)$ is the probability that, under the condition that actor i has the opportunity to make a change and the current digraph is x^0, the change will be to $x_{ij} = 1 - x^0_{ij}$ with the rest unchanged; while $p_{ii}(\beta, x^0)$ is the probability that, under the same condition, the digraph will not be changed.

The most usual models are based on objective functions that depend on x only. This has the possible interpretation that actors wish to maximize a function $f_i(\beta, x)$ (plus a random residual) that itself is independent of "where they come from", although the previous state does determine the set of possible next states. The greater generality of (5), where the objective function can depend also on the previous state x^0, makes it possible to model path-dependencies, or hysteresis, where the loss suffered from withdrawing a given tie differs from the gain from creating this tie, even if the rest of the network has remained unchanged. In the literature on this model, when the objective function does not depend on the previous state, it is also called an evaluation function.

Various ingredients for specifying the objective function were proposed in Snijders (2001). A linear form is convenient,

$$f_i(\beta, x^0, x) = \sum_{k=1}^{L} \beta_k \, s_{ik}(x^0, x) \ , \tag{6}$$

where the functions $s_{ik}(x^0, x)$, called *effects*, are determined by subject-matter knowledge, available scientific theory, and formulation of research questions. The effects should represent essential aspects of the network structure, assessed from the point of view of actor i, such as

$$s_{ik}(x^0, x) = \sum_j x_{ij} \qquad \text{(outdegree)} \tag{7}$$

$$\sum_j x_{ij}\, x_{ji} \qquad \text{(reciprocated ties)} \tag{8}$$

$$\sum_{j,k} x_{ij}\, x_{jk}\, x_{ik} \qquad \text{(transitive triplets)} \tag{9}$$

$$\sum_j x^0_{ij}\, x^0_{ji}\, x_{ij}\, x_{ji} \qquad \text{(persistent reciprocity);} \tag{10}$$

they can also depend on covariates – such as resources and preferences of the actors, or costs of exchange between pairs of actors – or combinations of network structure and covariates. See Snijders et al. (2010) for a basic list of effects, and Ripley et al. (2016) for an extensive list.

2.1 Intensity matrix; time-homogeneity

The model description given above defines $X(t)$ as a continuous-time Markov process with Q-matrix or intensity matrix (e.g., Norris, 1997) for $x \neq x^0$ given by

$$q(x^0, x) = \begin{cases} \lambda_i(\alpha, x^0)\, p_i(\beta, x^0, x) & \text{if } x \in \mathcal{A}_i(x^0),\ i \in \mathcal{N} \\ 0 & \text{if } x \notin \cup_i \mathcal{A}_i(x^0). \end{cases} \tag{11}$$

The assumptions do not imply that the distribution of $X(t)$ is stationary. The transition distribution, however, is time-homogeneous, except for time dependence reflected by time-varying components in the functions $s_{ik}(x^0, x)$, and except for multiplicative constants as explained in the next paragraph.

The time points $t_1, \ldots, t_M$ of the observations can be used for marking time-heterogeneity of the transition distribution. For example, covariates may be included with values allowed to change at the observation moments. A special role is played here by the time durations $t_m - t_{m-1}$ between successive observations.

Standard theory for continuous-time Markov chains (e.g., Norris, 1997) shows that the matrix of transition probabilities from $X(t_m)$ to $X(t_{m-1})$ is $e^{(t_m - t_{m-1})Q}$, where Q is the matrix with elements (11). Thus, changing the duration $t_m - t_{m-1}$ can be compensated by multiplication of the rate function $\lambda_i(\alpha, x)$ by a constant. The connection between an externally defined real-valued time variable t_m and the rapidity of network change is tenuous at best, and the parameter ρ_m in (3) absorbs effects of the durations $t_m - t_{m-1}$. This makes the numerical values t_m unimportant.

2.2 Simulation algorithm

The following steps can be used for simulating the process for the time period from t_m to t_{m+1}.

1. Set $t = t_m$ and $x = x(t_m)$.
2. Generate ΔT according to the exponential distribution with parameter $\lambda(\alpha, x)$.
3. If $t + \Delta T > t_{m+1}$ set $t = t_{m+1}$ and stop.
4. Select randomly $i \in \{1, ..., n\}$ using probabilities
$$\frac{\lambda_i(\alpha, x)}{\lambda(\alpha, x)} \ .$$
5. Select randomly $x' \in \mathcal{A}_i(x)$ using probabilities $p_i(\beta, x, x')$ (see (5)).
6. Set $t = t + \Delta T$.
7. Set $x = x'$.
8. Return to step (2).

If the continuous-time stochastic process $X(t)$ would be observed for $t_1 \leq t \leq t_M$ including all the time increments ΔT generated in the above simulation steps the model would be a generalized linear model with two parts, (3) and (6). Given that the stochastic process $X(t)$ is observed only for $t = t_1, t_2, \ldots, t_M$, the model can be regarded as a generalized linear model with a lot of missing data.

3 ESTIMATION

Exact calculations for this model are infeasible, except for uninterestingly simple special cases. Since the model can be simulated in a straightforward manner, it lends itself to simulation-based estimation procedures.

3.1 Method of Moments

The estimation method that is used mostly in practice is the Method of Moments procedure proposed in Snijders (2001). Denote the estimated parameter by $\theta = (\rho, \alpha, \beta)$. The statistics used for estimation are heuristically derived as follows. For each element of θ there is a one-dimensional statistic that is sensitive to this parameter. For ρ_m, influencing the total amount of change, the statistic used is the Hamming distance,

$$D\big(X(t_{m+1}), X(t_m)\big) \;=\; \sum_{i,j} \mid X_{ij}(t_m+1) - X_{ij}(t_m) \mid \;\; . \tag{12}$$

For parameter α_k, determining how strongly the rate of change for actor i is influenced by $u_{ik}(X)$, a sensitive statistic is

$$A_k\big(X(t_{m+1}), X(t_m)\big) \;=\; \sum_{i,j} u_{ik}\big(X(t_m)\big) \mid X_{ij}(t_m+1) - X_{ij}(t_m) \mid \;\; . \tag{13}$$

For linear predictors (6) where $s_{ik}(x^0, x)$ does not depend on x^0, higher values of β_k will tend to lead to networks x for which the value of $s_{ik}(x)$ is higher, for all i. Therefore, for estimating β_k, sensitive statistics are given by

$$s_k\big(X(t_m)\big) \;=\; \sum_{i} s_{ik}\big(X(t_m)\big) \;\; . \tag{14}$$

Combining these statistics and employing the Markov chain assumption for the observed data $x(t_1), \ldots, x(t_M)$, the estimating equations are

$$D\big(x(t_{m+1}), x(t_m)\big) = E_\theta\big\{D\big(X(t_{m+1}), x(t_m)\big) \mid X(t_m) = x(t_m)\big\} \qquad (m = 1, \ldots, M-1) \tag{15a}$$

$$\sum_{m=1}^{M-1} A_k\big(x(t_{m+1}), x(t_m)\big) = \sum_{m=1}^{M-1} E_\theta\big\{A_k\big(X(t_{m+1}), x(t_m)\big) \mid X(t_m) = x(t_m)\big\} \qquad (k = 1, \ldots, K_\alpha) \tag{15b}$$

$$\sum_{m=1}^{M-1} s_k\big(x(t_{m+1})\big) = \sum_{m=1}^{M-1} E_\theta\big\{s_k\big(X(t_{m+1})\big) \mid X(t_m) = x(t_m)\big\} \qquad (k = 1, \ldots, K_\beta), \tag{15c}$$

where K_α and K_β are the number of elements of α and β.

To find the solution of (15), stochastic optimization can be used. Snijders (2001) uses a multivariate version of the Robbins-Monro algorithm (Robbins and Monro, 1951; Ruppert, 1991; Kushner and Yin, 2003) with the improvements proposed by Polyak (1990) and Ruppert (1988). The 'double averaging' technique of Bather (1989) and Schwabe and Walk (1996) turns out to give further improvement. The R package RSiena (Ripley et al., 2016; Snijders, 2016a) implements this in a three-phase algorithm, where the first phase is for roughly determining the sensitivity of expected statistics to parameters; the second phase conducts Robbins-Monro parameter updates based on simulations of the network dynamics for the current parameter value; and the third is for determining how well (15) is indeed approximated, and for determining standard errors. For the latter purpose the derivatives of the expected values with respect to the parameters are needed, and these can be estimated by the score function method (Schweinberger and Snijders, 2007; Rubinstein, 1986); this is elaborated below. This algorithm is quite robust, although rather time-consuming.

3.2 Generalized Method of Moments

The statistics (14) used for the Method of Moments depend only on the networks observed at the end of each period $(t_{m-1}, t_m]$. It might be possible to obtain greater efficiency by utilizing more statistics, e.g., function of both the networks observed at the beginning and the end of each period. E.g., for estimating the parameter β_k for the reciprocity effect $s_k(x) = \sum_j x_{ij} x_{ji}$, (14) leads to

$$\sum_{i,j} X_{ij}(t_m)\, X_{ji}(t_m) \ ,$$

a measure for the amount of reciprocity observed at t_m. Another relevant statistic would be

$$\sum_{i,j} \big(1 - X_{ij}(t_{m-1})\big) X_{ji}(t_{m-1})\, X_{ij}(t_m)\, X_{ji}(t_m) \ ,$$

measuring the amount of observed reciprocation from t_{m-1} to t_m. Utilizing more statistics than parameters is possible with the generalized method of moments (Burguete et al., 1982; Hansen, 1982). The principle is, for a q-dimensional statistic S with outcome denoted s and a p-dimensional parameter θ with $q > p$, to estimate θ such that

$$\big(s - E_\theta S\big)' W \big(s - E_\theta S\big) \tag{16}$$

is minimal as a function of θ. When $S = \mathrm{cov}_\theta(S)$ this estimator will be at least as efficient as the method of Moments estimator for any subvector of S. The further elaboration and the numerical implementation of a simulated generalized method of moments for this model is presented in Amati et al. (2015).

3.3 Likelihood-based methods

Bayesian and frequentist likelihood-based estimators were developed by Koskinen and Snijders (2007) and Snijders et al. (2010). These estimators use data augmentation (Tanner and Wong, 1987). The procedure is explained here only for the case of constant rate functions, i.e., $\lambda_i(\alpha, x) \equiv \rho_m$ between t_m and t_{m+1}.

The observed data $x(t_1), \ldots, x(t_M)$ is augmented by the chain of all choosing actors i and intermediate states x' in the simulation algorithm; the time delays ΔT can be integrated out. The intermediate states are the results of the 'choices' by the actors described above; recall that the outcome of the choice may be that the state stays the same. The formulae are given for $M = 2$ waves; generalization to more waves is obtained by concatenating the chains from wave to wave. Denote the sequence of subsequent intermediate states, resulting from these choices, by $x_1 = x(t_1), x_2, \ldots, x_R = x(t_2)$, the actor making the choice for x_h by i_h, and the time delay between x_h and x_{h+1} by ΔT_h. Then R is determined by the requirement

$$\sum_{h=1}^{R-1} \Delta T_h < t_2 - t_1 \leq \sum_{h=1}^{R} \Delta T_h \ . \tag{17}$$

The model assumptions imply that

$$P\{X_h = x_h, i_h = i \mid X_{h-1} = x_{h-1}\} \ = \ \frac{\exp\Big(\sum_k \beta_k s_{ik}(x_h)\Big)}{n \sum_{x \in \mathcal{A}_i(x_{h-1})} \exp\Big(\sum_k \beta_k s_{ik}(x)\Big)} \ , \tag{18}$$

provided that the sequences (i_h) and (x_h) are compatible in the sense that for each h the difference (if any) between x_{h-1} and x_h is in row i_h of the adjacency matrix. Further, the continuous Markov chain assumption implies that the number $R-1$ of choices made between t_1 and t_2 has a Poisson distribution,

$$P\Big\{\sum_{h=1}^{R-1} \Delta T_h < t_2 - t_1 \leq \sum_{h=1}^{R} \Delta T_h\Big\} \ = \ e^{-n\rho(t_2 - t_1)} \frac{\big(n\rho(t_2 - t_1)\big)^{R-1}}{(R-1)!} \ . \tag{19}$$

Using the Markov assumption, the likelihood of the augmented data therefore is given by

$$P\big\{R = r, X_2 = x_2, \ldots, X_{R-1} = x_{r-1},\, I_2 = i_2, \ldots, I_R = i_r \mid X(t_1) = x_1,\, X(t_2) = x_R\big\}$$

$$= e^{-n\rho(t_2-t_1)} \frac{\big(\rho(t_2-t_1)\big)^{r-1}}{(r-1)!} \times \prod_{h=2}^{r} \frac{\exp\big(\sum_k \beta_k s_{i_h k}(x_h)\big)}{\sum_{x \in \mathcal{A}_i(x_{h-1})} \exp\big(\sum_k \beta_k s_{i_h k}(x)\big)} \tag{20}$$

provided that sequences (i_h) and (x_h) are compatible. In the data augmentation method, be it frequentist or Bayesian, the crucial step is to obtain a sample from the conditional distribution of the augmented data, here $X_2, \ldots, X_{R-1}, I_2, \ldots, I_R$, given the observed data, here $\big(x(t_1), x(t_2)\big)$. This is done in Snijders et al. (2010) by a Metropolis-Hastings scheme consisting of a mix of the following proposals. The proposals are formulated in terms of 'toggles', because effectively each change in the network dynamics is a change of some X_{ij} to $(1 - X_{ij})$, called a toggle of (i, j).

1. Paired deletions: two toggles of the same element (i, j) are randomly selected and both deleted.
2. Paired insertions: for a random pair (i, j) $(i \neq j)$, two toggles are inserted at random positions in the chain.
3. Single insertions: for a random node i, a choice by i resulting in a non-change is inserted at a random position.
4. Single deletions: a random choice resulting in a non-change is deleted from the chain.
5. Permutations: a section of the chain is randomly permuted.

This collection of proposals leads to a communicating random process on the space of all possible chains connecting $x(t_1)$ to $x(t_2)$. Using acceptance ratios determined by the proposal probabilities and the likelihood (20), the Metropolis-Hastings procedure converges to the conditional distribution given the observations.

This is used to estimate the posterior distribution in a Bayesian approach in Koskinen and Snijders (2007). To find the Maximum Likelihood (ML) estimate, Snijders et al. (2010) employ the missing data principle of Orchard and Woodbury (1972) and Louis (1982). Denoting the parameter by θ, observed data by X, and augmented data by V, let the observed data score function be $J_X(\theta; x)$ and the total data score function $J_{XV}(\theta; x, v)$. Then

$$E_\theta\{J_{XV}(\theta; x, V) \mid X = x\} = J_X(\theta; x) \ , \tag{21}$$

so that the likelihood equation can be expressed as

$$E_\theta\{J_{XV}(\theta; x, V) \mid X = x\} = 0. \tag{22}$$

Snijders et al. (2010) present a method for solving this equation by a Robbins-Monro stochastic approximation procedure, similar to the one sketched above for the Method of Moments.

3.3.1 Data augmentation and the score function For the calculation of standard errors by the delta method, derivatives of the type

$$\frac{\partial E_\theta S(X)}{\partial \theta}$$

are needed for various statistics $S(X)$; see Schweinberger and Snijders (2007). Under regularity conditions satisfied here, this is equal to

$$\frac{\partial E_\theta S(X)}{\partial \theta} \ = \ E_\theta J_X(\theta, X)\, S(X) \ = \ E_\theta J_{XV}(\theta, X, V)\, S(X) \tag{23}$$

for any data augmentation V. This is a very convenient equation, because in our case X is easily simulated but $J_X(\theta, X)$ is incalculable; this equation allows to choose a data augmentation by unobserved statistics V which is computed anyway for the simulation of V and for which $J_{XV}(\theta, X, V)$ is easy to calculate.

This is used for the standard errors of the Method of Moments estimates (regular as well as generalized), with the data augmentation being the i and x' of each step of the simulation algorithm for the process described in Section 2.2. These are the same statistics as in the data augmentation for the ML estimation, but computed in a different process. This is elaborated in Schweinberger and Snijders (2007), where it is also discussed that the derivatives can also be approximated by ratios of finite differences, but this has various disadvantages.

4 CO-EVOLUTION MODELS

The principle of a continuous-time discrete-state Markov chain can be extended from a changing network to a more complicated outcome space, consisting of the Cartesian product of several spaces.

4.1 Networks and Behavior

The interest in networks is based for a large part on their important consequences for nodal variables of actors – behavioral tendencies, attitudes, performance, etc. These will often not only be influenced by networks, but also themselves exert influence on networks. This leads to scientific interest in the interdependent dynamics of networks and actor variables, e.g., friendships and health-related lifestyle behaviors of adolescents, or collaboration and performance of organisations.

Let $X(t)$ be a changing network (directed graph) on n nodes, as above; and let $Z(t) = \big(Z_1(t), \ldots, Z_n(t)\big)$ be a vector of nodal variables for the same nodes. To keep the treatment in the discrete realm, assume that $Z_i(t)$ is an ordinal discrete variable with values $1, 2, \ldots, K$ for some integer $K \geq 2$. The model presented above can then be duplicated for these two as joint dependent variables: for X and also for Z there are rate functions λ_i^X and λ_i^Z, and objective functions f_i^X and f_i^Z. Parameters in f_i^X are denoted β^X, those in f_i^Z are β^Z. It is assumed that at any moment t, not more than one variable $X_{ij}(t)$ or $Z_i(t)$ can change, so that opportunities for change are either for the network, X, or the behavior, Z. Network changes are modeled as above, except that now the objective function f^X can also depend on the current state of the behavior. Given that actor i has an opportunity for change in behavior, for the current value $z_i(t) = z^0$ the options for the next possible state are $z^0 - 1$, z^0, $z^0 + 1$. The probabilities are

$$p_i^Z(\beta, x, z^0, z) = \begin{cases} \dfrac{\exp\big(f_i^Z(\beta, x, z^0, z)\big)}{\displaystyle\sum_{\delta=-1}^{1} \exp\big(f_i^Z(\beta, x, z^0, z^0 + \delta)\big)} & z = z^0 + \delta;\ \delta \in \{-1, 0, 1\} \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

If this would mean an excursion out of the permitted range $\{1, \ldots, K\}$, z stays the same.

The network-and-behavior dynamics can be simulated from t_m to t_{m+1} by the following steps.

1. Set $t = t_m$, $x = x(t_m)$, $z = z(t_m)$.
2. Generate ΔT according to the exponential distribution with parameter $\lambda_+ = \sum_i \big(\lambda_i^X(\alpha, x, z) + \lambda_i^Z(\alpha, x, z)\big)$.
3. If $t + \Delta T \geq t_{m+1}$, set $t = t_{m+1}$ and stop.

4. Select variable $V \in \{X, Z\}$ with probabilities

   $P\{V\} = \sum_i \lambda_i^V(\alpha, x, z)/\lambda_+$.

5. Select randomly $i \in \{1, ..., n\}$ using probabilities
$$\frac{\lambda_i^V(\alpha, x, z)}{\sum_{i'} \lambda_{i'}^V(\alpha, x, z)} \ .$$

6. If $V = X$, select randomly $x' \in \mathcal{A}_i(x)$ using probabilities $p_i(\beta, x, x')$ (see (5)); set $x = x'$.

   If $V = Z$, select randomly $z' \in \{z-1, z, z+1\}$ using probabilities $p_i^Z(\beta, x, z, z')$ (see (24)); if $1 \le z' \le K$, set $z = z'$.

7. Set $t = t + \Delta T$.

8. Return to step (2).

By letting the objective function f_i^X for the network depend on the network x as well as the behavior z, and doing the same for the objective function f_i^Z for the behavior, a dynamic interdependence between the network and the behavior can be modelled, so that the model represents social selection (where network changes depend on network as well as behavior) together with social influence (where behavior changes depend on behavior as well as network). The possibilities for jointly analysing social selection and social influence are discussed extensively in Steglich et al. (2010).

Estimation of these models by the Method of Moments was discussed in Snijders et al. (2007). We limit ourselves here to the case that the objective functions can be expressed without a dependence on the previous state; the mentioned reference also treats the general case. The objective function for the network then has the form
$$f_i^X(x, z) \ = \ \sum_k \beta_k^X \, s_{ik}^X(x, z) \tag{25}$$

and the objective function for behavior

$$f_i^X(x,z) \;=\; \sum_k \beta_k^Z \, s_{ik}^Z(x,z) \ . \tag{26}$$

Estimation by the Method of Moments works like was explained above, but the statistics still have to be specified. For the dependencies across the two variables, cross-lagged statistics are used, expressing the 'causal' arrow pointing from the, earlier, explanatory variable to the, later, dependent variable. Define the functions

$$s_k^X(x,z) \;=\; \sum_i s_{ik}^X(x,z)$$

and

$$s_k^Z(x,z) \;=\; \sum_i s_{ik}^Z(x,z) \ .$$

For parameters β^X and β_Z, respectively, the moment equations are

$$\begin{aligned}
&\sum_{m=1}^{M-1} E_\theta\big\{s_k^X\big(X(t_{m+1}), z(t_m)\big) \mid X(t_m) = x(t_m), Z(t_m) = z(t_m)\big\} \\
&\qquad\qquad = \sum_{m=1}^{M-1} s_k^X\big(x(t_{m+1}), z(t_m)\big) \qquad\qquad (k = 1, \ldots, K_\beta) \ , \qquad (27)\\
&\sum_{m=1}^{M-1} E_\theta\big\{s_k^Z\big(x(t_m), Z(t_{m+1})\big) \mid X(t_m) = x(t_m), Z(t_m) = z(t_m)\big\} \\
&\qquad\qquad = \sum_{m=1}^{M-1} s_k^Z\big(x(t_m), z(t_{m+1})\big) \qquad\qquad (k = 1, \ldots, K_\beta) \ . \qquad (28)
\end{aligned}$$

These estimators perform quite well, but improved performance is expected from the Generalized Method of Moments. This is the basis of current work, extending Amati et al. (2015).

An algorithm for Maximum Likelihood is also implemented in the R package RSiena, and explained in Snijders (2016a). The principles are the same as for the networks-only model, but there are more details because the changes have more aspects.

4.2 Multivariate Networks

Quite a similar approach can be used to model co-evolution of several networks. Examples are friendship and advice in task-oriented environments, or like and dislike. Multivariate actor-oriented network models were proposed in Snijders et al. (2013). Each network is a dependent variable with its own rate function, objective function, and parameter vector; the objective functions for each will depend on all networks to represent their dynamic interdependence. Denote the networks by $X_1 = (X_{1ij})$ and $X_2 = (X_{2ij})$. Here the dependence has a multilevel aspect (cf. Snijders, 2016b). We give here a number of dependencies at different levels with illustrative components of the objective function for X_2.

- The level of ties or dyads:
  - direct entrainment: $\sum_j x_{1ij}\, x_{2ij}$;
  - mixed reciprocity: $\sum_j x_{1ji}\, x_{2ij}$;
- The level of actors: degrees in one network affect degrees in the other network,
  - outdegree associations: $\big(\sum_j x_{1ij}\big)\big(\sum_j x_{2ij}\big)$;
  - indegree associations: $\big(\sum_{j,h} x_{1hj}\, x_{2ij}\big)$;
- The level of triads, which is also related to composition of relations and algebraic network models. Two examples are, with an interpretation for $X_1 =$ friendship, $X_2 =$ advice:
  - mixed closure, composition $X_1 \circ X_1 \Rightarrow X_2$: $\sum_{j,h} x_{1ih}\, x_{1hj}\, x_{2ij}$; friends of friends tend to become advisors;
  - mixed closure, composition $X_1 \circ X_2 \Rightarrow X_2$: $\sum_{j,h} x_{1ih}\, x_{2hj}\, x_{2ij}$ advisors of friends tend to become advisors.

4.3 Other extensions

The use of a simulation model as a model for data, with simulation-based estimates and tests, gives a lot of freedom to define details of the model and extend or adapt it for particular data structures. This section gives some of these extensions.

Changing composition In many situations it may occur that the actor set is not constant throughout, but changes between the start and end of the observation period. Students may leave or enter school, companies can go broke, organisations may be established. Provided such changes are exogenous, this can be straightforwardly specified in the simulation design, as discussed in Huisman and Snijders (2003). The researcher will have to specify the initial ties, and what 'happens to' the ties after an actor left the network.

Structurally determined values Some ties may be impossible. For example, in the network of advice between judges in Tubaro et al. (2017) some of the actors were available as advisors but could not make any advice choices themselves. It is also possible that some ties are prescribed, such as advice from a superior to a direct subordinate. This implies that some tie variables X_{ij} may be structural zeros or structural ones. Such restrictions can be implemented as specifications of the set $\mathcal{A}_i(x)$ of permitted values of the adjacency matrix for the next step in the simulations.

Ordered tie values Networks with ordinal discrete tie values can be represented by multivariate networks, using transformation to the level networks, defined as follows. Let $x = \big(x_{ij}\big)$ be the adjacency matrix of a network with possible tie values $x_{ij} = 0$ for no tie, and $x_{ij} = 1, \ldots, K$ for some $K \geq 2$, for increasingly stronger ties. This can be represented by K level networks $x^{(1)}, \ldots, x^{(K)}$ defined by

$$x_{ij}^{(k)} = \begin{cases} 0 & x_{ij} < k \\ 1 & x_{ij} \geq k \ . \end{cases} \tag{29}$$

Often an acceptable requirement for the underlying continuous-time process is that at any given time point t, $x_{ij}(t) = k$ can change to the values $k - 1$ or $k + 1$ — to the extent that these are within the range $[0, K]$ — but not make larger instantaneous jumps. An example with $K = 2$ is for weak ($k = 1$) and strong ($k = 2$) ties, and the requirement means that to go from no tie to a strong tie, or vice versa, it is necessary to pass through the intermediate state of a weak tie. This can be implemented by modeling the multivariate network $\big(X^{(1)}(t), \ldots, x^{(K)}(t)\big)$ with restrictions (depending on x, t and k) to the sets $\mathcal{A}_i^{(k)}(x)$, enforcing this requirement. This is elaborated in Snijders and Steglich (2017b).

Duration analysis Analysis of survival times dependent on a network where the network itself evolves in dependence on the set of survivors can be applied, e.g., to diffusion of innovations co-evolving with a network. This can be represented by the model of Section 4.1 with a binary 'behavior' variable that cannot decrease, implemented as a restriction on the permitted set of z' in the simulation model in that section. Such a model was proposed by Greenan (2015). Her model represents the dependence of survival on the network and the covariates in the rate function, not in the objective function. This leads to a network coevolution version of the proportional hazards model.

Non-directed networks A larger difference is the modeling of non-directed networks, defined by the restriction that $x_{ij} = x_{ji}$. Here the actor-oriented framework requires to express the coordination between actors i and j in the choice of the tie value $x_{ij}(t)$. Several models for this coordination are presented in Snijders and Pickup (2017).

5 SOME EXAMPLES

The actor-oriented model has been used in various publications, many of which are listed at `http://www.stats.ox.ac.uk/~snijders/siena/`.

The example presented here is based on social network data collected in the 'Teenage Friends and Lifestyle Study' (Pearson and Michell, 2000; West and Sweeting, 1996).[1] This data set was collected in view of prevention of smoking among adolescents, and it was also used, e.g., in the article explaining the joint analysis of social selection and social influence (Steglich et al., 2010), where smoking was the main behavioral variable. Here only results for the network dynamics are given.

Friendship network data were collected for a cohort of pupils in a school in the West of Scotland. The panel data were recorded over a three year period starting in 1995, when the pupils were aged 13, and ending in 1997. A total of 160 pupils took part in the study, 129 of whom were present at all three measurement points. The friendship networks analyzed here were observed by allowing the pupils to name up to six closest friends. A picture of the second wave is in Figure 1.

[1] The study was funded by the Chief Scientist Office of the Scottish Home and Health Department under their Smoking Initiative (grant K/OPR/17/8), and was executed by Lynn Michell and Patrick West of the Medical Research Council / Medical Sociology Unit, University of Glasgow.

The data set analyzed here consists of the pupils present at all three waves. Based on sociological theories about friendship combined with earlier studies, the model included the following set of effects:

1. The outdegree effect (7).
2. The reciprocity effect (8).
3. To account for transitive closure, i.e., the tendency for friends of friends to be friends, the so-called geometrically weighted edgewise shared partners (‘GWESP’) effect (Snijders et al., 2006; Hunter, 2007). This differs from the transitive triplets effect (9) in that the number of indirect connections does not have a linear effect on the log-probability for the creation of a direct tie, but rather a concave increasing effect:
$$\text{GWESP}(i, \alpha, x) \;=\; \sum_j x_{ij}\, e^{\alpha} \left\{ 1 \;-\; \left(1 - e^{-\alpha}\right)^{\sum_h x_{ih} x_{hj}} \right\} . \qquad (30)$$
Here α is a tuning parameter, which here is chosen at the usual fixed value of $\alpha = \ln(2)$. The GWESP effect fits better than the transitive triplets effect for this and many other data sets.
4. Four degree-related effects, useful for a good representation of the indegree and outdegree distributions:
$$s_{\mathrm{IP}i}(x) \;=\; \sum_j x_{ij}\, x_{+j} \qquad \text{indegree popularity}$$
$$s_{\mathrm{OA}i}(x) \;=\; \sum_j x_{ij}\, x_{i+} \qquad \text{outdegree activity}$$
$$s_{\mathrm{IA}i}(x) \;=\; \sum_j x_{ij}\, x_{+i} \qquad \text{indegree activity}$$
$$s_{\mathrm{RA}i}(x) \;=\; \sum_j x_{ij} \left(\sum_h x_{ih} x_{hi} \right) \qquad \text{reciprocated degree activity .}$$
Here the + denotes summation over the index.

5. Having the same gender,

$$s_{\mathrm{SG}i} = \sum_j x_{ij}\, I\{\text{gender } i = \text{ gender } j\} \ ,$$

   where $I\{A\}$ is the indicator function of the event A.

6. The distance between the dwelling places. This was operationalized as the centered logarithm of the distance between the centroids of the postcode regions:

$$s_{\mathrm{LD}i} = \sum_j x_{ij} \ln(d_{ij}) \ .$$

Estimation was done with the Method of Moments as in Snijders (2001), using the R package RSiena version 1.1-294. To account for the restriction that outdegrees were not larger than 6, the permitted set $\mathcal{A}_i^{\mathrm{r}}(x^0)$ in Section 2 was the set of all digraphs with outdegrees not larger than 6, and differing in no more than one tie variable from x^0.

The results are in Table 1. The interpretation is the following. For the rate parameters, actors have on average 13 and 10 opportunities for change, respectively, in periods 1 (wave 1 – wave 2) and 2 (wave 2 – wave 3). Recall that an opportunity for change does not need to lead to a change, and changes can be canceled before being observed; these estimated rate parameters are therefore higher than the observed average number of changes. The parameters in the objective function can be tested by referring the t-ratio of estimated divided by standard error to an approximate standard normal distribution. No proof is available for this, but it is supported by extensive simulation studies. Since all effects in this model operate for creation of new ties as well as for maintenance of existing ties, positive parameters have the interpretation that the effect works for both these aspects of the dynamics. The outdegree effect balances between creation and termination of ties, given all other effects in the model, and usually is taken for granted in the interpretation; it is negative, reflecting that the probability of befriending arbitrary others is low. There is a strong tendency toward reciprocity ($\hat{\beta}_{\mathrm{R}} = 3.297$) as well as transitive closure as represented by the GWESP effect ($\hat{\beta}_{\mathrm{GWESP}} = 1.856$). This is generally seen for friendship networks. Friendship ties become less likely as the distance between dwellings increases ($\hat{\beta}_{\mathrm{LD}} = -0.18$), and are more likely between same-gender pairs ($\hat{\beta}_{\mathrm{SG}} = 0.636$). High indegrees have a negative effect on receiving and keeping incoming friendships ($\hat{\beta}_{\mathrm{IP}} = -0.055$) and also on creating and maintaining outgoing friendships ($\hat{\beta}_{\mathrm{IA}} = -0.115$). High current outdegrees have a positive effect on creating and maintaining outgoing friendships ($\hat{\beta}_{\mathrm{OA}} = 0.101$). However, all these degree effects should be considered in the light of all other effects, in particular the reciprocated degree-activity effect ($\hat{\beta}_{\mathrm{RA}} = -0.244$). An interpretation is that those who have many reciprocated friendships are satisfied with their network position, this negative parameter showing that they are less active in creating new ties; the strongly positive reciprocity effect implies they keep their existing ties with a high probability; contrasting with this, those with few reciprocated friendships will try to establish more friendships. Note that the maximum outdegree of 6 is a bound which is used as a constraint in the simulation model; this should be taken into account in the interpretation of parameter estimates.

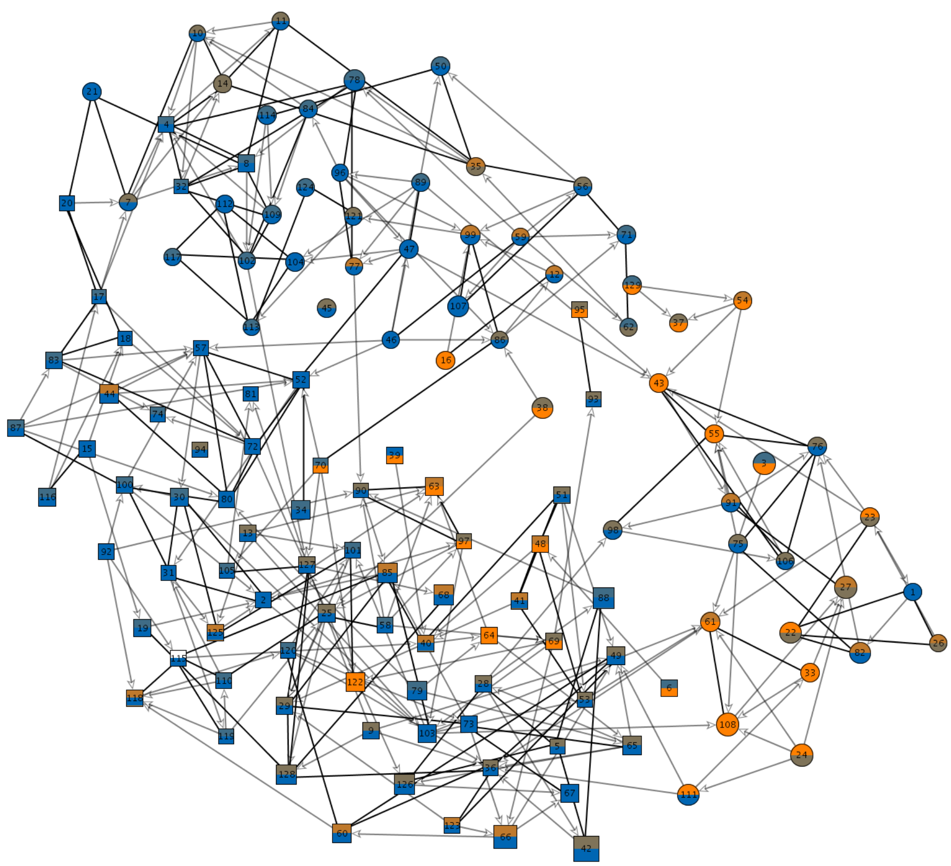

Figure 1: Second wave of West of Scotland friendship network. Arrows denote ties. Boys represented as squares, girls as circles. Color/shading represents smoking (bottom) and drinking (top): orange/light is high, blue/dark is low. Picture made by Krists Boitmanis.

Effect	parameter		(s.e.)
Rate period 1	$\hat{\rho}_1 =$	12.853	(1.375)
Rate period 2	$\hat{\rho}_2 =$	10.217	(1.020)
outdegree	$\hat{\beta}_{\mathrm{OD}} =$	–3.092	(0.224)
reciprocity	$\hat{\beta}_{\mathrm{R}} =$	3.297	(0.271)
GWESP	$\hat{\beta}_{\mathrm{GWESP}} =$	1.856	(0.083)
indegree-popularity	$\hat{\beta}_{\mathrm{IP}} =$	–0.055	(0.020)
outdegree-activity	$\hat{\beta}_{\mathrm{OA}} =$	0.101	(0.035)
reciprocated degree-activity	$\hat{\beta}_{\mathrm{RA}} =$	–0.244	(0.065)
indegree-activity	$\hat{\beta}_{\mathrm{IA}} =$	–0.115	(0.051)
log distance (centered)	$\hat{\beta}_{\mathrm{LD}} =$	–0.180	(0.040)
same gender	$\hat{\beta}_{\mathrm{SG}} =$	0.636	(0.083)

Table 1: Estimation results of actor-oriented model for Glasgow data: parameter estimates and standard errors.

Co-evolution of networks and behavior The models for co-evolution of networks and behavior have been used in various studies of how adolescent life styles and health behaviors are influenced by friends' behaviors, and how important they are for making and losing friends. An overview is the special issue 'Network and Behavior Dynamics in Adolescence' of the *Journal of Research on Adolescence*, see Veenstra et al. (2013).

Multivariate dependent networks A multivariate application was presented by Huitsing et al. (2014). Two interdependent networks of bullying and defending were studied in three elementary schools with a total of 354 children over three waves. The children could nominate others in the school who bullied them, and also others who defended them. Note that bullies can be themselves also defended, which points to cooperation in bullying. Of special interest was how bullying and defending influenced each other, with distinct attention for effects at the dyadic, actor, and triadic levels. At the dyadic level, unexpectedly, no significant effects were found and also at the actor level, where effects of in- and outdegrees were tested, results were not strong and differed somewhat between the three schools. Clear patterns did emerge, however, at the triadic level. For example, in these three groups, if i bullies j and k defends j, then the probability is higher that, later, i will also bully k: defending others leads to being victimized, in turn, by their bullies. Another pattern is that if i defends j and j bullies k, the probability will be higher that, later, i will also bully k. Here the defending is interpreted not as 'defense against bullies' but 'defending bullies against others who try to intervene'. Such patterns can be called mixed triadic closure, generalizing the more well-known pattern of transitive network closure.

6 SOME OTHER RELATED MODELS

There are a lot of non-statistical models for network dynamics, designed for theoretical or other purposes. These are too many to even start reviewing here. Surveys of statistical models for networks were given by Goldenberg et al. (2009) and Kolaczyk (2009). Reviews from an economics, statistical physics, and sociological point of view, respectively, are to be found in Jackson (2008) and Graham (2015); Newman (2010); and Snijders (2011). An overview of dynamic network models from a background of statistical mechanics, but with quite general applications, is Holme and Saramäki (2012). In the following, we briefly mention some statistical longitudinal network models for networks observed at two or more discrete time points.

Some of these are longitudinal versions of the Exponential Random Graph Model, the 'ERGM' of Wasserman and Pattison (1996) and Lusher et al. (2013). Of these, the longitudinal tie-oriented model of Koskinen and Snijders (2013) is closest to the stochastic actor-oriented models presented here. This model is called the longitudinal ERGM (LERGM) in Koskinen et al. (2015). It is a continuous time model, with changes restricted to sequences of changes of single tie variables. The opportunities for change occur randomly for pairs of nodes, and the probability of tie changes, given that some pair of nodes was selected, are defined by a logistic regression model conditional on the rest of the graph just like in the basic ERGM. The model specification can be done in quite a similar way as for the actor-oriented model proposed in this chapter, except that an analogue of the rate function is not straightforward to define and not implemented in current software.

Hanneke et al. (2010) proposed the temporal ERGM (TERGM). This is a discrete time network model for a network time series $X(t_1), \ldots, X(t_M)$ where the conditional distribution of $X(t_m)$ given $X(t_{m-1})$ has an ERGM distribution. They elaborated the case where the $X_{ij}(t_m)$ for $i, j = 1, \ldots, n$ are conditionally independent given $X(t_{m-1})$. This avoids the so-called near-degeneracy problems discussed in Snijders et al. (2006) and Schweinberger (2011), but Lerner et al. (2013) concluded that the restriction to models with conditional independence given the preceding observation does not give a good representation of network dependencies, unless inter-observation times are relatively short. A different presentation of basically the same model was given by Paul and O`Malley (2013).

The TERGM was extended by Krivitsky and Handcock (2014) to the 'separable temporal ERGM (STERGM)' which is a TERGM for which the newly created ties are conditionally independent of the terminated ties. This assumption will not always be tenable, but it may facilitate interpretation of parameters. Their article contains an application implementing triadic network dependence.

Other longitudinal models build on the latent space network models of Hoff et al. (2002). These models postulate a low-dimensional Euclidean space in which the nodes are positioned, with the log-odds of a tie being a linear function of covariates and the distance between the nodes. Durante and Dunson (2014) propose a longitudinal latent space model, using the inner product instead of the distance; whether this is well interpretable and provides a good fit will depend on the application. In their model the positions of the nodes evolve in continuous time according to a Gaussian process.

A different approach is taken by Westveld and Hoff (2011). Since the tie $x_{ij} = 1$ is interpreted as a connection going from i to j, it is customary to call i the *sender* and j the *receiver* of the tie. Westveld and Hoff specify a mixed effects model, with sender and receiver effects, which are correlated over time. In a sense this is a longitudinal version of the p_2 model of van Duijn et al. (2004), but with a less detailed representation of reciprocity.

For the Stochastic Actor-Oriented Model as well as all these other models, estimation procedures have been developed but up to now no mathematical proofs are available of properties such as unbiasedness or consistency. The relevant asymptotics here would be that, with a fixed number of waves M, the number of nodes n tends to infinity while the average degree remains bounded in probability. The network dependence is so difficult that until now it seems these models have resisted attempts at constructing such proofs, although various simulations have led to conjectures that indeed some consistency and asymptotic normality properties should hold. Graham (2015) reviews some identification properties.

7 DISCUSSION

The crucial issue for network modeling is how to represent dependence structures between network ties. For single ('cross-sectional') observations of one network this is very hard, but for longitudinal network data there is at least the arrow of time. The Stochastic Actor-Oriented Model is designed for longitudinal network data collected in a panel design, with 2 or more panel waves. It is a Markov chain model, consisting of a combination of two generalized linear models for the unobserved continuous-time process, but observed only at the times of the panel waves. If change between subsequent waves is not too large, then a strong dependence between the repeated observations may be presumed. A rule of thumb was proposed (Snijders et al., 2010) that Jaccard measures of similarity between successive waves preferably should be .3 or more. This means that the number of remaining ties is not much less than the average of the number of newly created ties and the number of terminated ties. Experience with applications has shown, however, that data sets with larger turnover, and Jaccard similarities between successive waves lower even than 0.2, may also be meaningfully analysed with this model. When the amount of change is somewhat limited in this way, and when it is reasonable to assume that, as an approximation, network change took place as an unobserved sequence of dyadic (as opposed to groupwise) changes, the Stochastic Actor-Oriented Model may be a good representation of network dynamics. As with other generalized linear models there is a wide flexibility to adapt the model specification to the domain of application, research question, and subject-matter knowledge. How to specify the model is discussed at length in Snijders and Steglich (2017a). Some examples are given in Snijders and Steglich (2017b).

The model is available in the R package RSiena. Its manual (Ripley et al., 2016) presents many specification possibilities. Published applications can be found at website `http://www.stats.ox.ac.uk/~snijders/siena/`.

The choice between models, comparing the Stochastic Actor-Oriented Model to other models such as those mentioned in Section 6, should be based on considerations such as goodness of fit, conceptual validity, specification possibilities, interpretability, and practicality. This will depend also on the purposes of the data analysis: e.g., studying network dependencies themselves, or controlling for network structure in tests of effects of actor or dyadic variables. Goodness of fit, in the sense of a good representation of dependence structures, can be assessed by comparing simulated data to observed data in the way discussed for cross-sectional network models by Hunter et al. (2008). Methods for doing this were developed by Lospinoso (2012) and are presented in Snijders and Steglich (2017a). The network effects that can be specified in the Stochastic Actor-Oriented Model and in the various longitudinal variants of the Exponential Random Graph Model give more flexible possibilities to explicitly specify and investigate the details of the network dependence than the latent space or the mixed effect models.

The Stochastic Actor-Oriented Model is based on a continuous-time probability model, while the data are assumed to be collected at discrete time moments, possibly as few as 2. With the free multiplicative parameter ρ_m in the rate function (3), this implies that as long as observation does not interfere with the process, the timing of the observations does not affect the validity of the model. The model refers to a process unfolding in continuous time with a stationary transition distribution (unless some of the parameters are time-dependent) of which the marginal distribution is not necessarily stationary; the meaning and values of parameters other than ρ_m are unrelated to the observation times. This may be an advantage over discrete-time models when modeling processes for which time elapsed between observations is irregular, or does not have a regular consequence on the network changes.

For modeling cross-sectional network data the Exponential Random Graph Model (Wasserman and Pattison, 1996; Lusher et al., 2013) is much used. This model is known to be plagued by the problem of near-degeneracy, which appears as the property that for many quite usual observed data sets, the maximum likelihood estimate is obtained for a distribution concentrating almost all of its probability mass on a few networks, some complete or very dense and some empty or very sparse, for which the expected value of the sufficient statistics is indeed equal to the observed value, but all of which are quite distant from the observed network. This problem is discussed, e.g., in Snijders et al. (2006), Rinaldo et al. (2009), and Schweinberger (2011), and the former reference proposes specifications that avoid near-degeneracy. Because of its longitudinal nature, this problem does not affect the Stochastic Actor-Oriented Model; depending on the model specification, it may, however, apply to its limiting stationary distribution. For practical application, this is of no direct concern. There is a relation with model specification, however, because the specifications for which the limiting distribution could be near-degenerate might also be prone to having a worse fit when applied to panel data. As an example, transitivity for cross-sectional networks cannot be modeled well by the Exponential Random Graph Model with the count of transitive triplets among the sufficient statistics; this statistic leads to near-degenerate models. For the purpose of modeling transitivity, it can be replaced by the geometrically weighted edgewise shared partners ('GWESP') effect which is analogous to (30), see Hunter (2007). The count of transitive triplets defines a linear effect of the number of indirect connections on the log-odds for the existence of a direct tie; the GWESP statistic defines a concave increasing effect, and its sublinearity counters the near-degeneracy. For the longitudinal Stochastic Actor-Oriented Model the transitive triplets effect is (9); this effect can be used mostly without a problem, but often a better fitting representation of transitivity is given by the GWESP effect (30).

Modeling network dynamics is quite demanding in a number of respects, and the area is itself highly dynamic. More models will have to be developed in response to needs of empirical researchers. New procedures should be constructed that may be more efficient statistically and/or computationally, and implemented in software. Mathematical results should be proved for properties of the statistical procedures. A major open question is the robustness of estimation results for misspecification, and the role played in this respect by the goodness of fit of a model.

Acknowledgements

I am grateful to Christian Steglich for many years of collaboration on the development of this model. Thanks to Ruth Ripley, Krists Boitmanis, and Felix Schönenberger for programming the RSiena package; to Krists, once more, for making Figure 1. Thanks to Patrick West and the Medical Research Council / Medical Sociology Unit, University of Glasgow, for permission to use the data.

8 REFERENCES

References

Amati, V., F. Schönenberger, and T. Snijders (2015). Estimation of stochastic actor-oriented models for the evolution of networks by generalized method of moments. *Journal de la Société Française de Statistique 156*, 140–165.

Bather, J. (1989). Stochastic approximation: A generalisation of the Robbins-Monro procedure. In P. Mandl and M. Hušková (Eds.), *Proceedings of the fourth Prague symposium on asymptotic statistics*, pp. 13–27. Prague: Charles University.

Bollobaŝ, B. (1985). *Random Graphs*. London: Academic Press.

Brandes, U., G. Robins, A. McCranie, and S. Wasserman (2013). What is network science? *Network Science 1*, 1–15.

Burguete, J., A. Ronald Gallant, and G. Souza (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews 1*, 151–190.

Durante, D. and D. B. Dunson (2014). Nonparametric bayes dynamic modelling of relational data. *Biometrika 101*, 883–898.

Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *A Matematikai Kutató Intézet Kőzleményei 5*, 17–61.

Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, BC: Empirical Press.

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

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics 7*, 465–485.

Greenan, C. C. (2015). Diffusion of innovations in dynamic networks. *Journal of the Royal Statistical Society, Series A 178*, 147–166.

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

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

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

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

Holme, P. and J. Saramäki (2012). Temporal networks. *Physics Reports 519*, 97–125.

Huisman, M. E. and T. A. B. Snijders (2003). Statistical analysis of longitudinal network data with changing composition. *Sociological Methods & Research 32*, 253–287.

Huitsing, G., T. A. B. Snijders, M. A. Van Duijn, and R. Veenstra (2014). Victims, bullies, and their defenders: A longitudinal study of the coevolution of positive and negative networks. *Development and Psychopathology 26*, 645–659.

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

Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association 103*, 248–258.

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

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

Koskinen, J. and T. A. B. Snijders (2013). Longitudinal models. In D. Lusher, J. Koskinen, and G. Robins (Eds.), *Exponential Random Graph Models*, pp. 130–140. Cambridge: Cambridge University Press.

Koskinen, J. H., A. Caimo, and A. Lomi (2015). Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to Foreign Direct Investments. *Network Science 3*, 58–77.

Koskinen, J. H. and T. A. B. Snijders (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference 13*, 3930–3938.

Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society, Series B 76*, 29–46.

Kushner, H. J. and G. G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications* (Second ed.). New York: Springer.

Lerner, J., N. Indlekofer, B. Nick, and U. Brandes (2013). Conditional independence in dynamic networks. *Journal of Mathematical Psychology 57*(0), 275–283.

Lospinoso, J. A. (2012). *Statistical Models for Social Network Dynamics*. Ph. D. thesis, University of Oxford, U.K.

Louis, T. (1982). Finding observed information when using the EM algorithm. *Journal of the Royal Statistical Society, Series* **B** *44*, 226–233.

Lusher, D., J. Koskinen, and G. Robins (2013). *Exponential Random Graph Models*. Cambridge: Cambridge University Press.

Maddala, G. (1983). *Limited-dependent and Qualitative Variables in Econometrics* (3rd ed.). Cambridge: Cambridge University Press.

Newman, M. (2010). *Networks. An Introduction*. Oxford University Press.

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

Orchard, T. and M. Woodbury (1972). A missing information principle: theory and applications. In *Proceedings 6th Berkeley Symposium on Mathematical Statistic and Probability*, Volume 1, pp. 697–715. Berkeley: University of California Press.

Paul, S. and A. J. O`Malley (2013). Hierarchical longitudinal models of relationships in social networks. *Applied Statistics 62*, 705–722.

Pearson, M. A. and L. Michell (2000). Smoke rings: Social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention and Policy 7*(1), 21–37.

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

Rinaldo, A., S. E. Fienberg, and Y. Zhou (2009). On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, 446–484

Ripley, R. M., T. A. B. Snijders, Z. Bóda, A. Vörös, and P. Preciado (2016). Manual for Siena version 4.0. Technical report, Oxford: University of Oxford, Department of Statistics; Nuffield College.

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

Rubinstein, R. (1986). The score function approach for sensitivity analysis of computer simulation models. *Mathematics and Computers in Simulation 28*, 351–379.

Ruppert, D. (1988). Efficient estimation from a slowly convergent Robbins-Monro process. Technical Report 781, School of Operations Research and Industrial Engineering, Cornell University.

Ruppert, D. (1991). Stochastic approximation. In B. K. Ghosh and P. K. Sen (Eds.), *Handbook of Sequential Analysis*, pp. 503–529. New York: Marcel Dekker.

Schwabe, R. and H. Walk (1996). On a stochastic approximation procedure based on averaging. *Metrika 44*, 165–180.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association 106*, 1361–1370.

Schweinberger, M. and T. A. B. Snijders (2007). Markov models for digraph panel data: Monte Carlo-based derivative estimation. *Computational Statistics and Data Analysis 51*(9), 4465–4483.

Snijders, T. (2016a). Siena algorithms. Technical report, University of Groningen, University of Oxford,
`http://www.stats.ox.ac.uk/~snijders/siena/Siena_algorithms.pdf`.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology 31*, 361–395.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology 37*, 131–153.

Snijders, T. A. B. (2016b). The multiple flavours of multilevel issues for networks. In E. Lazega and T. A. B. Snijders (Eds.), *Multilevel Network Analysis for the Social Sciences; Theory, Methods and Applications*, pp. 15–46. Cham: Springer.

Snijders, T. A. B. and P. Doreian (2010). Introduction to the special issue on network dynamics. *Social Networks 32*, 1–3.

Snijders, T. A. B. and P. Doreian (2012). Introduction to the special issue on network dynamics (part 2). *Social Networks 34*, 289–290.

Snijders, T. A. B., J. H. Koskinen, and M. Schweinberger (2010). Maximum likelihood estimation for social network dynamics. *Annals of Applied Statistics 4*, 567–588.

Snijders, T. A. B., A. Lomi, and V. Torlò (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks 35*, 265–276.

Snijders, T. A. B., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology 36*, 99–153.

Snijders, T. A. B. and M. Pickup (2017). Stochastic actor-oriented models for network dynamics. In J. N. Victor, M. Lubell, and A. H. Montgomery (Eds.), *Oxford Handbook of Political Networks*, pp. xxx–xxx. Oxford: Oxford University Press.

Snijders, T. A. B. and C. E. G. Steglich (2017a). *Actor-based Models for Analyzing Network Dynamics*. Cambridge: Cambridge University Press.

Snijders, T. A. B. and C. E. G. Steglich (Eds.) (2017b). *Social Network Dynamics by Examples*. Cambridge: Cambridge University Press.

Snijders, T. A. B., C. E. G. Steglich, and M. Schweinberger (2007). Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud, and A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences*, pp. 41–71. Mahwah, NJ: Lawrence Erlbaum.

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

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

Tanner, M. and W. Wong (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association 82*, 528–550.

Tubaro, P., L. Mounier, E. Lazega, and T. Snijders (2017). Dynamics of advice-seeking networks among judges at the commercial court of paris. In T. Snijders and C. Steglich (Eds.), *Social Network Dynamics By Examples*, pp. xxx–xxx. Cambridge: Cambridge University Press.

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

Veenstra, R., J. K. Dijkstra, C. Steglich, and M. H. Van Zalk (2013). Network–behavior dynamics. *Journal of Research on Adolescence 23*(3), 399–412.

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

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of small-world' networks. *Nature 393*(6684), 440–442.

West, P. and H. Sweeting (1996). Background, rationale and design of the West of Scotland 11 to 16 study. *MRC Medical Sociology Unit Working Paper no. 52*.

Westveld, A. H. and P. D. Hoff (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *Annals of Applied Statistics 5*, 843–872.