Journal de la Société Française de Statistique

Preprint

Estimation of Stochastic actor-oriented models for the evolution of networks by generalized method of moments

Titre: Estimation des modèles stochastiques orienté par l'acteur pour l'évolution des réseaux par la méthode des moments généralisées

Viviana Amati[1], Felix Schönenberger[1] and Tom A. B. Snijders[2]

Abstract: The stochastic actor-oriented model (Snijders, Sociological Methodology, 2001) models the evolution of networks over time, given panel data in a fixed group of actors, where at each panel wave the network between these actors (a digraph structure) as well as attribute variables for these actors are observed. The parameters of this model usually are estimated by a stochastic approximation version of the method of moments. Statistics that correspond to the parameters in a natural way are used for fitting the model. Here we present an estimator based on the generalized method of moments, i.e., using more statistics than parameters, for minimizing the distance between observed statistics and their expected values. Again, the resulting equation is solved by stochastic approximation. Several algorithmic issues arise that have to be solved in order to obtain a stable procedure. For some examples we study the resulting gain in statistical efficiency.

Résumé : Le modèle stochastique orienté par l'acteur (Snijders, Sociological Methodology, 2001) modèle l'évolution temporelle des réseaux, étant donné un panel dans un ensemble fixe d'acteurs, où à chaque vague de panel le réseau entre ces acteurs (une structure de graphe orienté) ainsi que les attributs des acteurs sont observés. Les paramètres de ce modèle sont, d'habitude, estimés par une version d'approximation stochastique de la méthode des moments. Des statistiques qui correspondent aux paramètres d'une manière naturelle sont utilisés pour l'ajustement du modèle. Nous présentons ici un estimateur basé sur la méthode généralisée des moments, utilisant plus de statistiques que de paramètres, pour minimiser la distance entre les statistiques observées et leurs espérances mathématiques. Ici encore, l'équation résultante est résolue par approximation stochastique. Plusieurs questions algorithmiques surviennent qui doivent être résolues afin d'obtenir une procédure stable. Pour quelques exemples, nous étudions le gain résultant de l'efficience statistique.

Keywords: Stochastic actor-oriented models, social networks, generalized method of moments, stochastic approximation algorithm
Mots-clés : Modèles stochastiques orienté par l'acteur, réseaux sociaux, méthode des moments généralisées, algorithmes d'approximation stochastique

1. Introduction

Network analysis provides methods to describe and analyse one or more relations between actors. When dealing with networks, relational or dyadic variables —referring to the ties existing among

[1] Department of Computer Science, University of Konstanz.
E-mail: viviana.amati@uni-konstanz.de and E-mail: felix.schoenenberger@uni-konstanz.de
[2] Department of Sociology, University of Groningen; Nuffield College and Department of Statistics, University of Oxford.
E-mail: tom.snijders@nuffield.ox.ac.uk

 http://www.sfds.asso.fr/journal – ISSN: 2102-6238

actors— are collected in addition to the classical monadic variables describing actor characteristics. These variables together define the network data.

Various methods have been proposed in the past (Wasserman and Faust, 1994; Carrington et al., 2005; Scott and Carrington, 2011) to analyse network data, including methods of statistical inference. An important class is constituted by models that take networks as the dependent variable and aim to explain why actors are related in a certain way. In this article we consider such models for longitudinal data, collected by observing network data over time in a panel design.

Doreian and Stokman (1997) classified studies focusing on network longitudinal data into three categories: studies that predict attributes from structural information (Robins et al., 2001a,b; Doreian et al., 1984; Leenders, 2002); descriptive studies of network change (Freeman, 1984); and studies where the observed networks are perceived as the outcome of an unobserved process going on between several time points (Wasserman and Iacobucci, 1988; Wasserman, 1980a,b). Following the general framework suggested by Holland and Leinhardt (1977), the evolution process is often modelled as a continuous-time Markov chain.

Stochastic actor-oriented models (SAOMs) (Snijders, 1996, 2001; Snijders et al., 2010b) are models for longitudinal network data based on continuous-time Markov chains. One intuitively convenient interpretation of these models relies on the general idea that ties are both beneficial and costly for actors. Consequently, expensive ties are likely to be terminated, whereas beneficial ties are likely to be maintained or created by the actors. The decisions of actors of creating or dissolving links are modelled according to a random utility model, requiring the specification of a utility function depending on a set of explanatory variables and parameters. This is only an interpretation; the real assumption made is a Markov chain model with transition matrix defined by (1) and (2) below.

In SAOMs, the most often used procedure for parameter estimation is the method of moments (MoM), which estimates the parameters using one observed statistic for each estimated parameter. Due to the complexity of the model, the system of moment equations cannot be solved analytically, and the Robbins-Monro algorithm is used in order to approximate such a solution. This was elaborated in Snijders (2001).

Here we propose estimation by a procedure including more information, using more statistics than parameters. The result is that the principle of the MoM gives rise to an over-identified system of equations, so that the regular MoM cannot be applied. A suitable method then is the generalized method of moments (GMoM; Hansen, 1982), an estimation technique mainly used in econometrics, and potentially more efficient than the MoM. Like the regular MoM, the GMoM is based on the differences between the expected values of the statistics and their sample counterparts, but the GMoM involves the minimization of a quadratic function of these differences rather than setting all differences to zero. In order to solve such a minimization problem, the Robbins-Monro algorithm is modified and adapted.

The remainder of the paper is organized as follows. In Section 2 we present the model, the MoM estimator, and the stochastic algorithm used to approximate the solution of the system of moment equations. Then, in Section 3, we introduce the idea of using new statistics, we define the GMoM estimator and discuss the related computational problems. Section 4 and Section 5 report results of a small simulation study and of an empirical example, respectively. These results and further development are then discussed in Section 6.

2. Stochastic-actor oriented models

2.1. Model formulation

Stochastic actor-oriented models are a class of models for the analysis of network panel data. Let $N = \{1,2,\ldots,n\}$ be a set of actors over which a dichotomous directed relation $\mathscr{R}$ is defined and observed at $M+1$ time points $t_0,\ldots,t_M$. Let $X_{ij}(t_m)$ denote the tie indicator variable between actors i and j, taking value 1 if there is a tie from i to j at t_m and 0 otherwise. The matrix $X(t_m)$ with cells $X_{ij}(t_m)$ is the adjacency matrix of the network at t_m.

Monadic and dyadic variables are also part of the network panel data. While monadic variables describe the characteristics of the actors and are measured on nodes, dyadic variables describe the characteristics of dyads and therefore are defined over pairs of actors. We denote by V the matrix of monadic covariates, and by R the array of $n \times n$ matrices describing the dyadic covariates.

The network panel data, denoted by Y, are then defined as the time series of networks $\{X(t_0),\ldots,X(t_M)\}$, together with the matrix V and the array R. In the following, the observed values of these quantities will be denoted by small letters.

The time series represents the dependent variable of the model and it is assumed to be the outcome of a continuous-time Markov chain $\{X(t),\ t \in [t_0,t_M]\}$, observed only for $t = t_0,\ldots,t_M$. The Markovian property means that for any point in time, the next state of the network is probabilistically determined by the current state, and there are no additional effects of previous states. The continuous time assumption allows to consider the observed panel data as the result of a sequence of unobserved changes, happening between two consecutive observation time points t_{m-1} and t_m.

In order to model such a series of changes, it is assumed that at any single moment only one probabilistically selected actor has the opportunity to make a change. Such an opportunity for change is called a micro-step. The selected actor can decide either to change one of his/her outgoing ties or maintain the current configuration of the network. Consequently, only one tie variable can change at any time point. This assumption that no more than one tie variable can change at any given point was proposed by Holland and Leinhardt (1977), and yields a great simplification in modelling. It follows that each micro-step is defined by the moment at which it occurs, the actor having the opportunity to make a change, and the change itself. As a result, the network change process can be decomposed into two sub-processes: the change opportunity process modelling the time between two opportunities of change, and the change determination process modelling the particular choice made by the actor who gets the opportunity to modify one of his/her outgoing ties.

Due to the Markov assumption the waiting times between consecutive changes must be exponentially distributed. We assume that for each actor, the waiting times between consecutive opportunities for change by this actor are exponentially distributed with parameter ρ_m, for $t_{m-1} \leq t < t_m$. In this paper we assume that ρ_m is constant across actors; for models with variable waiting times see Snijders (2001) and Snijders et al. (2010a).

At any moment t, for all actors independent exponentially distributed waiting times are generated with parameter ρ_m, and the minimum of these is the time for the next micro-step (the others are disregarded). It follows that the waiting time until the next micro-step by any actor is exponentially distributed with rate parameter $n\rho_m$, and the conditional probability that it is actor

i who gets the opportunity to make a change between two consecutive time points, given that a micro-step is made at this time point by anybody, is $1/n$.

The particular change that takes place at a micro-step is regarded as the result of a choice made by the actor i. In this sense the model is actor-oriented. One way of presenting this is a rational choice model, where actors choose their best action depending on their preferences among all feasible actions according to a random utility model. Suppose that an individual faces a alternatives, so that $A = \{1, \ldots, a\}$ is the set of the feasible actions. According to random utility models a decision maker i chooses the action $a \in A$ providing the highest reward. Suppose actors are myopic, i.e., they consider only the utility that will be the immediate result of their choice, and the reward associated to an action a is modelled by a utility function $u_i(a)$

$$u_i(a) = f_i(a) + \varepsilon_i(a)$$

composed of two terms: $f_i(a)$ is the deterministic part of the utility capturing all the factors that are observed, while $\varepsilon_i(a)$ is a random term accounting for all the factors that are not included in $f_i(a)$, e.g., because the researcher did not collect them. Assuming that the random terms are independent and identically distributed as a standard type I extreme distribution (the standard Gumbel distribution), the probability that an individual i chooses action a is expressed in a closed form (Luce and Suppes, 1965; McFadden, 1974) by

$$P(i \text{ chooses } a) = \frac{e^{f_i(a)}}{\sum_{b \in A} e^{f_i(b)}} .$$

Suppose that actor i has the opportunity to make a change. The set of the feasible choices consists of either creating or terminating one outgoing tie or doing nothing (i.e., not making any changes). The deterministic part of the utility function associated to each possible choice is modelled by the *network evaluation* function

$$f_i(\beta, x, v, r) = \sum_{k=1}^{K} \beta_k s_{ik}(x, v, r)$$

which is defined as a linear combination of statistical parameters β_k and so-called effects $s_{ik}(x, v, r)$, which are functions of the current state x and covariates v, r, computed from the point of view of the actor i. It is assumed that actors have the knowledge required to compute the network evaluation function for each possible choice. We denote by $x(i \rightsquigarrow j)$ the network resulting from the decision of i to change the outgoing tie variable towards actor j, i.e., replace x_{ij} by $1 - x_{ij}$, and formally define $x(i \rightsquigarrow i) = x$ as the result of doing nothing. The conditional probability that the network changes to state $x(i \rightsquigarrow j)$, given that actor i makes a micro-step and the current is state $x(t) = x$, then is given by

$$P\big(x \text{ changes to } x(i \rightsquigarrow j)\big) = p_{ij}(x, \beta) = \frac{e^{f_i(\beta, x(i \rightsquigarrow j), v, r)}}{\sum_{h \in N} e^{f_i(\beta, x(i \rightsquigarrow h), v, r)}} . \quad (1)$$

It is further assumed that actors in a network follow the same decision rule and thus the deterministic part of the evaluation function has the same mathematical formulation for every

actor, allowing for differences between the actors' covariates. This implies that the parameters β_k are homogeneous with respect to actors. Furthermore, it is assumed that the decision rule defined by the evaluation function is constant over time. Consequently, the parameters β_k are also time-homogeneous.

The definitions given above imply that the model is a continuous-time Markov chain with transition matrix defined for $x \neq x'$ by

$$Q(x,x') = \begin{cases} \rho_m \, p_{ij}(x,\beta) & \text{if } x' = x(i \rightsquigarrow j) \\ 0 & \text{if } \sum_{h,k} \mid x_{hk} - x'_{hk} \mid \geq 2 \, . \end{cases} \tag{2}$$

2.2. Examples of effects

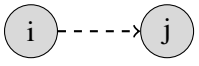
The effects $s_{ik}(x,v,r)$ are local patterns of ties and covariates used to explain the choices made by the actors, and are the focus of practical modelling for the SAOMs. In fact, each actor occupies a certain position in a network and being embedded in a particular pattern of ties (also called a local structure or configuration) might be more rewarding than being involved in a different one. According to the model's interpretation of actor-oriented optimization, when an actor gets the opportunity to make a change, he struggles for improving his position in the network, i.e., for being part of local structures particularly beneficial to him and for avoiding to be involved in those local configurations that are more cumbersome. The reward deriving from each local structure is measured by the parameter β_k. If β_k is high and positive (or low and negative), then the local configuration described by $s_{ik}(x(i \rightsquigarrow \cdot),v,r)$ is (or is not) appealing to i.

The effects are specified on the grounds of theories or previous knowledge and they may depend only on the network, or also on monadic or dyadic covariates. Here we present a list of commonly used effects, in particular, those occurring in the later examples (for a complete list available in the RSiena package see Ripley et al., 2015). Next to the mathematical formulation of each effect, we provide its graphical representation. A solid arrow denotes an existent tie; a dashed arrow represents an outgoing tie from i and indicates the tie that, when i has the opportunity to make a change, may be created if it does not yet exist or may be maintained if it does.

Some examples of network effects are outdegree, reciprocity, transitivity, and three-cycle. While outdegree and reciprocity are defined within dyads, transitivity and three-cycle involved a third actor and therefore model a more complex dependence assumption.

The *outdegree effect* measures the tendency for actors to create ties and is defined as the number of outgoing ties of actor i

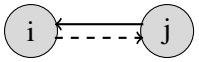
$$s_{i,\text{outd}}(x) = \sum_j x_{ij} \, .$$



Usually the corresponding parameter β_{outd} is negative since networks tend to be sparse (i.e., they have a low density). In the actor-oriented perspective this means that the cost of having a tie overcomes its benefit, unless there are special reasons for having this tie such as expressed by the following.

The *reciprocity effect* indicates the tendency to reciprocate incoming ties and is determined by the number of mutual dyads in which actor i is involved

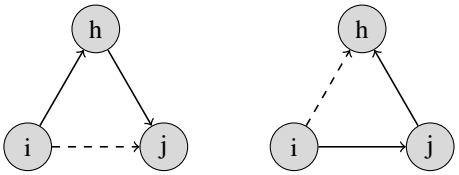
$$s_{i,\mathrm{rec}}(x) = \sum_j x_{ij}\, x_{ji} \ .$$



Since a lot of social relations show a tendency towards reciprocation, the associated parameter β_{rec} often assumes positive values. This suggests that even though the cost of creating or maintaining a tie is higher than its benefit, creating a tie that reciprocates an existing tie might lead to a positive reward for the actor i. This follows the general principle according to which ties should bring some benefits in order to overcome their costs and incentivize actors to create new ties.

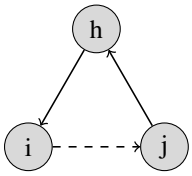
The *transitive triplets effect* models the tendency towards triadic closure, i.e., the tendency to relate to those others to whom actor i is related indirectly through an intermediary. This effect is defined as the number of transitive triplets in which an actor i is involved

$$s_{i,\mathrm{trans}}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{hj} \ .$$



The *three-cycle effect* is a sort of generalized reciprocity which involves three actors i, j and h. It expresses the idea that it may be not important if j directly reciprocates the tie from i, but it is important that j sends a tie to some actor h who has an outgoing tie towards i. This means that the tie between i and j is indirectly reciprocated through actor h. The effect is defined as the number of 3-cycles in the network:

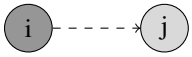
$$s_{i,\mathrm{cycle}}(x) = \sum_{j,h} x_{ij} x_{jh} x_{hi} \ .$$



Examples of effects depending on monadic covariates are the covariate-related activity, the covariate-related popularity, the same covariate, and the covariate-related similarity. These effects express the idea that creation and termination of ties in the network do not only depend on the position of actors in the network but also on their attributes. These effects, therefore, are defined as interactions between network variables and monadic covariates.

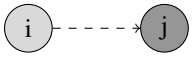
The *covariate-related activity* (or *covariate-ego*) effect indicates whether an attribute v affects the activity of actors in creating ties, i.e., whether actors with higher values of v tend to increase the number of their outgoing ties more rapidly than actors with lower values of v:

$$s_{i,\text{ego}}(x,v) = \sum_j x_{ij}\, v_i \ .$$



By contrast, the *covariate-related popularity* (or *covariate-alter*) effect measures whether actors tend to establish ties towards others showing a particular value of v, i.e., whether the indegree of actors with higher values of v increases more rapidly than the indegree of actors with lower values of v:

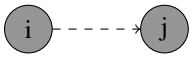
$$s_{i,\text{alter}}(x,v) = \sum_j x_{ij}\, v_j \ .$$



The *same covariate* and the *covariate-related similarity* model homophily. The former is specifically defined for categorical covariates and investigates whether ties between actors with exactly the same value of v are more likely to exist. The latter is defined for continuous (or at least ordinal) covariates and measures the tendency of actors to establish ties with others who are similar to themselves:

$$s_{i,\text{same}}(x,v) = \sum_j x_{ij}\, \mathbb{1}\{v_i = v_j\}$$

$$s_{i,\text{sim}}(x,v) = \sum_j x_{ij} \left(1 - \frac{|v_i - v_j|}{\max_{ij} |v_i - v_j|} \right) \ .$$



The above is the description of the simplest formulation of the SAOM. For the sake of completeness, it is worth mentioning that there are several extensions that render the SAOM more flexible and applicable in a large variety of situations. For instance, in the evaluation function the gain of creating a tie is the opposite of the loss of terminating a tie. In reality creating or terminating a tie might be evaluated in a different way by the actors. To model such a difference the *creation* and the *endowment* functions might be used next to the evaluation function (Snijders et al., 2010b). A further development of the SAOM allows the simultaneous modelling of the evolution of networks and behaviours (Steglich et al., 2010). The main idea is that actors in a network may adjust their behaviours to that of the people they are connected to and therefore actors can change their behaviour next to their ties. Other extensions are described in Lospinoso et al. (2011); Koskinen and Edling (2012); Snijders et al. (2013); Greenan (2015).

2.3. Estimating the parameter of SAOMs

2.3.1. The method of moments estimator

Let $\theta = (\rho_1, \cdots, \rho_M, \beta_1, \cdots, \beta_K)$ be the $P = (M+K)$-dimensional vector of parameters corresponding to a particular specification of the SAOM. Estimating θ allows inferences about whether

certain mechanisms drive the network evolution. Estimation is performed conditional on the first observation of the network, i.e., $x(t_0)$ is not modelled but considered as given. Therefore, no model assumptions concerning the mechanisms that might have generated $x(t_0)$ need to be formulated. Moreover, the estimated parameters refer exclusively to the evolution of the network in the time interval $[t_0, t_M]$.

Several estimation procedures might be used to estimate θ. Snijders (2001) proposed to use the method of moments (MoM), a Bayesian procedure was presented by Koskinen and Snijders (2007), and Snijders et al. (2010a) proposed maximum likelihood estimation procedures (MLE). Since the likelihood function of the SAOM cannot be computed explicitly, Bayesian and ML estimations are very time-consuming and the MoM is the most used estimation method in practice.

The logic of this method is quite straightforward and is based on the idea that the expected values of statistics depend on the parameters of the distribution. Specifically, let us assume that X is a random variable whose probability distribution depends on a parameter $\theta \in \mathbb{R}^P$. We are interested in estimating θ. The MoM requires a P-dimensional vector of statistics $S = (S_1, \ldots, S_P)$, one for each parameter θ_p $(1 < p < P)$, and estimates θ by the value $\widehat{\theta}_{MoM}$ for which the expected values equal their sample counterparts,

$$E_\theta[S] = s \ ,$$

where $s = (s_1, \ldots, s_P)$ is the observed value of S.

The statistics S_p are usually chosen using a formal method such as the reduction by sufficiency. When applying the MoM to the estimation of the parameter of the SAOM, however, this method cannot be invoked because of the complexity of the model. Therefore, the statistics are chosen in a heuristic way. The requirement is that they are sensitive to the parameter θ in the sense that higher values of the parameter θ_p will lead to higher expected values of the statistic S_p, i.e.,

$$\frac{\partial}{\partial \theta_p} E_\theta [S_p] > 0 \ .$$

Snijders (2001) proposed to use the following statistics.

The rate parameter models the frequency at which actors get an opportunity to make a change. When this frequency is higher, the total observed number of changes between two consecutive observations will tend to be higher. Therefore, a relevant statistic for ρ_m is the Hamming distance between $X(t_m)$ and the previous observation $X(t_{m-1})$,

$$S_{\rho_m} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left| X_{ij}(t_m) - X_{ij}(t_{m-1}) \right| \ .$$

A further motivation for the use of S_{ρ_m} is that it is a sufficient statistic for ρ_m when ties are changed totally at random by the actors in the network (i.e., when $\beta = 0$).

For parameters such as β_k, which are assumed to be constant for all M periods, an adaptation of this prescription is used. Functions $S_k(x)$ are used (with an implicit possible dependence also on V and R), with observed values $S_k\big(x(t_m)\big)$ for $m = 0, \ldots, M$, and the MoM equation for these S_k is

$$\sum_{m=1}^{M} E_\theta \left[S_k\big(X(t_m)\big) \mid X(t_{m-1}) = x(t_{m-1}) \right] = \sum_{m=1}^{M} S_k\big(x(t_m)\big) \ .$$

For any given k, higher values of β_k will presumably lead to higher expected values of the corresponding function $s_{ik}\big(X(t)\big)$ for $t > t_0$ and all $i \in N$, and a sensitive statistic for the parameter β_k is therefore

$$S_k\big(X(t_m)\big) = \sum_{i=1}^{n} s_{ik}\big(X_{ij}(t_m)\big) \ .$$

Summarizing, the estimator $\widehat{\theta}_{MoM}$ for θ is the solution of the system of $M+K$ moment equations:

$$\begin{cases} E_\theta\left[S_{\rho_m} \mid X(t_{m-1}) = x(t_{m-1})\right] = \ s_{\rho_m} & m = 1,\ldots,M \ , \\ \\ \sum\limits_{m=1}^{M} E_\theta\left[S_k\big(X(t_m)\big) \mid X(t_{m-1}) = x(t_{m-1})\right] = \ \sum\limits_{m=1}^{M} s_k\big(x(t_m)\big) & k = 1,\ldots,K \ ; \end{cases} \tag{3}$$

where s_{ρ_m} is the observed value of S_{ρ_m}.

In the following, we will use the more compact notation

$$E_\theta[S] - s = 0 \tag{4}$$

Application of the delta method (Lehmann, 1999, p.315) allows the approximation of the covariance matrix of $\widehat{\theta}_{MoM}$ according to the following formula (Bowman and Shenton, 1985):

$$\Sigma_{\widehat{\theta}_{MoM}} = D_\theta^{-1}\,\Sigma_S\,(D_\theta^{-1})' \tag{5}$$

where

$$D_\theta = \left(\frac{\partial}{\partial\theta} E_\theta[S]\right) \tag{6}$$

is the matrix of partial derivatives of the expected values of the statistics and

$$\Sigma_S = \mathrm{cov}_\theta(\mathrm{S})$$

is the covariance matrix of the statistics.

2.3.2. Stochastic approximation

Analytical and usual numerical procedures cannot be applied to solve the system of moment equations in (4), since the expected values of the statistics cannot be computed explicitly except for some trivial cases (Snijders and van Duijn, 1997). For this reason, a stochastic approximation algorithm is used.

The algorithm is an iterative procedure consisting of two main steps: *i)* simulations of network evolution for a given value θ_i of θ; *ii)* update of the value of θ according to the multivariate version of the Robbins-Monro step (Robbins and Monro, 1951)

$$\theta_{i+1} = \theta_i - a_i \widetilde{D}_0^{-1}(S_i - s)$$

where S_i is the value of the statistics generated in step (i) for $\theta = \theta_i$; $\widetilde{D}_0^{-1}$ is the diagonal matrix (Ruppert, 1988; Polyak, 1990) of the matrix of partial derivatives in (6), measuring the sensitivity of the expected value of S_p to perturbation of the parameter θ_p; a_i is a sequence of positive numbers slowly converging to 0 and controlling for the convergence of the algorithm. The final estimate is a tail average of the values θ_i (Ruppert, 1988; Polyak, 1990).

The algorithm as implemented (Ripley et al., 2015) consists of three phases and can be sketched as follows (for more details see Snijders (2001)). Let θ_0 be an initial value of θ. In the first phase a small number n_1 of simulations is used to obtain a rough estimate of (6). Either the finite difference method or the score function method (Schweinberger and Snijders, 2007) can be used for estimating the derivatives. At the end of the first phase the initial parameter value is updated using the Newton-Raphson step

$$\theta_1 = \theta_0 - a_0 D_0^{-1}(\overline{S} - s)$$

where $\overline{S}$ is the mean of the simulated values of the statistics and thus a Monte Carlo approximation of the expected values of the statistics.

The second phase determines the estimate of θ via a small number (usually four) of sub-phases. Each sub-phase h consists of n_{2h} iterations of the steps *i)* and *ii)* as defined previously. The number of iterations n_{2h} is bounded by a minimum value $\underline{n}_{2h}$ and a maximum value $\overline{n}_{2h}$. Each sub-phase ends when either the upper bound $\overline{n}_{2h}$ is reached; or the first order autocorrelation of the sequence S_i is negative for all i (Pflug, 1990), and also the lower bound $\underline{n}_{2h}$ was reached. At the end of each sub-phase the value of θ is estimated using the mean for this subphase of the sequence θ_i.

In the third phase the network evolution is simulated a large number n_3 of times using the value of θ determined at the end of the second phase to estimate the covariance matrix of the estimator $\Sigma_{\widehat{\theta}_{MoM}}$; and to check the convergence of the algorithm. According to (5), the matrix of partial derivatives D and the covariance matrix Σ_S should be estimated to compute $\Sigma_{\widehat{\theta}_{MoM}}$. The former is estimated as in phase 1, while the latter using its direct Monte Carlo approximation.

The check of convergence of the algorithm is based on the deviations between simulated values of the statistics and their observed values, i.e., the observed left hand side of (4). These deviations should be equal to 0 because of the logic of the MoM. However, since the algorithm only approximates the solution of the system of moment equations, the deviations are small but not exactly equal to 0 when the process has properly converged. Consequently, the convergence of the algorithm is tested using the ratio between the mean simulated values of the statistics and their standard deviations. As a rule of thumb, when all these ratios are in absolute value lower than 0.1 (0.15) the convergence is excellent (reasonable).

The Robbins-Monro algorithm for the parameter estimation of the SAOM is implemented in the R package `RSiena` (Ripley et al., 2015).

3. The generalized method of moments

3.1. New statistics

The definition of the statistics S_k used in (3) for the MoM depends only on the network at time t_m. However, the network observed at time t_{m-1} can provide valuable information on how the configurations observed at time t_m originated.

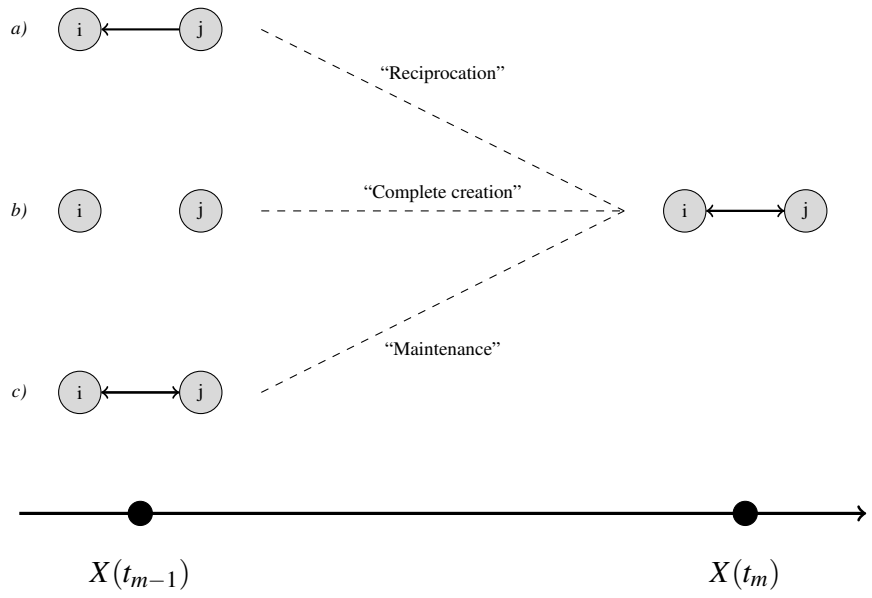


FIGURE 1. *Different situations from which a reciprocal dyad can originate.*

For instance, let us assume that actors i and j constitute a reciprocal dyad at time t_m (Figure 1). The mutual dyad noticed at t_m may have originated from different configurations existing between i and j at time t_{m-1}. In particular, three situations are possible. *a)* At time t_{m-1} an asymmetric dyad was observed. Then, actor i decided to reciprocate the tie from j. *b)* At time t_{m-1} a null dyad existed. Then, it may be that i sent a tie towards j and, later on, j decided to reciprocate the relation. Vice versa, it may also be that j was first the sender and then the receiver of the two ties. *c)* A mutual dyad was already observed at time t_{m-1} and both ties between actors i and j remained unchanged. It follows that the mutual dyad observed at t_m may be the result of *"reciprocation"* (a), *"complete creation"* (b) or *"maintenance"* (c).

These three situations contribute in a different way to the reciprocity effect. Although the real sequence of tie changes is unknown (the process is observed only at discrete time points), situation *a* provides the reciprocity effect with more evidence than situations *b* and *c*. Therefore, it seems reasonable to include the information deriving from the starting configuration of a dyad in the estimation procedure of the reciprocity parameter.

A new statistic S^*_{rec} depending on both networks observed at t_{m-1} and t_m is defined to this end. The statistic counts the number of mutual dyads originated according to a mechanism of reciprocation

$$S^*_{\text{rec}}(X(t_m), X(t_{m-1})) = \sum_{m=1}^{M} \sum_{i,j=1}^{n} (1 - X_{ij}(t_{m-1}))\, X_{ji}(t_{m-1})\, X_{ij}(t_m)\, X_{ji}(t_m)$$

A similar argument applies to transitive triplets. Let us assume that we observed a transitive triplet at time t_m among the actors i, j and h (right side of Figure 2). Taking into account the configuration of ties existing between i, j and h at time t_{m-1}, there are eight different situations from which the

observed transitive triplet could have originated (left side of Figure 2). Due to the actor-oriented perspective of the model and the evidence provided to the mechanism of transitivity, the most interesting situations are *a* and *b*. In the former, the transitive triplet observed at t_m is the result of the closure of a two-path between i and j. In the latter the transitive triplet derives from the agreement between i and h on j.

As a result, two new statistics are defined for the estimation of the parameter related to the transitive triplets. One statistic refers to the closure mechanism

$$S^*_{\text{clos}}(X(t_m), X(t_{m-1})) = \sum_{m=1}^{M} \sum_{i,j=1}^{n} (1 - X_{ij}(t_{m-1}))\, X_{ih}(t_{m-1})\, X_{hj}(t_{m-1})\, X_{ij}(t_m)\, X_{ih}(t_m)\, X_{hj}(t_m)$$

and the other to the agreement mechanism

$$S^*_{\text{agr}}(X(t_m), X(t_{m-1})) = \sum_{m=1}^{M} \sum_{i,j=1}^{n} X_{ij}(t_{m-1})\, (1 - X_{ih}(t_{m-1}))\, X_{hj}(t_{m-1})\, X_{ij}(t_m)\, X_{ih}(t_m)\, X_{hj}(t_m)\ .$$

Similar arguments apply to all the effects that can be included in the evaluation function. Thus, we can specify new statistics for each effect according to the information provided by the network observed at t_{m-1}.

At this point, it is important to make the following remark. In the SAOM every statistic S_k corresponds to an effect in the evaluation function and thus to a parameter β_k. The new statistics are not part of the evaluation function, but they are only used during estimation. As a result, there is no parameter uniquely associated to them, and one should think about them only as additional information to improve the parameter estimation.

The new statistics can be included in the estimation process using the principle of the MoM. This means that further MoM equations based on the new statics are added to the system in (4). This operation results in an over-identified system of equations

$$\begin{cases} E_\theta\left[S_{\rho_m} \mid X(t_{m-1}) = x(t_{m-1})\right] = s_{\rho_m} & m = 1, \ldots, M \\ \sum_{m=1}^{M} E_\theta\left[S_k\left(X(t_m)\right) \mid X(t_{m-1}) = x(t_{m-1})\right] = \sum_{m=1}^{M} s_k\left(x(t_m)\right) & k = 1, \ldots, K \\ \sum_{m=1}^{M} E_\theta\left[S^*_l\left(X(t_m), X(t_{m-1})\right) \mid X(t_{m-1}) = x(t_{m-1})\right] = \sum_{m=1}^{M} s^*_l\left(x(t_m), x(t_{m-1})\right) & l = 1, \ldots, L \end{cases} \tag{7}$$

defined by more equations (Q=M+K+L) than unknowns (P=M+K); these equations take the place of equations such as (3). In the following, we use the more compact notation

$$E_\theta[S^*] = s^*$$

where S^* is the Q-dimensional vector comprised of both the new and the old statistics.

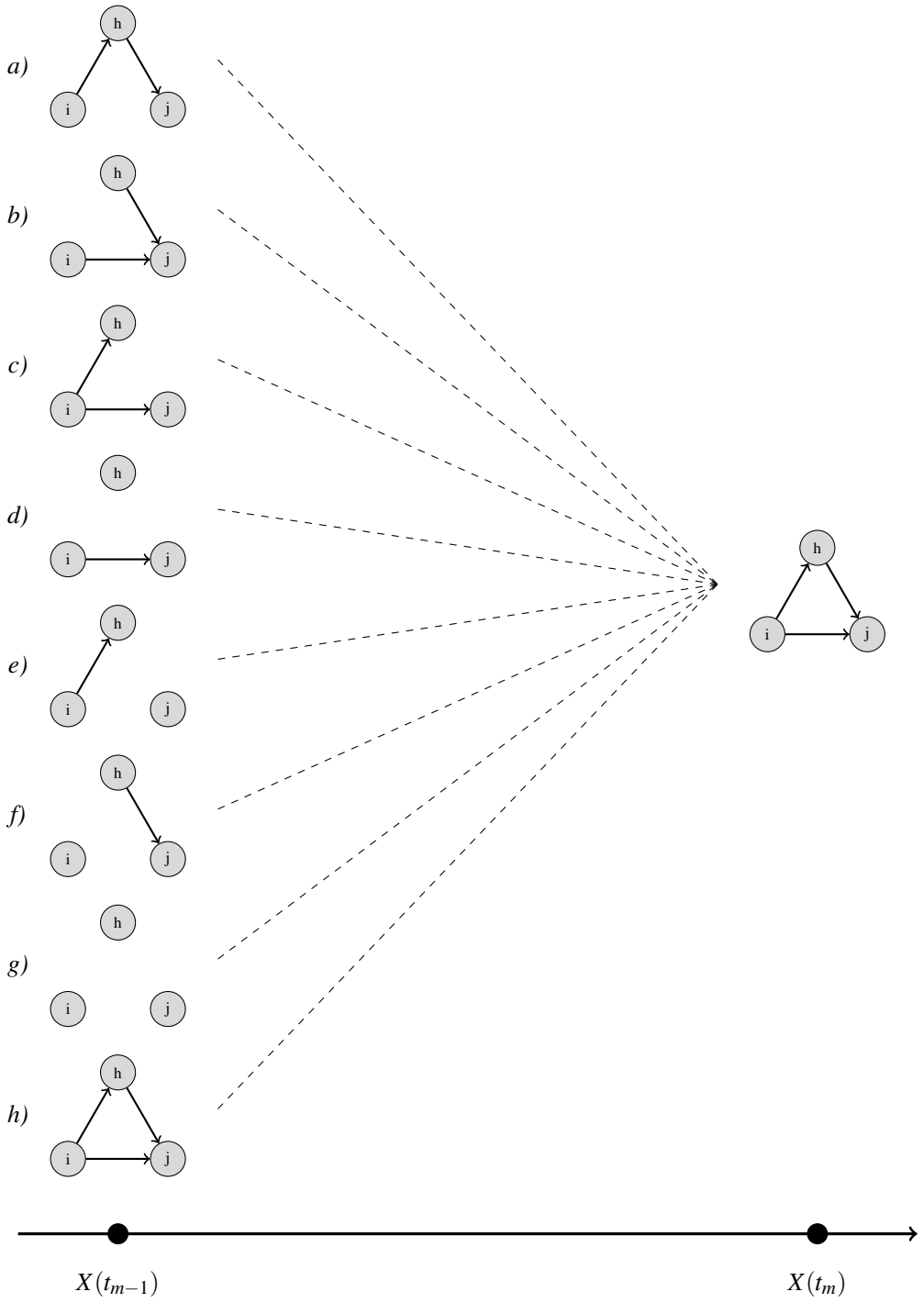


FIGURE 2. *Different situations from which a transitive triplet can originate.*

From mathematical theory it follows that such a system has no solution unless there are linearly dependent equations that can be neglected so that the number of equations corresponds to the number of unknowns. If this does not happen, the usual methods for solving a system do not work and approximated solutions are the only alternative.

In system (7), the equations are not linearly dependent and therefore an approximated solution must be calculated. When there are more moment conditions than parameters the regular MoM cannot be applied, but its generalization can be used. This generalization is the generalized method of moments.

3.2. The generalized method of moments estimator

The generalized method of moments (GMoM) refers to a class of estimators introduced into the econometric literature by Hansen (1982) and developed contemporaneously and independently by Burguete et al. (1982). Since then, the GMoM has become very popular both in applications and in theoretical analysis for some main reasons. First, the GMoM supplies a way to estimate parameters of partially specified models where maximum likelihood estimation is not feasible. Second, the GMoM can be applied under very weak conditions on the process that may have generated the observed data, thereby allowing wide application in cross-sectional, panel, and time-series data. Third, the GMoM offers computational advantages. Even in fully specified models, the MLE can be cumbersome to compute, whereas the GMoM provides a method of obtaining consistent and asymptotically normal estimators of parameters in non-linear dynamic models. Finally, the GMoM is a very general framework for statistical analysis, since it subsumes least squares (LS), instrumental variables (IV) and maximum likelihood (ML) estimation techniques as special cases.

Let X be a random variable whose probability distribution depends on a P-dimensional parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_P)$. Let $S^* = (S_1, \ldots, S_Q)$ be a Q-dimensional vector ($Q \geq P$) of statistics relevant to the estimation of θ. Like the MoM, the GMoM is based on the differences between the expected values of the statistics and their sample counterparts, but the GMoM involves the minimization of a quadratic function of these differences rather than setting all differences to zero. Let

$$f(\theta, S^*) = (E_\theta[S^*] - s^*)' \, W \, (E_\theta[S^*] - s^*)$$

be the quadratic function of these differences and W a (symmetric) semi-positive definite $Q \times Q$ matrix. The restriction on W ensures that the quadratic form $f(\theta, S^*)$ is a meaningful measure of distance between the expected values of the statistics and their sample counterparts. It can be proved (Hansen, 1982; Hall, 2005) that an optimal choice for W is the concentration matrix, i.e.,

$$W = [\mathrm{cov}(\mathrm{S}^*)]^{-1} \ . \tag{8}$$

The GMoM estimator $\widehat{\theta}_{GMoM}$ for θ is the value which minimizes $f(\theta, S^*)$, i.e.,

$$\widehat{\theta}_{GMoM} = \arg\min_{\theta \in \Theta} f(\theta, S^*) \ . \tag{9}$$

Since $f(\theta, S^*)$ depends on the expected values of the statistics, analytical and numerical procedure cannot be used when applying the GMoM to the estimation of the parameter of a SAOM. Moreover, compared to the MoM estimator, an extra problem arises: the determination of W based on one

single observation of the network evolution. To overcome these difficulties a modified version of the Robbins-Monro algorithm introduced in Section 2.2 will be presented.

The solution of the minimization problem in (9) corresponds to the root of

$$\frac{\partial}{\partial \theta} f(\theta, S^*) = \Gamma' W \left(E_\theta[S^*] - s^*\right) = 0 \,, \tag{10}$$

where

$$\Gamma = \frac{\partial}{\partial \theta} \left(E_\theta[S^*] - s^*\right) \tag{11}$$

is the $Q \times P$ matrix of first order derivatives of the statistics involved in the estimation process. Let B an approximation of $\Gamma' W$, then $\widehat{\theta}_{GMoM}$ can be approximated by solving

$$B \left(E_\theta[S^*] - s^*\right) = 0 \,. \tag{12}$$

The matrix B defines the normal equations for the solution of the optimizing problem in (12) and can be thought of a matrix of weights attaching differential importance to each statistic in the estimation process. The relevance of a statistic depends on both the sensitivity of the statistic to perturbation of the parameter θ (through Γ) and the variance of the statistic (through W). The higher the sensitivity and the lower the variance, the higher the importance. The next section describes how to solve (12) using the Robbins-Monro algorithm.

3.3. Stochastic approximation

Above we sketched the implementation of the Robbins-Monro algorithm used for MoM estimation of parameters in the SAOM. Some changes in this algorithm are required in order to account for the matrix $B = \Gamma' W$. In this "modified" version of the Robbins-Monro algorithm the two components Γ and W are estimated at the end of phase 1. Then, phase 2 and phase 3 are executed like in the "regular" Robbins-Monro algorithm, aiming to solve (12) instead of (4). In the following we refer to the "regular" Robbins-Monro algorithm as the "MoM algorithm", and to its modified version as the "GMoM algorithm".

In the first phase of the MoM algorithm a small number n_1 of simulations is used to roughly estimate the matrix of partial derivatives D and to update the initial value of θ. In a similar way, Γ can be approximated using either the finite difference method or the score function method.

Phase 1 should also be adapted in order to include the estimation of the matrix of weights W. Given the n_1 simulations, W is estimated computing the inverse of the covariance matrix of the simulated statistics. The network change process is decomposed into the change opportunity process, modelled by the rate functions, and the change determination process, modelled by the evaluation function, as described above. The parameters for these two subprocesses are distinct, and estimated using distinct statistics. Therefore, in the 2×2 block structure defined by the statistics for the rate parameters and those for the evaluation function parameters, the off-diagonal blocks in the matrix W are set to 0. This is meant to stabilize estimation of W.

Denoting by Γ_0' and W_0 these approximations, the matrix B is estimated by row-normalization applied to the matrix $B_0 = \Gamma_0' W_0$. The row-normalization (i.e., dividing all rows by their Euclidean length) is applied to improve the convergence of the algorithm and control for high values in the matrix B_0. Note that dividing the rows of B by a constant does not change equation (12).

At the end of phase 1, the parameter θ_0 then is updated using the Newton-Raphson step, which is in turn modified as follows:

$$\theta_1 = \theta_0 - D_0^{-1} B_0 (\overline{S}^* - s^*) \ ,$$

where $D_0^{-1} = B_0 \Gamma_0$. The second phase remains the same. Only the Robbins-Monro step is modified to account for B, and is given by

$$\theta_{i+1} = \theta_i - a_i \widetilde{D}_0^{-1} B_0 (S_i^* - s^*) \ ,$$

where $\widetilde{D}_0^{-1}$ is a diagonal matrix.

The third phase of the algorithm estimates the covariance matrix of the GMoM estimator and checks for the convergence of the algorithm. Application of the delta method allows the computation of the covariance matrix of $\widehat{\theta}_{GMoM}$ according to the following formula (Hall, 2005):

$$\Sigma_{\widehat{\theta}_{GMoM}} \approx (B_0 \Gamma)^{-1} (B_0 \Sigma_{S^*} B_0') \left((B_0 \Gamma)^{-1} \right)' \ ,$$

where $\Sigma_{S^*} = \mathrm{cov}_\theta(\mathrm{S}^*)$. Following phase 1 of the GMoM algorithm and using the value of θ computed during phase 2, a high number n_3 of simulations is performed to re-estimate B, Γ and Σ_{S^*}, and consequently $\Sigma_{\widehat{\theta}_{GMoM}}$.

The check of convergence of the algorithm is again based on the deviations between simulated values of the statistics and their observed values. Since the approximation of the GMoM estimator is the solution of equation (12), the deviations weighted with the matrix B should (ideally) be equal to 0. Consequently, the convergence of the algorithm is tested using the ratio between the simulated values of the statistics $B_0 S^*$ and their standard deviations. Just like for the MoM algorithm, when these ratios are in absolute value lower than 0.1 (0.15) the convergence is excellent (reasonable).

It must be noted that the number of simulations n_1 and n_3 of the GMoM algorithm should be higher than those of the MoM algorithm since more quantities need to be estimated. In fact, the matrix of first order derivatives Γ is larger than D and the GMoM requires the additional estimation of the matrix of weights W. Also, estimating B with good precision for the GMoM is more important than estimating $\widetilde{D}_0$ with good precision for the MoM. The Robbins-Monro algorithm for the GMoM estimator will be implemented in the R package `RSiena` (http://r-forge.r-project.org/R/?group_id=461).

4. Simulation examples

Several simulations were conducted to investigate the relative efficiency of the GMoM estimator with respect to the MoM estimator. A nested simulation design was used. Three repeated observations of a network consisting of 50 actors were simulated 500 times according to a particular model specification. Estimates by both methods were calculated twice for each simulated sequence of networks. When estimating the standard errors of the two estimators, such a design has the advantage of accounting for the variation in both the simulated network evolution and the stochastic algorithm.

The network evolution was simulated using a model specified by the outdegree, reciprocity, transitive triplets, and 3-cycle effects. Since actors were characterized by a binary attribute

taking values 0 and 1, the covariate-related activity, the covariate-related popularity and the same covariate effects were also included in the model as explanatory variables for the network evolution. Consequently, the evaluation function of the model is

$$\begin{aligned} f_i(\beta, x(i \rightsquigarrow j), v) = {} & \beta_{\text{outd}} \sum_j x_{ij} + \beta_{\text{rec}} \sum_j x_{ij} x_{ji} \\ & + \beta_{\text{trans}} \sum_{j,h} x_{ij} x_{ih} x_{hj} + \beta_{\text{cycle}} \sum_{j,h} x_{ij} x_{jh} x_{hi} \\ & + \beta_{\text{ego}} \sum_j x_{ij} v_i + \beta_{\text{alter}} \sum_j x_{ij} v_j \\ & + \beta_{\text{same}} \sum_j x_{ij} \mathbb{1}\{v_i = v_j\} \end{aligned} \tag{13}$$

and

$$\theta = \left(\rho_1, \rho_2, \beta_{\text{outd}}, \beta_{\text{rec}}, \beta_{\text{trans}}, \beta_{\text{cycle}}, \beta_{\text{ego}}, \beta_{\text{alter}}, \beta_{\text{same}}\right) .$$

The choice of the parameter values of the evaluation function is based on empirical results deriving from the estimation of SAOMs in friendship networks. Therefore, the outcome of the simulation study cannot be generalized to different type of relations. The value of the parameter related to the density of the network was tuned so that the average degrees of the actors are in a reasonable desired range, i.e., between 3 and 6.

The choice of the rate parameters was based on the consideration that the new statistics may add more information when observations are not too far apart. For instance, if the observed outcome is a transitive triplet and we would like to distinguish between closure and agreement, we need to know in which order ties were created. When observations are too far apart, the data do not provide such an information and situations like *d-g* in Figure 2 are more likely to be observed than the situations *a* and *b*. To empirically verify this expectation, different rate parameters were combined with the same parameter values of the evaluation function.

The results of the simulations are presented in Tables 1 and 2. The second column of the tables shows the value of the parameter θ used to simulate the network evolution. To estimate the parameters of the SAOM using the GMoM and to compare the efficiency of the two estimation methods only the new statistics S^*_{rec} and S^*_{clos} and S^*_{agr} were considered. The number of iterations in phase 1 was equal to 300 and 35 for the GMoM and the MoM algorithms, respectively. For the MoM, this is the default used for this value of P in RSiena (Ripley et al., 2015); for the GMoM, experience showed that for the determination of B_0 a higher number is needed. The number of iterations in phase 3 was set to 4,000.

Table 1 reports the results of the simulations from three SAOMs having the same parameter values for the evaluation function but different rates. While the algorithm always converged for the MoM, its modified version for the GMoM did not always reach convergence. To define a proper estimator, we defined the GMoM estimator in the case of non-convergence as the MoM estimator. For the results in Table 1 non-convergence of the GMoM occurred in 5, 6 and 17 simulations, respectively. The GMoM and the MoM estimators yield similar estimates. The efficiency of the MoM and the GMoM was compared using a Wilcoxon signed rank test. The last column of Table 1

	θ	$\widehat{\theta}_{GMoM}$	$RMSE(\widehat{\theta}_{GMoM})$	$\widehat{\theta}_{MoM}$	$RMSE(\widehat{\theta}_{MoM})$	
rate t_0–t_1	3.000	2.786	0.348	2.784	0.329	†
rate t_1–t_2	4.000	3.985	0.521	3.987	0.458	†
density	–2.000	–1.980	0.091	–1.980	0.096	*
reciprocity	2.000	1.925	0.177	1.919	0.182	*
transitivity	0.400	0.381	0.088	0.382	0.107	*
3-cycle	–0.120	–0.120	0.174	–0.121	0.189	*
ego v	–0.200	–0.152	0.151	–0.153	0.150	
same v	0.600	0.537	0.145	0.537	0.141	†
alter v	0.400	0.363	0.146	0.360	0.144	
rate t_0–t_1	4.500	4.475	0.573	4.494	0.504	†
rate t_1–t_2	5.500	5.481	0.767	5.536	0.612	†
density	–2.000	–2.003	0.082	–2.003	0.085	*
reciprocity	2.000	2.015	0.150	2.008	0.154	*
transitivity	0.400	0.388	0.062	0.388	0.079	*
3-cycle	–0.120	–0.125	0.134	–0.126	0.148	*
ego v	–0.200	–0.206	0.137	–0.205	0.133	†
same v	0.600	0.621	0.126	0.619	0.120	†
alter v	0.400	0.409	0.127	0.406	0.126	
rate t_0–t_1	6.500	6.343	0.915	6.412	0.720	†
rate t_1–t_2	7.500	7.424	1.040	7.574	0.771	†
density	–2.000	–1.992	0.081	–1.990	0.080	
reciprocity	2.000	2.006	0.138	1.998	0.138	
transitivity	0.400	0.384	0.048	0.380	0.063	*
3-cycle	–0.120	–0.110	0.114	–0.113	0.123	*
ego v	–0.200	–0.189	0.126	–0.190	0.122	†
same v	0.600	0.617	0.113	0.611	0.106	†
alter v	0.400	0.413	0.113	0.408	0.112	

TABLE 1. *Simulation results, 3 waves for 50 actors: true value of the parameter (θ), mean ($\widehat{\theta}$), and root mean squared errors (RMSE) of the generalized method of moments (GMoM) and the method of moments (MoM). SAOMs with the same parameter values for the effects in the evaluation function but different rates. An asterisk (*) in the last column denotes the effects for which the RMSE is lower for the GMoM than for the MoM; a dagger (†) denotes the effects for which the RMSE is higher for the GMoM than for the MoM (Wilcoxon signed rank test, p-value < 0.01).*

indicates that for the first combination of parameters, the GMOM has significantly lower MSE than the MoM for the parameters related to density, reciprocity, transitive triplets, and 3-cycles; and significantly higher RMSE for the parameters related to the rates and the same-covariate effect. Similar patterns were observed when higher rates of changes were used to simulate the network evolution: performance of the GMoM for the effects of network structure was somewhat better, for the effects of rates and of covariates somewhat worse compared to the MoM. Therefore, the small simulation study did not provide evidence for a higher gain in efficiency for the GMoM when network observations are closer in time. The ratio between the estimated mean squared errors ranges between 0.97 and 1 for the reciprocity and between 0.76 and 0.82 for the transitive triplets parameter, the two parameters for which additional statistics were used in the GMoM. However, for some other parameters the GMoM resulted in higher standard errors, in particular for the rate parameters, where this ratio went up to 1.35. This was unexpected, and will be examined in future work.

The efficiency of the GMoM and the MoM estimators was also compared for some other

	θ	$\widehat{\theta}_{GMoM}$	$RMSE(\widehat{\theta}_{GMoM})$	$\widehat{\theta}_{MoM}$	$RMSE(\widehat{\theta}_{MoM})$	
rate t_0–t_1	3.000	2.971	0.327	2.973	0.303	†
rate t_1–t_2	4.000	4.076	0.436	4.089	0.372	†
density	–1.500	–1.576	0.091	–1.575	0.096	*
reciprocity	1.500	1.427	0.154	1.426	0.160	*
transitivity	0.400	0.411	0.057	0.410	0.063	*
3-cycle	–0.120	–0.113	0.097	–0.112	0.104	*
ego v	–0.200	–0.148	0.138	–0.149	0.137	
same v	0.600	0.522	0.129	0.522	0.123	†
alter v	0.400	0.357	0.129	0.355	0.125	†
rate t_0–t_1	3.000	2.951	0.325	2.950	0.309	†
rate t_1–t_2	4.000	4.125	0.446	4.130	0.396	†
density	–1.500	–1.575	0.086	–1.576	0.091	*
reciprocity	1.000	0.927	0.173	0.925	0.176	*
transitivity	0.400	0.416	0.061	0.416	0.074	*
3-cycle	–0.120	–0.129	0.107	–0.128	0.117	*
ego v	–0.200	–0.128	0.129	–0.128	0.129	
same v	0.600	0.500	0.128	0.499	0.124	†
alter v	0.400	0.344	0.126	0.343	0.125	
rate t_0–t_1	3.000	2.785	0.347	2.785	0.322	†
rate t_1–t_2	4.000	3.990	0.521	4.007	0.435	†
density	–2.000	–1.973	0.095	–1.972	0.097	*
reciprocity	2.000	1.926	0.180	1.921	0.180	
transitivity	0.500	0.482	0.079	0.481	0.090	*
3-cycle	–0.120	–0.105	0.152	–0.105	0.162	*
ego v	–0.200	–0.154	0.150	–0.154	0.149	
same v	0.600	0.554	0.145	0.551	0.138	†
alter v	0.400	0.370	0.145	0.367	0.142	†
rate t_0–t_1	3.000	2.865	0.358	2.874	0.306	†
rate t_1–t_2	4.000	3.976	0.433	4.020	0.359	†
density	–2.000	–1.984	0.139	–1.978	0.119	†
reciprocity	2.000	1.915	0.203	1.906	0.192	†
transitivity	0.700	0.699	0.113	0.689	0.085	†
3-cycle	–0.120	–0.084	0.177	–0.082	0.162	†
ego v	–0.200	–0.174	0.196	–0.175	0.189	†
same v	0.600	0.569	0.159	0.565	0.144	†
alter v	0.400	0.375	0.159	0.376	0.148	†

TABLE 2. *Simulation results, 3 waves for 50 actors: true value of the parameter (θ), mean ($\widehat{\theta}$), and root mean squared errors (RMSE) of the generalized method of moments (GMoM) and the method of moments (MoM). SAOMs with different parameter values for the reciprocity and the transitive triplet effects. An asterisk (*) in the last column denotes the effects for which the RMSE is lower for the GMoM than for the MoM; a dagger (†) denotes the effects for which the RMSE is higher for the GMoM than for the MoM (Wilcoxon signed rank test, p-value < 0.01).*

values of the parameters associated to reciprocity and transitivity triplets. The same procedure was followed, with results reported in Table 2. Again the algorithm always converged for the MoM, while modified version for the GMoM did not always reach convergence, and in such cases the MoM estimate was used instead. For the results in Table 2 non-convergence occurred in 4, 4, 6, and 22 simulations, respectively. The mean squared errors associated to the density, reciprocity, and transitive triplet, and 3-cycle parameters are mostly smaller for the GMoM than

	rate t_0–t_1	rate t_1–t_2	density	recip.	real recip.	trans.	real trans.	3-cycle	ego v	same v	alter v
rate t_0–t_1	1.000	0.132	0.261	0.214	0.110	0.163	0.053	0.156	0.079	0.191	0.129
rate t_1–t_2	0.132	1.000	0.476	0.374	0.124	0.471	0.174	0.439	0.153	0.304	0.251
density	0.261	0.476	1.000	0.799	0.549	0.762	0.523	0.702	0.163	0.510	0.352
recip.	0.214	0.374	0.799	1.000	0.699	0.685	0.465	0.693	0.229	0.513	0.263
real recip.	0.110	0.124	0.549	0.699	1.000	0.450	0.332	0.457	0.131	0.339	0.161
trans.	0.163	0.471	0.762	0.685	0.450	1.000	0.680	0.943	0.356	0.517	0.460
real trans.	0.053	0.174	0.523	0.465	0.332	0.680	1.000	0.634	0.195	0.340	0.261
3-cycle	0.156	0.439	0.702	0.693	0.457	0.943	0.634	1.000	0.372	0.509	0.419
ego v	0.079	0.153	0.163	0.229	0.131	0.356	0.195	0.372	1.000	0.281	0.672
same v	0.191	0.304	0.510	0.513	0.339	0.517	0.340	0.509	0.281	1.000	0.162
alter v	0.129	0.251	0.352	0.263	0.161	0.460	0.261	0.419	0.672	0.162	1.000
rate t_0–t_1	1.000	0.247	0.224	0.196	0.131	0.090	0.059	0.091	0.046	0.161	0.068
rate t_1–t_2	0.247	1.000	0.519	0.447	0.241	0.524	0.357	0.513	0.090	0.328	0.165
density	0.224	0.519	1.000	0.798	0.546	0.778	0.632	0.737	0.043	0.455	0.216
recip.	0.196	0.447	0.798	1.000	0.723	0.721	0.566	0.736	0.084	0.457	0.131
real recip.	0.131	0.241	0.546	0.723	1.000	0.478	0.357	0.500	0.057	0.250	0.082
trans.	0.090	0.524	0.778	0.721	0.478	1.000	0.830	0.963	0.110	0.480	0.223
real trans.	0.059	0.357	0.632	0.566	0.357	0.830	1.000	0.783	0.118	0.375	0.219
3-cycle	0.091	0.513	0.737	0.736	0.500	0.963	0.783	1.000	0.130	0.478	0.190
ego v	0.046	0.090	0.043	0.084	0.057	0.110	0.118	0.130	1.000	0.190	0.646
same v	0.161	0.328	0.455	0.457	0.250	0.480	0.375	0.478	0.190	1.000	0.094
alter v	0.068	0.165	0.216	0.131	0.082	0.223	0.219	0.190	0.646	0.094	1.000
rate t_0–t_1	1.000	0.312	0.327	0.223	0.191	0.195	0.185	0.180	0.050	0.175	0.095
rate t_1–t_2	0.312	1.000	0.594	0.476	0.255	0.641	0.443	0.612	0.145	0.306	0.262
density	0.327	0.594	1.000	0.779	0.484	0.824	0.715	0.779	0.138	0.416	0.366
recip.	0.223	0.476	0.779	1.000	0.687	0.747	0.632	0.777	0.241	0.432	0.282
real recip.	0.191	0.255	0.484	0.687	1.000	0.450	0.387	0.482	0.152	0.271	0.171
trans.	0.195	0.641	0.824	0.747	0.450	1.000	0.841	0.960	0.342	0.476	0.503
real trans.	0.185	0.443	0.715	0.632	0.387	0.841	1.000	0.789	0.238	0.386	0.389
3-cycle	0.180	0.612	0.779	0.777	0.482	0.960	0.789	1.000	0.381	0.479	0.460
ego v	0.050	0.145	0.138	0.241	0.152	0.342	0.238	0.381	1.000	0.334	0.557
same v	0.175	0.306	0.416	0.432	0.271	0.476	0.386	0.479	0.334	1.000	0.183
alter v	0.095	0.262	0.366	0.282	0.171	0.503	0.389	0.460	0.557	0.183	1.000

TABLE 3. *Mean correlation matrices of the statistics according to the blocks in Table 1.*

for the MoM, while those for the rate and covariate parameters are larger; except for the last case, where the reciprocity and transitivity parameters are 2 and 0.7, respectively. In this case the MoM outperforms the GMoM estimator.

Table 3 and Table 4 show the correlations between the new and the old statistics for the simulation examples. Correlations between network statistics are usually quite high because of the intrinsic properties of networks. Table 3 and Table 4 indicate that the correlation between new and the old statistics related to the same effect varies between 0.6 and 0.94. The highest correlation is observed for the simulation study whose results are shown in the last block of Table 2.

	rate t_0–t_1	rate t_1–t_2	density	recip.	real recip.	trans.	real trans.	3-cycle	ego v	same v	alter v
rate t_0–t_1	1.000	0.227	0.251	0.191	0.170	0.150	0.120	0.126	0.057	0.135	0.060
rate t_1–t_2	0.227	1.000	0.487	0.370	0.235	0.514	0.308	0.459	0.044	0.220	0.176
density	0.251	0.487	1.000	0.682	0.474	0.820	0.635	0.742	0.021	0.251	0.237
recip.	0.191	0.370	0.682	1.000	0.762	0.671	0.513	0.699	0.132	0.259	0.126
real. recip.	0.170	0.235	0.474	0.762	1.000	0.443	0.340	0.468	0.055	0.162	0.057
trans.	0.150	0.514	0.820	0.671	0.443	1.000	0.797	0.907	0.125	0.259	0.294
real trans.	0.120	0.308	0.635	0.513	0.340	0.797	1.000	0.702	0.044	0.140	0.174
3-cycle	0.126	0.459	0.742	0.699	0.468	0.907	0.702	1.000	0.182	0.256	0.224
ego v	0.057	0.044	0.021	0.132	0.055	0.125	0.044	0.182	1.000	0.213	0.389
same v	0.135	0.220	0.251	0.259	0.162	0.259	0.140	0.256	0.213	1.000	0.074
alter v	0.060	0.176	0.237	0.126	0.057	0.294	0.174	0.224	0.389	0.074	1.000
rate t_0–t_1	1.000	0.186	0.198	0.092	0.111	0.137	0.093	0.078	0.031	0.101	0.090
rate t_1–t_2	0.186	1.000	0.349	0.202	0.178	0.344	0.197	0.266	0.006	0.148	0.172
density	0.198	0.349	1.000	0.554	0.396	0.780	0.571	0.622	0.062	0.243	0.273
recip.	0.092	0.202	0.554	1.000	0.771	0.535	0.363	0.562	0.140	0.253	0.160
real recip.	0.111	0.178	0.396	0.771	1.000	0.408	0.268	0.431	0.127	0.199	0.157
trans.	0.137	0.344	0.780	0.535	0.408	1.000	0.784	0.772	0.210	0.294	0.381
real trans.	0.093	0.197	0.571	0.363	0.268	0.784	1.000	0.572	0.180	0.227	0.299
3-cycle	0.078	0.266	0.622	0.562	0.431	0.772	0.572	1.000	0.249	0.280	0.283
ego v	0.031	0.006	0.062	0.140	0.127	0.210	0.180	0.249	1.000	0.226	0.283
same v	0.101	0.148	0.243	0.253	0.199	0.294	0.227	0.280	0.226	1.000	0.096
alter v	0.090	0.172	0.273	0.160	0.157	0.381	0.299	0.283	0.283	0.096	1.000
rate t_0–t_1	1.000	0.242	0.204	0.197	0.167	0.171	0.128	0.157	0.076	0.156	0.114
rate t_1–t_2	0.242	1.000	0.399	0.339	0.189	0.422	0.225	0.410	0.146	0.256	0.196
density	0.204	0.399	1.000	0.721	0.518	0.764	0.580	0.686	0.120	0.434	0.255
recip.	0.197	0.339	0.721	1.000	0.766	0.656	0.458	0.683	0.181	0.397	0.222
real recip.	0.167	0.189	0.518	0.766	1.000	0.443	0.300	0.472	0.098	0.273	0.140
trans.	0.171	0.422	0.764	0.656	0.443	1.000	0.815	0.917	0.307	0.425	0.405
real trans.	0.128	0.225	0.580	0.458	0.300	0.815	1.000	0.655	0.190	0.299	0.284
3-cycle	0.157	0.410	0.686	0.683	0.472	0.917	0.655	1.000	0.320	0.408	0.366
ego v	0.076	0.146	0.120	0.181	0.098	0.307	0.190	0.320	1.000	0.255	0.594
same v	0.156	0.256	0.434	0.397	0.273	0.425	0.299	0.408	0.255	1.000	0.176
alter v	0.114	0.196	0.255	0.222	0.140	0.405	0.284	0.366	0.594	0.176	1.000
rate t_0–t_1	1.000	0.195	0.383	0.323	0.246	0.218	0.226	0.184	0.028	0.145	0.124
rate t_1–t_2	0.195	1.000	0.638	0.515	0.280	0.730	0.563	0.662	0.045	0.173	0.168
density	0.383	0.638	1.000	0.729	0.480	0.751	0.672	0.653	-0.004	0.228	0.202
recip.	0.323	0.515	0.729	1.000	0.863	0.676	0.577	0.729	0.057	0.211	0.039
real recip.	0.246	0.280	0.480	0.863	1.000	0.386	0.350	0.433	0.039	0.090	0.046
trans.	0.218	0.730	0.751	0.676	0.386	1.000	0.944	0.940	0.066	0.091	0.183
real trans.	0.226	0.563	0.672	0.577	0.350	0.944	1.000	0.776	0.009	0.068	0.114
3-cycle	0.184	0.662	0.653	0.729	0.433	0.940	0.776	1.000	0.087	0.076	0.091
ego v	0.028	0.045	0.004	0.057	0.039	0.066	0.009	0.087	1.000	0.178	0.483
same v	0.145	0.173	0.228	0.211	0.090	0.091	0.068	0.076	0.178	1.000	0.056
alter v	0.124	0.168	0.202	0.039	0.046	0.183	0.114	0.091	0.483	0.056	1.000

TABLE 4. *Mean correlation matrices of the statistics according to the blocks in Table 2.*

	$\widehat{\theta}_{MoM}$	$s.e.(\widehat{\theta}_{MoM})$	$\widehat{\theta}_{GMoM}$	$s.e.(\widehat{\theta}_{GMoM})$
rate 1	4.913	0.685	4.912	0.691
rate 2	3.879	0.585	3.876	0.579
density	–1.883	0.145	–1.882	0.139
reciprocity	1.470	0.295	1.461	0.282
transitivity	0.296	0.058	0.294	0.057
3-cycle	–0.193	0.149	–0.185	0.155
ego	0.350	0.212	0.349	0.212
similarity	0.476	0.207	0.474	0.207
alter	0.054	0.206	0.055	0.207

TABLE 5. *Model estimation for the van de Bunt data using the MoM and the GMoM.*

5. Empirical example

An illustrative example of the application of SAOMs is presented next to simulation results. Data were collected by Van de Bunt (1999) and are also discussed in Van de Bunt et al. (1999) and Snijders (2001).

The network is defined over a set of 32 university freshmen, 24 female and 8 male students, attending the same study program at a Dutch university. The relation analysed is friendship. Data were collected at 7 time points, but here only the last three waves were analysed. The covariate is a centered dummy variable for gender, equal to –0.25 for females and +0.75 for males.

Following the model-based approach typical of SAOMs, the model specification is based on theories concerning the mechanisms that might determine the creation and dissolution of ties. Earlier research on friendship networks helps in specifying the model. In particular, several studies (e.g. van Duijn et al., 2003; Knecht et al., 2010; Ko and Buskens, 2011) showed that friendship relationships show tendencies towards reciprocity and tend to triadic closure. Moreover, homophily with respect to gender is often observed. Therefore, the specification of the model in (13) can also be used in this illustrative example.

Table 5 shows the MoM and the GMoM estimates. The new statistics S^*_{rec}, S^*_{clos} and S^*_{agr} were used to estimate θ. For the GMoM, the number of iterations in phase 1 was $n_1 = 200$ and convergence was reached after running the estimation twice, using the results of the first run as starting values for the second. For the MoM, the number of iterations in phase 1 was $n_1 = 34$ and convergence was reached after running the algorithm once. For both estimators, convergence of the algorithm was checked by simulating the model for the observed parameters $n_3 =$4,000 times and computing the (weighted) ratio between the mean value of the simulated statistics and their standard deviations.

The GMoM and the MoM estimates lead to similar values of the parameters and their standard errors. In particular, the standard errors associated to the density, reciprocity, and transitive triplet effects are smaller for the GMoM than for the MoM. But some of the other standard errors are larger.

Given the results in Table 5, it is possible to make inference on the mechanisms that might have generated the observed sequence of networks (conditional on the first of the analyzed waves). The significance of each parameter is tested using the ratio between the corresponding estimate and standard error. Assuming that the estimators are approximately normally distributed, the

test statistic can be tested in the standard normal distribution (Snijders, 2001). This shows that several parameters of the evaluation function are significant. The negative parameter associated to the density effect corresponds to the sparseness of the observed networks and suggests that the cost of an isolated tie is higher than its benefit. Such a cost can be attenuated when a tie is reciprocated, closes a transitive triplet, and/or is directed to a person of the same gender. Indeed, the parameters associated to reciprocity and transitive triplets are positive, providing evidence of direct reciprocation of friendship ties and triadic closure, as suggested by earlier literature.

Finally, the positive value of the parameter of the similarity effect supports the theory of homophily of friendship ties with respect to gender, i.e., there is a stronger tendency to create and maintain friendship ties between actors having the same gender than between actors with different gender. The 3-cycle effect is not significant and therefore data do not provide evidence for or against generalized reciprocity. Furthermore, there is no significant effect of gender on the activity (ego) or popularity (alter) of the freshmen. Finally the rate parameters indicate that on average freshmen got the opportunity to change their outgoing ties 5 times in the first period and 4 times in the second period.

6. Discussion

This article proposed an estimator based on the generalized method of moments for the parameter of stochastic actor-oriented models (SAOMs) as an alternative to the estimators based on the regular method of moments and the maximum likelihood estimation, respectively. Since the log-likelihood function of SAOMs cannot be computed explicitly, the maximum likelihood estimator is hard to compute and the method of moments (MoM) is still the most used estimation method. Therefore, only the method of moments and its generalization were considered in this paper.

Compared to the regular method of moments, the generalized method of moments allows using more statistics than parameters, exploiting more information in the observed data. In the case of SAOMs, the additional information derives from the use of new statistics which do not depend solely on the network observed at time t_m, like the statistics that have been used until now, but also on the network observed at t_{m-1}. Although the real sequence of changes between two consecutive observations of a network is not known, the advantage of the new statistics is that in a way they trace the pattern of ties from which a certain configuration originated.

For each regular statistic several new statistics can be defined. Following the principle of the method of moments, the use of the new statistics (next to the regular statistics) results in an over-identified system of equations whose solution is the generalized method of moments (GMoM) estimator. This estimator is defined as the value minimizing a quadratic function of the differences between the expected values of the statistics and their sample counterparts, or equivalently, the root of the first order derivative of the quadratic function. The latter can be approximated by pre-multiplying the differences between the expected values of the statistics and their sample counterparts by a matrix of weights B, measuring the relevance of each statistic to the estimation of the corresponding parameter. It follows that the Robbins-Monro algorithm, currently used to approximate the solution of the system of moment equations for the SAOM according to the MoM, can be modified to account for the matrix B and compute the estimator based on the GMoM. Various adaptations, mentioned above, were needed to make this algorithm work in practice.

The algorithm for the GMoM is slower than the one for the MoM, primarily because the estimate of the matrix B and of the covariance matrix of the estimator require a higher number of simulations during both phase 1 and phase 3 of the algorithm. Moreover, a common practice in estimating the parameter of the SAOM when the algorithm does not reach convergence, is to re-run the Robbins-Monro algorithm skipping the first phase and using the estimates deriving from the first run of the algorithm as the initial parameter value (Ripley et al., 2015). During the simulations it was almost always necessary to use this procedure when using the GMoM.

The small simulation studies presented in the previous pages shows that the GMoM and the MoM yield similar estimates. In general, the simulations do not support the intuition that the relative advantage of the GMoM estimator is greater when the observations of the network are close in time. The comparison of the RMSEs shows that the estimator based on the GMoM is slightly more efficient than that based on the MoM only for the effects related to the new statistics. However, the GMoM does not generally outperform the MoM, and in one of our examples the MoM performs even better than the GMoM. One reason may be the use of a weight matrix W with zero off-diagonal blocks for entries combining rate parameters and objective function parameters. Another reason might be *redundancy*, i.e., a high correlation between the new and the old statistics. Simulations in economic studies showed that redundancy might affect the quality of the GMoM. Therefore, the over-identifying statistics should be carefully chosen in order not to incur collinearity.

In network analysis, the statistics are highly correlated because of the intrinsic properties of networks and establishing which correlations are "too high" is not so obvious. Nevertheless, the inspection of the correlation matrices of the statistics used in the simulation examples indicates that the higher the correlation between the new and old statistics for a certain effect, the lower the gain of the GMoM. When some correlations get closer to 1, the MoM can even outperform the GMoM. Redundancy might also affect the convergence of the algorithm, as suggested by the number of simulations for those the GMoM algorithm did not reach convergence: this seemed to increase with the correlations between the new and the old statistics.

It follows that further investigations into the effect of redundancy on the efficiency of the GMoM, and the development of procedures for choosing new statistics, are desirable for the practical use of the GMoM estimator.

The choice of the parameters for the simulation study was based on empirical results from the estimation of the SAOM in friendship networks, and the use of new statistics defined in the previous pages. However, SAOMs are used to analyse many other relations, and new statistics can be defined with respect to more complex formulations of the model; e.g. for analysing the co-evolution of networks and behaviours (Steglich et al., 2010). For some of these models we expect that the GMoM may be seriously better than the MoM, and we are continuing investigations of the efficiency of the GMoM, as depending of the choice of additional statistics, for estimating more complex versions of the SAOM.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 319209.

References

Bowman, K. O. and Shenton, L. R. (1985). Method of moments. In Kotz, S. and Johnson, N. L., editors, *Encyclopedia of Statistical Sciences*, volume 5, pages 467–473. Wiley, New York.

Burguete, J. F., Ronald Gallant, A., and Souza, G. (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews*, 1(2):151–190.

Carrington, P. J., Scott, J., and Wasserman, S. (2005). *Models and Methods in Social Network Analysis*. Cambridge University Press, New York.

Doreian, P. and Stokman, F. N. (1997). The dynamics and evolution of social networks. In Doreian, P. and Stokman, F., editors, *Evolution of Social Networks*, pages 1–18. Routledge, Amsterdam.

Doreian, P., Teuter, K., and Wang, C.-H. (1984). Network autocorrelation models: Some Monte Carlo results. *Sociological Methods & Research*, 13(2):155–200.

Freeman, L. C. (1984). The impact of computer based communication on the social structure of an emerging scientific specialty. *Social Networks*, 6(3):201–221.

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

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford.

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

Holland, P. W. and Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20.

Knecht, A., Snijders, T. A. B., Baerveldt, C., Steglich, C., and Raub, W. (2010). Friendship and delinquency: Selection and influence processes in early adolescence. *Social Development*, 19:494–514.

Ko, P.-C. and Buskens, V. (2011). Dynamics of adolescent friendships: the interplay between structure and gender. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 313–320. IEEE.

Koskinen, J. and Edling, C. (2012). Modelling the evolution of a bipartite network-peer referral in interlocking directorates. *Social Networks*, 34(3):309–322.

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

Leenders, R. T. A. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47.

Lehmann, E. (1999). *Elements of Large-Sample Theory*. Springer, New York.

Lospinoso, J. A., Schweinberger, M., Snijders, T. A. B., and Ripley, R. M. (2011). Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Advances in Data Analysis and Classification*, 5(2):147–176.

Luce, R. and Suppes, P. (1965). Preference, utility, and subjective probability. *Handbook of Mathematical Psychology*, 3:249–410.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Pflug, G. C. (1990). Non-asymptotic confidence bounds for stochastic approximation algorithms with constant step size. *Monatshefte für Mathematik*, 110(3-4):297–314.

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

Ripley, R. M., Snijders, T. A. B., Bóda, Z., Vörös, A., and Preciado, P. (2015). Manual for RSiena. Technical report, Oxford: University of Oxford, Department of Statistics; Nuffield College.

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

Robins, G., Elliott, P., and Pattison, P. (2001a). Network models for social selection processes. *Social Networks*, 23(1):1–30.

Robins, G., Pattison, P., and Elliott, P. (2001b). Network models for social influence processes. *Psychometrika*, 66(2):161–189.

Ruppert, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

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

Scott, J. and Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. SAGE publications, London.

Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, 21(1-2):149–172.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395.

Snijders, T. A. B., Koskinen, J., and Schweinberger, M. (2010a). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2):567–588.

Snijders, T. A. B., Lomi, A., and Torlò, V. (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., Van de Bunt, G. G., and Steglich, C. E. G. (2010b). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60.

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

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

Van de Bunt, G. G. (1999). *Friends by choice: An actor-oriented statistical network model for friendship networks through time*. Interuniversity Center for Social Science Theory and Methodology, Groningen/Amsterdam.

Van de Bunt, G. G., Van Duijn, M. A. J., and Snijders, T. A. B. (1999). Friendship networks through time: An actor-oriented dynamic statistical network model. *Computational & Mathematical Organization Theory*, 5(2):167–192.

van Duijn, M., Zeggelink, E. P. H., Huisman, M., Stokman, F. N., and Wasseur, F. W. (2003). Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology*, 27:153–191.

Wasserman, S. (1980a). Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, 75(370):280–294.

Wasserman, S. (1980b). A stochastic model for directed graphs with transition rates determined by reciprocity. *Sociological Methodology*, 11:392–412.

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

Wasserman, S. and Iacobucci, D. (1988). Sequential social network data. *Psychometrika*, 53(2):261–282.