BAYESIAN INFERENCE FOR LONGITUDINAL SOCIAL NETWORKS

JOHAN KOSKINEN

ABSTRACT. A natural approach for modeling stochastic processes on social networks is by using continuous-time Markov chains, examples of which have been given by Wasserman (1977, 1980b,a) and Leenders (1995b,a). Snijders (1996) proposed a class of models that allow for greater flexibility in defining the dynamic components, relaxing the restrictions on the type of dependence structures that could be modeled. Previously, estimation of the parameters in such models has been based on a Markov chain Monte Carlo (MCMC) implementation of the method of moments. In this paper we generalize the class of stochastic actor-oriented models, and propose an MCMC algorithm for exploring the posterior distribution of the parameters. The generalized class of stochastic actor oriented models can handle un-directed, bipartite and valued social networks in addition to the dichotomous directed networks of the stochastic actor oriented models. The MCMC procedure explicitly models the changes in-between observations as latent variables.

1. INTRODUCTION

Social network analysis is concerned with the patterns of relationships between actors. Typically the social network is conceived as graphs, directed or un-directed, in which the vertices represent actors and the edges or arcs represent the relation of interest. For the actors $i, j = 1, \ldots, n$ the relational ties are recorded as x_{ij}, representing the strength of the relation from i to j. For symmetric relations such as friendship, acquaintance, and collaboration, x_{ij} often is a binary variable indicating the presence or absence of the relation with $x_{ij} = x_{ji}$. Being for example the pattern of giving/receiving of advice among actors one has to discriminate between x_{ij} and x_{ji}, the first saying whether j receives advice from i whereas the latter indicates whether j gives advice to i. In addition, both are instances of non-reflexive relations, since relations from an actor to himself is not meaningful. The proximity of structural concepts and measurements has contributed to making quantitative analysis a popular tool in social network analysis and has found many applications since its introduction, generally attributed to Moreno (1934). An introduction and fairly exhaustive review is given by Wasserman and Faust (1994).

Key words and phrases. Longitudinal social networks; data augmentation; Bayesian inference; valued relations; random graphs.

The author is grateful for comments and suggestions by Tom A. B. Snijders, with whom a previous version with the title "Bayesian and Likelihood Based Inference for Longitudinal Social Network Data" was presented at the 2003 Sunbelt XXIII, Cancun, Mexico.

Whereas quantitative analysis has been the prevalent mode of investigation of social networks, realistic statistical modeling is notoriously complicated by the interdependencies naturally arising between the tie variables. Building on the work of Besag (1974), the nature of the structural dependencies in social networks were thoroughly investigated in Frank and Strauss (1986), leading to a class of exponential random graph models. Popularized and further elaborated by Wasserman and Pattison (1996) much work has been done in the field of estimation (Strauss and Ikeda, 1990; Wasserman and Pattison, 1996; Corander et al., 1998; Besag, 2000; Hancock, 2000; Snijders, 2002). The study of the dependence structures that underpin social networks and has been further investigated by Frank and Nowicki (1993); Robins (1998); Pattison and Robins (2002); Robins and Pattison (2004).

Another line of attack has been to model the tie variables as conditionally independent conditional on latent structures. Originating in models for stochastic block models (Fienberg and Wasserman 1981; Holland et al. 1983; see also Frank 2004 for some recent elaborations of the p_1 model), via recent approaches to parameter inference (Snijders and Nowicki, 1997; Nowicki and Snijders, 2001; Tallberg, 2004), the fixed latent blocks have been elaborated to include latent settings (see e.g. Schweinberger and Snijders, 2003, on ultra metrics) and Euclidean or arbitrary metric (social) spaces (Hoff et al., 2002).

The study of longitudinal social network data offers a wealth of information about the network (see e.g. Wasserman, 1980a; Doreian and Stokman, 1997). Instead of a single observation it is assumed that we have a sequence of M observations $X(0), X(t_1), \ldots, X(t_{M-1})$. Several methods have been proposed for analyzing repeated observations on social networks using models where changes are made in discrete steps from one moment to the next (Katz and Proctor, 1959; Wasserman and Iacobucci, 1988; Sanil et al., 1995; Banks and Carley, 1996; Robins and Pattison, 2001). As argued in for example Snijders (1996), considerable advantages can be had from modeling longitudinal social networks in continuous time. Especially using continuous-time Markov chains, as proposed by Kalbfleisch and Lawless (1985) for non-network data. Early models for longitudinal social networks using continuous-time Markov chains include Holland and Leinhardt (1977a,b) and Wasserman (1977). A model taking the dependence between x_{ij} and x_{ji} into consideration, called the reciprocity model, was proposed by Wasserman (1977, 1980b,a) and further investigated by Leenders (1995b,a). The reciprocity model assumes that the dyads in a directed graph, (X_{ij}, X_{ji}), evolve as mutually independent Markov chains. Although providing a computationally attractive model, the assumption of independent dyads is somewhat contrived in most instances of social networks, since other structural features of the network are neglected. In response to this Snijders (1996) proposed a class of model, stochastic actor-oriented models. In this and subsequent papers (Snijders and van Duijn, 1997; Snijders, 2001), a class of models were developed where the rate of change as well as the type of changes taking place are allowed to depend on higher order features of the network structure as well as known covariates. This allows for a great flexibility in modeling data while not making concessions in terms of empirical testability.

The relevance of this class of models was also shown in that the reciprocity model could be seen as a special case (Snijders and van Duijn, 1997) and that the p^*-model can be obtained as the limiting distribution for certain parameterizations (Snijders, 2001).

The likelihood function for stochastic actor-oriented models is however only available in closed form in a few special cases wherefore estimation has had to rely on a Markov chain Monte Carlo (MCMC) implementation of the method of moments. More specifically, a stochastic approximation algorithm, a version of the Robbins-Monro (1951) algorithm was used to solve the moment equations numerically. Generally for these types of models the moment estimators do not coincide with the Maximum likelihood estimate. Parameter-estimate uncertainty as measured by the approximate standard errors obtained from the delta method (Bishop et al., 1975), rely on asymptotics and numerical approximations in several phases. The performance of these asymptotics for small samples are unknown but a reasonable assumption is that they are not any better than in the case of probit models (c.f. Griffiths et al., 1987).

In this paper we provide procedures for conducting "exact" Bayesian inference. In addition we extend the class of stochastic actor oriented models, both with respect to what kind of structures can be analysed as well as providing means for relaxing some of the assumptions, notably, the axiom of independence of irrelevant alternatives (usually abbreviated IIA).

The proposed inference scheme consists of augmenting the observed data with a latent variable describing the network evolution in-between observations. To employ data augmentation to facilitate inference was first suggested by Tanner and Wong (1987) and is nowadays a commonly used tool in Bayesian statistics. Once data has been augmented with the latent variable a conventional Metropolis (Metropolis et al., 1953) algorithm can be implemented, making the necessary modifications to handle variable model dimensionality (Green, 1995; Richardson and Green, 1997). MCMC algorithms have been used to analyse binary longitudinal data in epidemiology with various restrictions on the indicators of being in infection-states. Gibson (1997) considered the order of infections for a simple epidemic, with non-recurrent infections. This has been elaborated to handle increasing degrees of complexity (Gibson and Renshaw, 1998; O'Neill and Roberts, 1999; Auranen et al., 2000; Eerola et al., 2003). To the best of our knowledge, none of them are suited to the kind of dependencies that are usually associated with social network analysis. The mechanisms of the evolution of social networks arguably differ a great deal from the mechanisms of infections, which are to their nature passively received as opposed to consciously sought. This has consequences both for the design of the models and the inference procedure. In this paper, for example, a certain emphasis is put on the generation of the latent evolutionary paths.

We now proceed to give the general formulation of a network process along with examples of specifications and their interpretations. The main components of the proposed inference procedure are presented in the following section and the details

of some implementational issues are commented and illustrated in the concluding empirical section.

2. Notation and model formulation

In the following we consider a fixed set of actors represented by the vertex set $V = \{1, \ldots, n\}$ and a fixed set of (ordered) pairs of actors $\mathscr{N} \subseteq V \times V$, $N = |\mathscr{N}|$. For example, symmetric, non-reflexive relational data would have $\mathscr{N} = \{(i,j) \in V \times V : 1 \leq i < j \leq n\}$. For an affiliation network with a set C of organizations and a set M of members, the vertices of the corresponding bi-partite graph would be $V = C \cup M$, and the set of pairs of interest $\mathscr{N} = C \times M$. We limit the study to relations with ordinal or nominal scale that takes values in the label set $\mathscr{R} = \{0, 1, \ldots, R-1\}$. A network can be described by a (di-) graph with vertices V and set of coloured (arcs) edges $\mathscr{N}$, with colours in $\mathscr{R}$. We let the generalized adjacency matrix of this (di-) graph consists of a collection $X = (X_e : e \in \mathscr{N})$ of variables with range space $\mathscr{X} = \mathscr{R}^N$.

For the elements of $\mathscr{X}$ define the distance metric

$$|x - y| = \sum_{e \in \mathscr{N}} |x_e - y_e|,$$

which in the case of binary $\mathscr{R}$ becomes the Hamming metric. The network evolution model is a continuous-time Markov chain on $\mathscr{X}$, but for our purposes is best defined in terms of the embedded chain. The transition probabilities in the embedded chain are

$$\pi(\theta, x, y), \tag{2.1}$$

and the time spent in $x \in \mathscr{X}$ exponentially distributed with rate

$$\lambda(\theta, x). \tag{2.2}$$

It is assumed that $\pi(\theta, x, y) > 0$ for $x, y \in \mathscr{X}$ such that $|x - y| = 1$ and 0 otherwise. Note that the underlying graph of the state-space graph of the embedded chain is an R-ary N-cube. The event that the embedded chain traverses one of the edges of the state space graph constitutes a *mini-step*. Processes such as those of Mayer's (1984) where more than one mini-step type change can occur simultaneously are not considered. The $p \times 1$ parameter vector $\theta \in \Theta \subseteq \mathbb{R}^p$ includes all unknown parameters. These two functions, π and λ, determine the rate functions

$$q(\theta, x, y) = \lambda(\theta, x)\, \pi(\theta, x, y)$$

which defines the generator of the continuous-time process on $\mathscr{X}$.

Assume that we have observations on the network $X(t)$ for fixed time points $t_0 < t_1 < \cdots < t_{M-1}$. The analysis is throughout made conditional on the first observation at t_0. Because of the Markov property we can drop the notational dependency on the observation points, in order to make the notation more lucid. Note also that we need not concern ourselves with the kind of stationarity assumptions for the marginal distributions $X(t)$ needed for estimation of parameters in

the expansiveness and popularity models (Wasserman, 1977, 1980a). In the sequel we refer by s and t to two generic consecutive observation points, $s < t$, with $T = (t - s)$.

For t and s, denote the distance between these two observations by

$$H = |x(t) - x(s)|.$$

To construct a sequence of m ministeps that transforms $x(s)$ into $x(t)$, we let the times at which these ministeps occur be a sequence of m event times $s_1, s_2, \ldots, s_m$, with $s < s_1 < \cdots < s_m < t$. At time s_h the value of only one element e_h is changed by $a_h \in \{-1, 1\}$. Hence the value on the edge e immediately after the change at time s_h is made, is $x_e(s_h) = x_e(s_{h-1}) + a_h$ if $e = e_h$, and otherwise it is unchanged, i.e. $x_e(s_h) = x_e(s_{h-1})$. The time between consecutive changes, the holding times, are the differences $u_h = s_h - s_{h-1}$. For the last change at time s_m we have $x_e(s_m) = x(t)$, for $e \in \mathscr{N}$. The total number of changes m is at least equal to the distance between the observations at time s and t, but can also include an even number of extra steps. Thus we can describe the complete observation consisting of the $m = H + 2k$, for some $k \in \{0, 1, 2, \ldots\}$, changes transforming $x(s)$ into $x(t)$, in terms of a series of holding times, indicators of what elements are changing, and the direction of change. A complete observation of the continuous-time stochastic process can be expressed by m triples

$$(u_h, e_h, a_h)$$

where $h = 1, 2, \ldots, m$, $e_h \in \mathscr{N}$, and $a_h \in \{-1, 1\}$, (for all h), subject to the following constraints:

$$\begin{array}{ll} (1) & u_h > 0 \text{ (all } h); \ \sum_{h=1}^{m} u_h \leq T; \\ (2) & \text{for each } e \in \mathscr{N}, \text{ and for, } h = 1, 2, \ldots, m, \text{ the partial sums} \\ & y_{e,h} = x_e(s) + \sum_{v=1}^{h} a_v \mathbf{1}\{e_v = e\} \\ & y_{e,h} \in \mathscr{R} \\ & y_{e,m} = x_e(t) \end{array} \tag{2.3}$$

For binary $\mathscr{R}$, a_h is redundant and condition 2 can be expressed as that the number $m_e = \sharp\{h \in \{1, \ldots, m\} : e_h = e\}$, is odd if and only if $x_e(t) \neq x_e(s)$. This can be expressed in terms of the a_h's as that when a change is made to element e_h, if the last change to that particular element was $a_{h'} = a$, then $a_h = -a$. Adding the number of individual changes we must have $\sum_e m_e = m$.

A sequence of triples subject to the constraints 2.3, defines an observation on a completely observed process. With this representation, the likelihood of augmented observations

$$x(s) = x(s_0), x(s_1), x(s_2), \ldots, x(s_m) = x(t)$$

of the continuous-time stochastic process can be written as

$$\exp\left\{-\sum_{h=1}^{m+1} u_h\,\lambda(\theta, x(s_{h-1}))\right\} \times \prod_{h=1}^{m} \pi\left(\theta, x(s_{h-1}), x(s_h)\right)\lambda\left(\theta, x(s_{h-1})\right), \tag{2.4}$$

where $u_{m+1} = t - s_m$, given that all constraints mentioned are satisfied.

The focus on mini-steps might seem too restrictive for valued relational ties and perhaps larger steps would seem more intuitive. In the current framework, though, this restriction is a necessity since larger steps would introduce indeterminacy in the model - a big step is hard to distinguish from many mini-steps.

2.1. **Applications.** Social network data comes in many shapes and sizes and each structural form bring with it different kinds of dependencies between the elements of $\mathscr{N}$. The network evolution models gives the researcher great freedom in modeling these dependencies. Since this is done through focusing on the mini-steps and the mini-step has a different interpretation for different relational structures, different specifications apply. Here follows a few examples.

2.1.1. *Directed graphs.* For directed graphs and vertex set V, $R = 2$, and $\mathscr{N} = \{(i,j) \in V^2 : i \neq j\}$. For a stochastic actor-oriented model (Snijders, 2001) the following is assumed for the behaviour of the actors and the process for t in the time window $[t_0, t_1)$:

- **i):** Control. Every actor has complete control over his or her out-going ties
- **ii):** Relative myopia. The decisions of the actors are based only on the present state and the states that can be reached by a single change to their composition
- **iii):** Complete information. Each actor is assumed to have full knowledge about the state of the network at each given time.

Given the present state $x \in \mathscr{X}$, the rate at which actor i's performs a mini-step is modeled by the individual change rate $\lambda_i(x, \theta)$. Given the state $x \in \mathscr{X}$, independently for each $i \in V$, the time until actor i decides to make a change to the composition of his out-going ties is exponentially distributed with rate $\lambda_i(x, \theta)$. Since the event that someone makes a change changes the composition of the network, all actors have to re-evaluate the network when such an event takes place. From the properties of the exponential distribution we have that given that a change is made, the actor who may change his out-relations is actor i with probability $\lambda_i(x, \theta)/\lambda(x, \theta)$, where

$$\lambda(x, \theta) = \sum_{j \in V} \lambda_j(x, \theta). \tag{2.5}$$

Denote by $x(i \rightsquigarrow j)$ the adjacency matrix that differs from x in exactly the element $(i, j) \in \mathscr{N}$. Now, make the assumption that for given x, when actor i evaluates the configurations obtainable $\mathscr{X}_i(x) = \{y \in \mathscr{X} : y = x(i \rightsquigarrow j) \text{ for some } j \in V \backslash \{i\}\}$ by making a change, i assigns a measure $U_i(t, x, j)$ of the attractiveness of each

configuration relative to the present configuration x. In addition to assumption i), ii), and iii), we now assume that

iv): Given configuration x, and that actor i may change his out-going relations, i changes his out-relations to the actor j that maximizes $U_i(t, x, j)$.

The utilities are conceived as consisting of two basic components, a systematic and a random component, which we express as

$$U_i(t, x, j) = r(\theta, i, j, x) + \epsilon_i(t, x, j),$$

where $\epsilon_i(t, x, j)$ is the random component. Assume further that $\epsilon_i(t, x, j)$ are independently and identically distributed according to the Type 1 extreme value distribution for all t, x, and j. This is a common and convenient specification which leads to the familiar Conditional Multinomial Logit (CML) model for each choice situation (see e.g. McFadden, 1974). For the CML formulation it is straightforward to show that given configuration x and that actor i may change his out-going relations, i changes his out-relations to actor j with probability

$$\frac{e^{r(\theta,i,j,x)}}{\sum_{k\in V\setminus\{i\}} e^{r(\theta,i,k,x)}}. \tag{2.6}$$

It terms of the social network evolution process we see that Eq. (2.5) corresponds to Eq. (2.2) and that the one-step probability 2.1 of going from x to $x(i \rightsquigarrow j)$, is given by $\lambda_i(x, \theta)/\lambda(x, \theta)$ times (2.6). For various sociological aspects of stochastic actor oriented models see for example Zeggelink (1994).

The previously mentioned reciprocity model (Wasserman, 1977, 1980b,a; Leenders, 1995b,a), can be seen as a special case of the actor oriented model (c.f. Snijders and van Duijn, 1997) and thereby a special case of the network evolution model. For a parameter vector $\theta = (\theta_1, \theta_2, \theta_3)$, with $\theta_1 > 0$, define

$$\lambda_{ij}(\theta, x) = x_{ij}e^{\theta_2}\left[x_{ji}e^{\theta_3} + (1 - x_{ji})\right] + (1 - x_{ij})\, e^{-\theta_2}\left[x_{ji}e^{-\theta_3} + (1 - x_{ji})\right]$$

for each $(i, j) \in \mathscr{N}$. This constitutes the generator of the continuous-time process, i.e. $q(\theta, x, x(i \rightsquigarrow j)) = \lambda_{ij}(\theta, x)$. The individual change rates are then obtained through

$$\lambda_i(\theta, x) = \frac{\theta_1}{n-1} \sum_{j\in V:(i,j)\in\mathscr{N}} \lambda_{ij}(\theta, x),$$

and the systematic part of the utility is modeled as a weighed sum of the number of out-going ties and the number of reciprocated ties for i in the new configuration

$$\theta_2 \sum_{j=1}^{n} x_{ij} + \theta_3 \sum_{j=1}^{n} x_{ij}x_{ji}.$$

The transition probabilities in the embedded chain incorporate information on how many ties each node has and how many of them are reciprocated. Specifically for the reciprocity model, the general propensity to change is modeled by θ_1, if all other parameters are zeros, the expected time until a specific actor makes a change is $[\theta_1/(n-1)]^{-1}$. For $\theta_2 > 0$ actors with few relations to others are most eager to change and when making a change, the change is likely to increase the actors

out-going relations. The worth, from the perspective of the ego, is taken to be the number of such ties weighted by θ_3. Note that whereas the method of moments estimates coincide with the maximum likelihood estimates for the independent arcs model, this is not the case for the reciprocity model. The reason for this is that the former belongs to the exponential family of distribution and the latter to the curved exponential family of distributions (c.f. Snijders and van Duijn, 1997).

An example of higher order sub-graph counts is the indirect relations effect, defined as the number of actors one actor is only indirectly related to. For an actor i, this can be defined for one intermediary

$$\sharp\{j : x_{ij} = 0, \max_k(x_{ik}x_{kj}) > 0\},$$

as well as for two intermediaries

$$\sharp\{j : x_{ij} = 0, \max_k(x_{ik}x_{kj}) = 0, \max_h(x_{ik}x_{kh}x_{hj}) > 0\},$$

or for an arbitrary number of intermediaries less than $n-1$. The contribution of the number of indirect relations to the utility for actor i of configuration x can be given the form

$$\theta \sum_{j \in V\setminus\{i\}} (1 - x_{ij}) \max_k(x_{ik}x_{kj}).$$

The relative desirability of having structurally similar (Burt, 1976) alters could be expressed by including

$$\theta \sum_{j \in V\setminus\{i\}} x_{ij} \sum_{k \in V\setminus\{i,j\}} |x_{ik} - x_{jk}|.$$

in the utility for i of configuration x, for some θ.

Typically, stochastic actor-oriented models include predictors based upon actor-bound covariates, dyad-covariates, and interactions between these and the structural statistics.

2.1.2. *Graphs.* For graphs with a fixed set of vertices V, having $R = 2$, since the relation is symmetric and $\mathscr{N} = \{(i,j) \in V \times V : 1 \leq i < j \leq n\}$, we can not, as in the case of directed graphs, assume that the actors control their out-going ties. The individual change rates have to be replaced by dyad rates $\lambda_{ij}(\theta, x)$ for each dyad $(i,j) \in \mathscr{N}$, that measures the relative stability of each dyad. Alternatively, (2.2) can be defined directly for the whole graph, using for example standard sociometric measures of stability, balance, etc. The one-step transition probabilities, can be defined using a multinomial logit or probit link function.

As an example, consider the case of modeling friendship formation. It seems natural to view friendship as something that is driven by mutual friends (as suggested for example in Leenders, 1995b, ch. 4). The simplest model with $\theta = (\rho, \beta)$, could have a constant rate $\lambda(\theta, x) = \rho$, for all $x \in \mathscr{X}$, and

$$\pi(\theta, x, y) = \frac{\exp\left\{\beta \sum_{(i,j,k)\in\binom{V}{3}} y_{ij}y_{ik}y_{jk}\right\}}{\sum_{z\in\mathscr{X}:|x-z|=1} \exp\left\{\beta \sum_{(i,j,k)\in\binom{V}{3}} z_{ij}z_{ik}z_{jk}\right\}},$$

which simplifies to

$$\frac{\exp\left\{\beta \sum_{k\in V\setminus\{i,j\}} (1-x_{ij})\, x_{ik}x_{jk}\right\}}{\sum_{(k,\ell)\in\mathscr{N}} \exp\left\{\beta \sum_{v\in V\setminus\{k,\ell\}} (1-x_{k\ell})\, x_{kv}x_{\ell v}\right\}},$$

for y that differs from x in exactly the element (i,j). For $\beta > 0$, the more common friends two actors have, the more probable it is that they will become friends. This simplistic transitivity model can well be extended with covariates - both dyad specific and actor specific - as well as more sophisticated network effects. In particular, the structural position of a tie in the network should be incorporated, for example structural holes (Burt, 1992) or Simmelian ties (Krackhardt, 1999).

2.1.3. *Bipartite graphs.* In the field of criminal networks, a common unit of study is the co-offending network (Reiss, 1986, 1988; Reis and Farrington, 1991; Sarnecki, 1999; Frank, 2001). For a set of offenders A, the co-offending network has the adjacency matrix $Y = (Y_{ij} : 1 \leq i < j \leq n)$, where the entry i, j, is 1 or 0 according to whether i and j has been recorded to commit a crime together. For a set of crimes C, Y is obtained from the offender by crime matrix X, where the entry (i,k), is 1 if $i \in A$ has been recorded to commit crime $k \in C$, and 0 otherwise. The rate of change $\lambda(\theta, x)$ is subdivided into, $\sum_{i\in A} \lambda_i(\theta, x)$, modeling variable levels of activity among the criminals and the duration of the crimes, crimes being conceived as long-term projects rather than fleeting liaisons. It is natural, from the time perspective to define the values as non yet committed, in progress and completed, denoting them 0,1, and 2 respectively. For the purpose of the evolution model consider now defining vertices as $V = A \cup C$, and pairs $\mathscr{N} = A \times C$, and $R = 3$. We are able to model the evolution of the criminal network taking for example third party effects into account, i.e. how the "decision" to commit a crime together with someone previously unknown is influence by a third person (or group) that both offenders are related to. When events occur transition are made with probabilities $\pi(\theta, x, x')$, decomposed in a way similar to the stochastic actor oriented models. This is interpreted such that given that a change is made, the probability that it is criminal i who decides to act is $\lambda_i(\theta, x)/\lambda(\theta, x)$. Conditional on i making a change, i chooses to become involved in a new crime, decides to participate in a crime in progress or terminates the involvement in a crime. The relative merits of each cause of action in the eyes of the criminals can be modeled as a random utility model through $r(\theta, x, y)$, assigning utilities to each option.

Another important field of study that concerns bi-partite graphs is sexual contacts between heterosexuals (often with claims to policy implications Bearman et al., 2004, see e.g.).

2.1.4. *Valued relations.* As noted in for example Robins et al. (1999), measurements of social networks are often in valued form, but typically data is dichotomized. Working directly with valued data complicates statistical modeling if not because many standard network concept do not apply, then because of the combinatorial explosion. Within the proposed modeled framework, using the original valued relations instead of the dichotomized network comes at a low computational cost,

on the contrary one can argue that the loss of detail entailing dichotomization seldom motivates disregarding the labels. Consider the case when the original data takes values $\mathscr{R} = \{0, 1, 2, 3, 4\}$, measuring say, the strength of acquaintances, and that we have dichotomized data, letting 0 and 1 denote the absence of a tie and 2 through 4 the presence of a tie. The dichotomized variable $X_e(t)$ then takes values 1 or zero according to whether there exists a tie or not for $e \in \mathscr{N}$, t in the time window $[t_0, t_1)$. If we for $e \in \mathscr{N}$, have recorded the dicotomized values $X_e(t_0) = 0$, and $X_e(t_1) = 1$, we do not take into account whether the strength has really increased from 1 to 2 or from 0 to 4. Even more serious, some changes might not be recorded at all. If the dichotomized data is recorded as $X_e(t_0) = 1$, and $X_e(t_1) = 1$, this could mean that no change has been made, but there is great difference between a dyad that stays at a high level of intensity - a strong tie - and one sustaining a lower level of intensity. There might in fact also have been a decline or a growth in intensity.

A word of caution is that when labels refer to frequencies of interaction (e.g., EIES, number of messages sent Freeman and Freeman, 1979), one has to ascertain the time-dimension of the labels is negligible in comparison to the time between observations. Standard statistical considerations naturally comes into play, such as assuring that the scale of the colors are comparable across pairs in $\mathscr{N}$. An example of an analysis of valued relations is given in Section 4.4.

3. Parameter estimation

The purpose in this section is to describe a procedure for obtaining the posterior distribution of the parameter vector θ. Denote the prior distribution for the parameters $\pi(\theta)$, the form of which is not elaborated upon further here. Since the likelihood given data is available in a closed form only in a limited number of specifications on the longitudinal model (for example for the reciprocity model Leenders, 1995b,a; Snijders, 1999), we employ a reversible jump Metropolis-Hastings algorithm (Green, 1995), which augments observed data with latent data.

For data $x(s)$, $x(t)$, and specification $\lambda(\theta, x)$ and $\pi(\theta, x, y)$, denote the data likelihood $L_D(\theta; x)$. For a sequence $(u_h, e_h, a_h)_{h=1}^m$ latent variables $w = (y_h, u_h)_{h=1}^m$ can be constructed to form a walk from $x(s)$ to $x(t)$ in the state space graph of the embedded chain with holding times. The range space of the latent variables is defined to comply with the constraints 2.3. For given m let $\mathscr{S}_m = \{(u_1, \ldots, u_m) \in (0, T)^m : u_1 + \cdots + u_m < T\}$, and let $\mathscr{W}_m(x, y)$ be the set of all walks in the R-ary N-cube from x to y, $x, y \in \mathscr{X}$. Where it is clear from the context we set $\mathscr{W}_m = \mathscr{W}_m(x(s), x(t))$. For a fixed m, the latent variables has range space $\mathscr{C}_m = \mathscr{W}_m \times \mathscr{S}_m$, and in general w takes values in $\mathscr{C} = \cup_{k=0}^{\infty} \mathscr{C}_{H+2k}$. The complete data likelihood, or augmented likelihood, function $L(\theta; w, x(s))$ conditional on $x(s)$, $x(t)$ and a w, is given by (2.4). The MCMC scheme consists of constructing a sequence $(\theta^{(r)}, w^{(r)})_{r=0}^G$ that converges to a sample from the joint posterior distribution of θ and w given data by in each iteration successively employing the two move types

(a): updating the latent variables w
(b): updating the parameters θ.

Of course, this is a very general formulation but it is important to point out that (a) involves altering the dimension of (θ, w) since $(u_h)_{h=1}^m$ contains m continuous variables and m is variable. We assume in the following that (b) can be performed using a conventional Metropolis step. In some instances it is worthwhile specifying the mini-steps with a probit link function, in which case w can be augmented even further with latent variables (Albert and Chib, 1993). We will not dwell on this possibility here but merely point out that in that case (b) can be performed with direct sampling from standard distributions.

Both (a) and (b) are designed to sample from the full conditional posterior given data and conditional on all other parameters and latent variables. The full conditional posterior of w conditional on data and θ, is given by

$$\pi(w|\theta, x(t), x(s)) = L(\theta; w, x(s))/L_D(\theta; x) \propto L(\theta; w, x(s)).$$

For given present state $w \in \mathscr{C}$, let $q(\cdot|w)$ be a candidate distribution for drawing w^* with range space $\mathscr{C}(w) \subset \mathscr{C}$, conditional on w. Hence for move type (a), for the current value $w \in \mathscr{C}_m$, a move to $w^* \in \mathscr{C}_{m^*}$ sampled from $q(w^*|w)$, is accepted with probability $\min\{1, A\}$, where

$$A = \frac{L(\theta; w^*, x(s))}{L(\theta; w, x(s))} \times \Pi, \tag{3.1}$$

and

$$\Pi = \frac{q(w|w^*)}{q(w^*|w)} J,$$

in which J is the Jacobian of the transformation of going from w to w^*, to make sure that the continuous part of $\pi(w^*|\theta, x(t), x(s))q(w|w^*)$ has a finite density with respect to a symmetric measure on $\mathscr{S}_{m^*} \times \mathscr{S}_m$ for each $(y_h^*)_{h=1}^{m^*} \in \mathscr{W}_{m^*}$ and $\{y_h\}_{h=1}^m \in \mathscr{W}_m$ (Green, 1995). As usual, in the ratio of full conditional posteriors in the RHS of (3.1) the normalizing constants, in this case L_D, cancels out. In a similar fashion, for move type (b), given the current parameter value θ, a move to θ^*, sampled from $q(\theta^*|\theta)$, is suggested. The expression for A in the acceptance probability corresponding to (3.1) becomes

$$\frac{L(\theta^*; w, x(s))\pi(\theta^*)}{L(\theta; w, x(s))\pi(\theta)} \times \Pi, \tag{3.2}$$

where π denotes the prior distribution, and since this move preserves the dimensions, Π simplifies to the ratio $q(\theta|\theta^*)/q(\theta^*|\theta)$.

This is so far fairly standard procedures for Bayesian inference and for general aspects of implementation we refer to the extensive MCMC literature (e.g. Gilks et al., 1996). Most issues are likely to depend on the particular specification of the transition probabilities (in the embedded chain), and the rate functions (for example what proposal distribution to use, whether to thin the sample or not, etc.) and we touch on a few of these issues in the empirical section. We need however, to elaborate on the form of the proposal distribution for the latent variables. There are many candidates for proposal distributions, particular in the case of $\mathscr{R}$ binary (c.p. e.g. Auranen et al., 2000), and it is especially simple to construct proposals

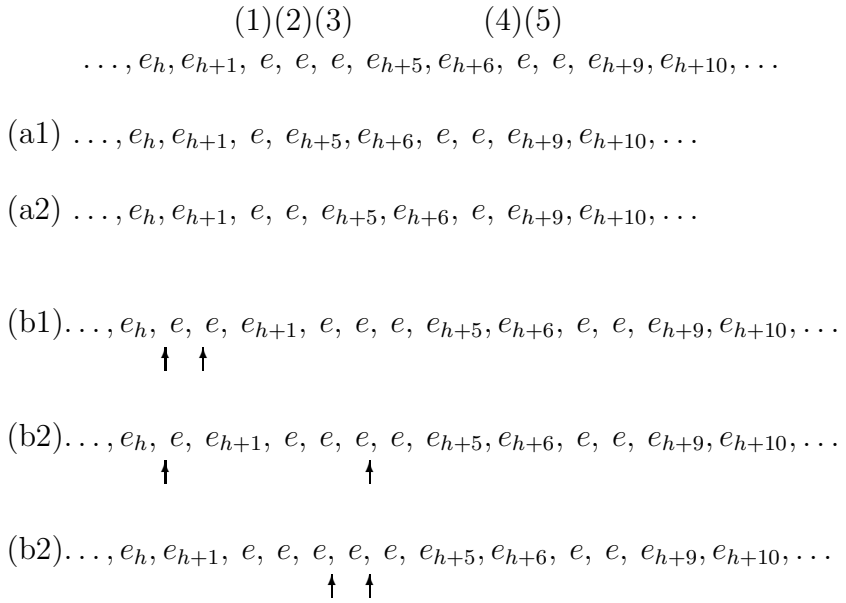

FIGURE 1. Illustration of shortenings (a1 and a2) and prolongings (b1, b2, b3) by making a change to the element e

that draw either separate coordinates or entire walks independently of the previous state of the algorithm. For the relatively complex network evolution models however, the algorithm fails if too large jumps in $\mathscr{C}$ are proposed. We propose a nearest neighbor candidate distribution that for binary $\mathscr{R}$ takes the topology of the binary N-cube into consideration and respects the importance of order in the sequence of changes. For valued data, the nearest neighbour proposal is supplemented by independently of the previous state conditional on the proposed sequence of changes, drawing lattice walks for specific coordinates. Although the supplemented candidate distribution no longer has the nearest neighbour interpretation, there is sufficient dependence between proposed walks and previous walks.

3.1. **Nearest neighbour proposal.** Recall that the longitudinal social network process with $\mathscr{R} = \{0, 1\}$, has a representation as random walk on a hypercube $\mathscr{G}$ with vertices consisting of all N-bit binary strings and a step from one vertex to another is performed by adding 1 modulo 2 to the bit in which the strings differ. For a given walk we can obtain a new one by inserting two "unnecessary" changes to a bit. The walk then makes a detour and the vertices visited in between these changes are identical to the corresponding vertices in the original walk bar for the bit that was changed. If for the present walk $m \geq H + 2$, there exists such a detour that can be removed. For a walk in $\mathscr{G}$ between two vertices we define three types of neighbourhoods and endeavor to design a proposal which samples conditionally uniformly on these. The three neighbourhoods consists of the prolonged, shortened, or swapped walks that can be obtained from a walk by the operations defined below.

Given $w \in \mathscr{C}_m$ with the representation $(u_h, e_h)_{h=1}^m$, define for each $e \in \mathscr{N}$ and $\ell \in \{1, \ldots, m\}$

$$K_{e\ell} = \sharp\{h \in \{1, \ldots, m-\ell+1\} : e_{h+v} = e, \text{ for } v = 0, \ldots, \ell-1$$
$$\text{subject to } e_{h+\ell} \neq e, \text{ or } h+\ell-1 = m, \text{ and } (e_{h-1} \neq e \text{ or } h = 1)\},$$

which counts the number of un-interrupted sequences of changes to the same element e, e-runs, of a given length ℓ. Note that $\sum_\ell \ell K_{e\ell} = m_e$, and for $\mathscr{R} = \{0, 1\}$, $a_i = r$ implies $a_{i+1} = -r$. A sequence of changes $(e_h)_{h=1}^m$ subject to the above mentioned constraints can be shortened by removing two elements e from two distinct e-runs or removing two changes from the same e-run. This is illustrated in Figure 1, where the sequence at the top has two e-runs, one of length 3, and one of length 2. The shortened walk (a1) is obtained by removing any two of the three consecutive changes, (1), (2), and (3), from the first e-run. By removing one element from each of the two e-runs, say (3) and (4), we obtain (a2). The total number of distinct walks that can be obtained in this manner is

$$\gamma_e^- = \binom{\sum_{\ell \geq 1} K_{e\ell}}{2} + \sum_{\ell > 1} K_{e\ell}$$

ways, where $\binom{k}{2}$ is defined as 0 for $k < 2$. The first term counts the number of ways there is to remove two elements from distinct e-runs whereas the second term is the number of ways of removing elements from the same e-runs.

For inserting two changes to a fixed element e in an walk $(e_h)_{h=1}^m$, let $I_{\tilde{e}} = \{h \in \{1, \ldots, m\} : e_h \neq e\}$ be the positions not occupied by a change to the element e. An extra change to e can be inserted in a space immediately before a position $h \in I_{\tilde{e}}$ or in the spot immediately after $\max I_{\tilde{e}}$. The sequence (b1) in Figure 1, has been obtained by inserting two e's in the space in front of position $h+1$. Similarly, (b3) can be obtained by inserting two changes e in the e-run in front of position $h+5$. By instead inserting one e in each space we obtain (b2). The number of insertion points, or distinct bins, is given by

$$|I_{\tilde{e}}| + 1 = m - m_e + 1.$$

Addition of the number of prolonged walks for which both changes are inserted in the same spot and when the changes are inserted into distinct spots gives the number of walks that can be obtained from prolonging by adding a change to e

$$\gamma_e^+ = \binom{m - m_e + 1}{2} + m - m_e + 1 = \binom{m + 2 - m_e}{2}.$$

The sequence can be re-ordered by swapping places for two e_h and $e_{h'}$, $e_h \neq e_{h'}$. This can be done in

$$\binom{m}{2} - \sum_{e: m_e > 1} \binom{m_e}{2}, \tag{3.3}$$

ways.

We are now equipped with three move-types for up-dating sequences. The number of changes can be increased from m to $m+2$ chosen uniformly at random

from either one of the

$$\gamma^+ = \sum_{e \in \mathscr{N}} \gamma_e^+,$$

distinct sequences obtainable from prolonging. The number of changes can be reduced from m to $m-2$ chosen uniformly at random from either one of the

$$\gamma^- = \sum_{e \in \mathscr{N}} \gamma_e^-,$$

distinct sequences obtainable from shortening. Furthermore, swaps can be affected with probability the reciprocal of (3.3). In the sequel we call these move type 1,2 and 3 respectively. In practice, move type 1 is performed by choosing the element e to be affected with probability γ_e^-/γ^-, and then proceeding sequentially by choosing to remove e from the same e-run or from two distinct e-runs, with probabilities $\sum_{\ell>1} K_{e\ell}/\gamma_e^-$ and $1 - \sum_{\ell>1} K_{e\ell}/\gamma_e^-$, respectively. If the former removal type is selected, one can simply draw two distinct e-runs uniformly at random. For a removal of the latter type, choose one e-run from among the e-runs with a lengths greater or equal than two at random and remove two of the changes. For move type 2, a similar sequential procedure is used in practice. The element e to be inserted is chosen with probability γ_e^+/γ^+. With the appropriate probabilities chose to either insert the two e's at the same insertion point or place them in two distinct insertion points. Given a walk $(y_h)_{h=1}^m \in \mathscr{W}_m$, move type 1 proposes a move to a walk in $\mathscr{W}_{m+2}$, type 2 a walk in $\mathscr{W}_{m-2}$ and move type 3 proposes a walk in $\mathscr{W}_m$. To take the cardinality of the reachable subsets of $\mathscr{W}_{m+2}$, $\mathscr{W}_{m-2}$, and $\mathscr{W}_m$, into consideration has proved to be of some importance for the performance of the algorithm for complex models.

We choose to perform move type j with probability ζ_j if all three move types are possible and $\zeta_j' = \zeta_j/(\zeta_1 + \zeta_3)$, for $j = 1, 3$ if $m = H$. Conditional on the move type, a change to $(e_h)_{h=1}^m$ is made as described above, and conditional on this change and $u = (u_h)_{h=1}^m$, $u^* = (u_h^*)_{h=1}^{m^*}$ is proposed in the following manner. The technique described here is closely related to the way in which Richardson and Green (1997) up-date the weights in a mixture model. The mains differences are that whereas mixture weights can be up-dated by proposing a new weight for a single coordinate and the requirement that the weights sum to unity, we propose a pair of new "weights" with the requirement that all the weights sum to T.

For move type 1 and the new sequence $(e_h^*)_{h=1}^{m+2}$ of length $m+2$ with additions of an element e in positions h_1^* and h_2^*, $e_{h_1^*}$ and $e_{h_2^*}$, we need to propose holding times corresponding to these positions, $u_{h_1^*}$ and $u_{h_2^*}$. When one or both of the new additions have been placed in positions adjacent to changes to the same element, in an e-run, the holding times are up-dated starting with the last position in the run. These are drawn from a rescaled Dirichlet distribution with parameters $(1, 1, m+1)$, with density

$$\frac{\Gamma(m+3)(T - u_{h_1^*} - u_{h_2^*})^m}{\Gamma(m+1)T^{m+2}},$$

where $T = t - s$ as defined earlier. To "make space" for the new holding times the old holding times are scaled down $u^*_{h'} = u_h(T - u_{h^*_1} - u_{h^*_2})/T$, where the position h' in relation to h is determined unambiguously by the position of h in relation the insertion points (again with some modifications as mentioned above). The Jacobian of this transformation is $[(T - u_{h^*_1} - u_{h^*_2})/T]^m$ and hence

$$\frac{q\left(u|(e_h)_{h=1}^m, w^*\right)}{q\left(u^*|w, (e^*_h)_{h=1}^{m+2}\right)} J = \frac{T^2}{(m+2)(m+1)}. \tag{3.4}$$

Thus Π equals (3.4) times $\zeta_2\gamma^+/(\zeta_1\gamma^-)$, if $k \geq 1$, and $\zeta_2\gamma^+/(\zeta_1'\gamma^-)$ otherwise.

For move type 2 and the removal of $e_{h^*_1}$ and $e_{h^*_2}$, from a sequence of length $m+2$, set the new elements $u^*_{h'} = u_h T/(T - u_{h^*_1} - u_{h^*_2})$, where the indices h' are matched to h for $h \neq h^*_1, h^*_2$. $q\left(u|(e_h)_{h=1}^{m+2}, w^*\right)/q\left(u^*|w, (e^*_h)_{h=1}^m\right) J$ then becomes the reciprocal of the RHS of (3.4). For move type 3, u^* is drawn from the symmetric distribution on $\mathscr{S}_m$, and $\Pi = 1$. For some applications, conditional on move type 1 (or 2) sampling the entire vector u^* independently of u works well. This is accomplished in a similar way to how holding times are up-dated for the move type 3, by drawing u^* from the symmetric distribution on $\mathscr{S}_{m+2}$. Because of the independence, the ration of proposal distributions becomes the ratio of scaled Dirichlet distributions with the appropriate dimension, the expression of which incidentally coincides with (3.4).

To ensure that the procedures presented thus far produces a Markov chain that is aperiodic and irreducible (with respect to the joint posterior distribution of the latent walks and parameters), we note firstly that aperiodicity is guaranteed by the construction of Metropolis up-dating steps (c.f. Tierney, 1994). For irreducibility, observe that each part of $\bigcup_{k=0}^\infty \mathscr{W}_{H+2k}$, can be reached by a sequence of prolongings and shortenings, and for each swap the support of the holding time proposal is $\mathscr{S}_m$.

3.2. **Proposal for colours.** For valued data, i.e. now the edges take values in $\mathscr{R} = \{0, \ldots, R-1\}$, the direction of a change to an element e is not uniquely determined by the previous value. The proposed holding times and change indices $(u^*_h, e^*_h)_{h=1}^{m^*}$ are drawn using the nearest neighbour proposal as described above. For coordinates in $\mathscr{N}$ that have been effected the directions of change are drawn conditionally independent of $(a_h)_{h=1}^m$ conditionally on $(u^*_h, e^*_h)_{h=1}^{m^*}$. In other words, if the old position of e^*_h was h', we set $a^*_h = a_{h'}$. For move type 1 and 2, and a change to $e \in \mathscr{N}$, denote by $I = \{h \in \{1, \ldots, m^*\} : e^*_h = e\}$ all the positions occupied by a change to e. The idea is to propose a new sequence of directions, a^*_h for $h \in I$, transforming $x_e(s)$ into $x_e(t)$. Given that a swap has been made the procedure described below is repeated for both the affected coordinates.

The total number of sequences $(a_{i_j} : a_{i_j} = \pm 1)$, of length m^* giving paths from $x_e(t_0)$ to $x_e(t_1)$ is given by $\binom{m^*_e}{N^+}$, where $N^+ = \sharp\{i_j \in I : a_{i_j} = 1\}$. As in (2.3) denote the partial sums $y_{i_j} = y_{i_{j-1}} + a_{i_j}$, for $i_j \in I$, and $y_0 = x_e(s)$. From standard lattice path counting techniques (Fray and Roselle, 1971) we have that the number

of paths such that $0 \leqslant y_{i_j} < R$, for all $i_j \in I$, is given by

$$C_{m_e^*}(y_0, y_{i_{m_e^*}}) = \sum_{k\in\mathbb{Z}} \left[\binom{m_e^*}{N^+ + k\,(\ell + v)} - \binom{m_e^*}{N^+ + k\,(\ell + k) + \ell} \right],$$

where $\ell = x_e(s)+1$, $v = R-x_e(s)$. This formula is obtained by iteratively applying the reflexion principle. Sampling a path $y_0, y^*_{i_1}, \ldots, y^*_{i_{F_e^*}}$, uniformly at random is easily done by drawing a sequence of N^+ up-steps and m_e^* down-steps, discarding sequences for which $\max_{i_j}\{y_{i_j}\}$ is greater than $R-1$, or $\min_{i_j}\{y_{i_j}\}$ is less than 0. The probability mass function of any accepted sequence is $P(y_0, y^*_{i_1}, \ldots, y^*_{i_{m_e^*}})$ $= C_{m_e^*}(y_0, y^*_{i_{m_e^*}})^{-1}$. An alternative, when rejection sampling is inefficient, is to draw the steps in a sequential manner. Let $N_j^+ = N^+(a_{i_1}, \ldots, a_{i_j})$ denote the number of positive steps until the j'th step and set $a_{i_{j+1}} = 1$, given the previous $a_{i_1}, \ldots, a_{i_j}$, with probability

$$\begin{cases} 1 & \text{if } y_{i_j} = 0 \\ 0 & \text{if } y_{i_j} = R-1 \\ \left(N^+ - N_j^+\right) / \left(m_e^* - j\right) & \text{o.w.} \end{cases}.$$

The probability mass function of a realization is given by $P(y_0, y^*_{i_1}, \ldots, y^*_{i_{m_e^*}}) = P(y^*_{i_1}|y_0) \cdots P(y^*_{i_{m_e^*}}|y_{i_{m_e^*-1}})$. Note that the p.m.f. of the previous sequence of changes$P(y_0,y_{i_1},\ldots,y_{i_{m_e}})$ of colours is inexpensive to compute if done in the process of proposing the new sequence.

4. Empirical illustration

In this section we analyse the Electronic Information Exchange System (EIES) data collected by Freeman and Freeman (1979) to illustrate specifications for the network evolution model and highlight some implementation issues. There were two measurements of acquaintanceship taken 8 months apart, and the complete data contains $n = 32$ actors. Since there are only 2 observations, the time scale is arbitrary and we set $T = 1$. Each actor was required to give the status of his relation to each other actor and thus $\mathscr{N} = V^{(2)}$, and the number of non-redundant entries in each adjacency matrix is $n(n-1) = 992$. Data was originally coded as 0 (person unknown to me, or no reply); 1 (person I've heard of but not met); 2 (person I've met); 3 (friend); and, 4 (close personal fiend). In addition to the network data, an individual-bound covariate w_i was recorded for each actor being a measure of the attractiveness of the individuals research. If individual i had 12 or less than citations of his work in SSCI in the year before the first network measure was taken $w_i = 1$, and $w_i = 0$ if he had more than 12 citations.

Of the three main parts of this section, the first two deals with the analysis of the dichotomized data, while the third works directly with the original values. In the analysis of the dichotomized data, the standard form of the adjacency matrix $X(t) = (x_{ij}(t))$ is used in which $x_{ij}(t)$ is equal to 1 or 0 according to whether (i, j) is a friendship relation at time t or a null relation. In accordance with Snijders and van Duijn (1997) null relations are defined as those with strengths less or equal to 1. The diagonal elements are set to structural zeros. For the analysis

of the valued data, $x_{ij}(t)$ is the strength of the relation from i to j at time t, as expressed by the original values.

4.1. **Random utilities.** In this section we analyse the EIES data using one of the models fitted in Snijders and van Duijn (1997) but conducting likelihood based inference rather than the method of moments.

The intention is to model the evolution of the network as driven by the actors conscious efforts at seeking better network configurations. To this end we employ a random utility effect where the utility of a network configuration to an actor is combination of covariate and network effects as well as a random component, as described in Section 2.1.1..

For the parameter $(p+2) \times 1$ vector $\theta = (\rho, \alpha, \beta_1, \ldots, \beta_p)'$, ρ and α only occur in the expression for the individual change rates and β only appear in the utility part.

The most basic structural effect to consider is the activity of actors. An actor is said to be active if he or she seeks to have many ties to other actors, which suggests the out-degree statistic

$$s_{i1}(x) = \sum_j x_{ij}.$$

The interpretation of a tie from i to j often has completely different interpretations depending upon whether j also has a tie to i. This is captured by the reciprocity statistic

$$s_{i2}(x) = \sum_j x_{ij}x_{ji}.$$

To study the influence on friendship seeking of the covariates, we include

$$s_{i3}(x) = \sum_j x_{ij}w_j,$$

to capture whether actors with low citation rates are less popular, and

$$s_{i4}(x) = \sum_j x_{ij}\left|w_i - w_j\right|,$$

to investigate whether there is any clusterings based on the citation rate.

In addition to the statistics based on individuals and dyads, we include the transitivity statistic

$$s_{i5}(x) = \sum_{j,h} x_{ij}x_{ih}x_{jh}.$$

A transitive triple or triad in which i is a head, is the sub (di-) graph consisting of three vertices, where i has ties to both other vertices and where there is a tie between these. In terms of a friendship relation, a transitive triple is a configuration in which the friends of i are also friends.

To make allowances to differential preferences in the eagerness to make changes to ones personal network we define

$$\lambda_i(x, \theta) = \rho\{e^{\alpha} + \frac{s_{i1}(x)}{n-1}(e^{-\alpha} - e^{\alpha})\},$$

which is to be interpreted as that actors with few ties are more eager to change, for $\alpha > 0$. When $\alpha < 0$, it is actors with many ties that are more active in making changes to their composition.

For specifying the systematic part of the utility, we let $\mathbf{s}_i(x) = (s_{i1}(x), \ldots, s_{ip}(x))$, and write

$$r(\theta, i, j, x) = \mathbf{s}_i(x(i \rightsquigarrow j))\beta. \tag{4.1}$$

Given data and the latent variable $w \in \mathscr{C}_m$, the complete data likelihood of θ is easily obtained from (2.4) and (2.6) as proportional to

$$\begin{aligned} &\exp\left\{-\sum_{h=1}^{m+1} u_h \sum_{i\in V}\left[e^{\alpha} + \frac{s_{i1}(y_{h-1})(e^{-\alpha}-e^{\alpha})}{n-1}\right]\right\}\rho^m \\ &\times \prod_{h=1}^{m}\left[e^{\alpha} + \frac{s_{i_h 1}(y_{h-1})(e^{-\alpha}-e^{\alpha})}{n-1}\right]\frac{e^{\mathbf{s}_{i_h}(y_h)\beta}}{\sum_{z\in\mathscr{X}_i(y_{h-1})} e^{\mathbf{s}_{i_h}(z)\beta}}, \end{aligned} \tag{4.2}$$

where for each h, $e_h = (i_h, j_h)$. The full conditional posterior of θ given w and $x(t_0)$, is proportional to (4.2) times the prior $\pi(\theta)$.

4.1.1. *Prior distributions.* Throughout the empirical section, all prior distributions for the parameters are vague. To ascertain analytically whether the joint posterior distribution of the parameters is proper is in general difficult. A commonly applied ad-hoc criterion for the appropriateness of the posteriors is to monitor the progress of the MCMC-algorithm. The idea is that the algorithm would fail to converge if the posterior were improper. A danger of using this ad-hoc criterion is that conclusions regarding posterior moments might not be valid. No doubt, the elicitation of prior distributions is an important issue for further studies. A particularly important aspect is the prior relation between the parameters. The nature of the relation between the statics, e.g. between out-degree and transitivity, suggests some sort of prior dependence between the corresponding parameters.

4.1.2. *Implementation.* Proposals for moves of type (a), were chosen according according to the procedure in Section 3.1., with $(\zeta_1, \zeta_2, \zeta_3) = (2/7, 4/7, 1/7)$. The relative weight assigned to shortenings, prolongings and swaps, respectively, does not seem affect the performance to any great extent other than that it is a desirable property that the algorithm is allowed to move fairly freely up and down.

For the parameters and moves of type (b), a Metropolis up-dating step was employed. Given the current parameter vector $\theta = (\rho, \alpha, \beta_1, \ldots, \beta_p)'$, a candidate vector θ^* was proposed from

$$\left(\alpha^*, \beta_1^*, \ldots, \beta_p^*\right)' \sim N_{p+1}\left((\alpha, \beta_1, \ldots, \beta_p)', \Omega\right),$$

and independently thereof

$$\rho^* \sim Gamma(\rho k, k^{-1}).$$

The proposal ratio, Π in (3.2), hence becomes

$$k^{k(\rho^*-\rho)}\Gamma(k\rho)/\Gamma(k\rho^*)\, e^{k(\rho-\rho^*)}\rho^{k\rho^*-1}/(\rho^*)^{k\rho-1}.$$

The performance of this MCMC scheme crucially depends on the choice of Ω and setting this requires a fair amount of fine-tuning (setting k proves a much less demanding task than setting Ω, and therefore we concentrate on the latter). The dangers and pit-falls of the choice of proposal distribution are well known (c.f. the review given by Chib, 1995), yet there does not seem to be any consensus on automatization strategies. In the empirical examples, we have relied on a combination of test runs and adaptive scaling. It is commonly suggested that the variance of the normal proposal be set to some constant times the variance, or approximation to the variance, in the target distribution (e.g. c times the inverse information matrix evaluated at the posterior mode Tierney, 1994; Roberts et al., 1997). In this particular case, there are a variety of options for choosing Ω. To obtain a provisional proposal variance, consider conditioning on a path $w \in \mathscr{C}_F$, in which case the augmented likelihood can be evaluated. One can get a rough estimate by evaluating the likelihood in a grid and then fitting a normal kernel using restricted least squares or using the properties of the multinomial probit (see below and the full conditional posteriors); or by maximizing the augmented likelihood with the quasi-Newton algorithm which produces an estimate of the Hessian en passant.

Once a provisional value is obtained, one, a couple or a few test-runs are performed to get an estimate $\widehat{\mathbf{\Sigma}}$ of the posterior covariance matrix of $(\alpha, \beta)'$. For the analysis carried out here, we set $\Omega = \frac{\gamma}{\sqrt{p+1}}\widehat{\mathbf{\Sigma}}$, for γ roughly in the interval $(.8, 3)$.

4.2. **Normal random components.** Despite the attractiveness of the expression (2.6) for the multinomial probabilities in the CML, this formulation is open to criticism. First and foremost we have axiom of independence of irrelevant alternatives (IIA, c.f. McFadden, 1974, ; note however that IIA only applies whitin the mini-step for the logit specification). For fixed V this should not be a limitation but there is a definite advantage to be had from taking *leavers* and *joiners* into consideration (c.p. Huisman and Snijders, 2003). Additionally, for affiliation networks there are substantial reasons for allowing for correlated utilities and varying spread. An interesting question is whether the skewness of the extreme value type 1 distribution is theoretically warranted.

Apart from the increased flexibility with a multinomial probit model (MNP), there are some computational advantages.

For a MNP formulation of the model described above, we retain the systematic part in Equation (4.1) but make other assumptions for the random component ϵ. Collect the utilities in an $n-1$ by 1 vector $\widetilde{\mathbf{v}}_h=(\widetilde{v}_j : j \neq i_h)$, where for each h, $e_h = (i_h, j_h)$, and given the $(n-1) \times p$ matrix of predictors $\mathbf{S}_h=(\mathbf{s}_{i_h,j} : j \neq i_h)$, $\mathbf{s}_{i_h,j} = \mathbf{s}_{i_h}(y_{h-1}(i_h \rightsquigarrow j))$, where p is the number of rows in β, we can write

$$\widetilde{\mathbf{v}}_h = \mathbf{S}_h\beta + \epsilon.$$

As in the case of the CML formulation of the evolution model, the random components are assumed to be independent over time but whereas it was previously assumed ϵ_j was independently and identically distributed according to the Gumbel distribution, we now assume that $\epsilon \sim \mathbf{N}_{n-1}(\mathbf{0}, \mathbf{\Gamma}(i_h, y_{h-1}))$. Correlated random

components can be motivated in analysis of for example affiliation networks, where the choice categories do not vary over actors, but in the present context an unrestricted covariance matrix is over specifying the model. Hence we follow the formulation of the independent multinomial probit model and set $\mathbf{\Gamma}(i,x) = \tau^2\mathbf{I}$ for all $i \in V$ and all $x \in \mathscr{X}$. Following standard practice we normalize on one of the alternatives, k, to achieve identification, by taking the differences $\varepsilon_j = \epsilon_j - \epsilon_k$, $v_j = \widetilde{v}_j - \widetilde{v}_k$ and $z_j = \mathbf{s}_{i_h,j} - \mathbf{s}_{i_h k}$. The system of equations now becomes $\mathbf{v}_h = \mathbf{z}_h\beta + \varepsilon_h$, and thus

$$\mathbf{v}_h \sim \mathbf{N}_{n-2}(\mathbf{z}_h\beta, \mathbf{\Sigma}),$$

where the diagonal elements of $\mathbf{\Sigma}$, $\sigma_{jj} = \gamma_{kk} + \gamma_{jj} - 2\gamma_{jk}$, and for the off-diagonal elements $\sigma_{j\ell} = \gamma_{kk} + \gamma_{j\ell} - \gamma_{kj} - \gamma_{1\ell}$. Naturally, for the independent multinomial probit model $\mathbf{\Sigma} = \mathbf{1}_{(n-2)\times(n-2)} + \mathbf{I}_{(n-2)}$.

4.2.1. *Sampling from the full conditional posteriors.* With normally distributed random components sampling from the posterior, in each iteration draws are made from 5 full conditional posteriors:

(a): $(u_h, e_h, \mathbf{v}_h)_{h=1}^m |\, x(t_0), x(t_1)\rho, \alpha, \beta$
(b): $\rho |\, (u_h, e_h, \mathbf{v}_h)_{h=1}^m, x(t_0), x(t_1), \alpha, \beta$
(c): $\alpha |\, (u_h, e_h, \mathbf{v}_h)_{h=1}^m, x(t_0), x(t_1), \rho, \beta$
(d): $\beta |\, (u_h, e_h, \mathbf{v}_h)_{h=1}^m, x(t_0), x(t_1), \rho, \alpha.$

In addition it is suggested that a step (a2) is added where only the latent differences ($\mathbf{v}_h$) are up-dated. Whereas step (a) does not differ much from in the logit formulation, and is carried out using a Metropolis step, as is step (c), we can sample directly from the full conditional posteriors of ρ and β in steps (b) and (d)

The full conditional posterior of a complete observation $(u_h, e_h, \mathbf{v}_h)_{h=1}^F$ is proportional to

$$\begin{aligned}
&(2\pi)^{-m(n-2)} \left|\mathbf{\Sigma}\right|^{-m/2} \exp\left[-\sum_{h=1}^{m+1} u_h \,\lambda\left(\theta, y_{h-1}\right)\right] \\
&\times \prod_{h=1}^{m} \lambda_i\left(\theta, y_{h-1}\right) \exp\left[-\frac{1}{2}\left(\mathbf{v}_h - \mathbf{z}_h\beta\right)' \mathbf{\Sigma}^{-1}\left(\mathbf{v}_h - \mathbf{z}_h\beta\right)\right] \\
&\times \prod_{h=1}^{m} \mathbf{1}\left\{\mathbf{v}_h \simeq y_{h-1}, y_h\right\},
\end{aligned}$$

where $\mathbf{1}\{\mathbf{v} \simeq x, y\}$ simply is an indicator function which is one or zero according to whether utilities are concordant with the transition from x to y.

Since explicitly modeling the latent utilities introduces extra randomness, the performance of the algorithm is even more sensitive to the choice of candidate distribution for complete observations. We will only mention one possible candidate. The holding times and element indicators (u_h, e_h) are proposed according to the nearest neighbor proposal distribution as described in Section 3.1. For each h, conditional on (u_h^*, e_h^*), $\mathbf{v}_h^*$ is drawn from $\mathbf{N}_{n-2}(\mathbf{z}_h^*\beta, \mathbf{\Sigma})$, truncated such that $\mathbf{1}\left\{\mathbf{v}_h^* \simeq y_{h-1}^*, y_h^*\right\} = 1$. How to draw random vectors with the desired truncation is

more closely described below. Suffice it to show at this point that when $\mathbf{\Gamma}(i,x)=\mathbf{I}$, we only really need a realization of $\mathbf{v}_h^*$ if the proposed move is accepted. Since the functional form of the pdf is not altered by the truncation, and the density of $\mathbf{v}_h^*$ appears in the expression for the full conditional posterior, all that remains of the ratio of full conditional posteriors and proposal distributions is, suppressing the indicator functions,

$$\exp\left[-\sum_{h=1}^{m^*}u_h^*\,\lambda\left(\theta,y_{h-1}^*\right)-\sum_{h=1}^{m}u_h\,\lambda\left(\theta,y_{h-1}\right)\right]$$
$$\times\frac{\prod_{h=1}^{m^*}\lambda_i\left(\theta,y_{h-1}^*\right)\psi\left(\mathbf{S}_h^*\beta,j_h^*\right)}{\prod_{h=1}^{m}\lambda_i\left(\theta,y_{h-1}\right)\psi\left(\mathbf{S}_h\beta,j_h\right)}\Pi,$$

where Π is the ratio of proposal densities arising from drawing $(u_h^*,e_h^*)_{h=1}^{m^*}$ conditional on $(u_h,e_h)_{h=1}^{m}$ and

$$\begin{aligned}\psi\left(\mathbf{S}_h\beta,j_h\right) &= \int_{\mathbb{R}^{n-2}}\mathbf{1}\left\{\mathbf{v}\simeq y_{h-1},y_h\right\}\phi\left(\mathbf{v};\mathbf{z}_h\beta,1\right)d\mathbf{v}\\ &= \int_{-\infty}^{\infty}\left[\prod_{j\neq j_h,i_h}\Phi\left(r-\mathbf{s}_{i_h,j}\theta\right)\right]\phi\left(r;\mathbf{s}_{i_h,j_h}\beta,1\right)dr.\end{aligned}$$

The quantity ψ can easily be estimated to any degree of accuracy using Monte Carlo integration or importance sampling (in the calculations carried out on the data material a normal distribution shifted to the right by 1, $N(\mathbf{s}_{i_h,j_h}\beta+1,1)$, was used as the importance function, and to reduce variance an antithetical sampling scheme was used, with 1000 sampling points).

For drawing $\mathbf{v}_h$ given (u_h,e_h), ρ, and α, concordant with data, we employ the Gibbs sampling scheme described in McCulloch and Rossi (1994). A draw from the distribution $\mathbf{N}_{n-2}(\mathbf{z}_h\beta,\mathbf{\Sigma})$, is achieved by cycling through all the components and drawing values from the univariate conditional distributions. From standard statistical theory we have that $v_j|v_1,\ldots,v_{j-1},v_{j+1},\ldots,v_{n-2},\theta$ is $N(\mu_j,\eta_j^2)$,

$$\mu_j=z_j\beta+\mathbf{\Sigma}_{j(-j)}\mathbf{\Sigma}_{(-j)(-j)}^{-1}\left(\mathbf{v}_{h(-j)}-\mathbf{z}_{h(-j)}\beta\right),$$

and

$$\eta_j^2=\sigma_{jj}-\mathbf{\Sigma}_{j(-j)}\mathbf{\Sigma}_{(-j)(-j)}^{-1}\mathbf{\Sigma}_{(-j)j},$$

with the appropriate parts of a suitably partitioned $\mathbf{\Sigma}$, and where $\mathbf{v}_{h(-j)}$, and $\mathbf{z}_{h(-j)}$ are $\mathbf{v}_h$, and $\mathbf{z}_h$ respectively, with the jth rows removed. For the independent multinomial probit model, these quantities simplify to

$$\mu_j=z_j\beta+\left(1-(n-3)/(n-2)\right)\sum_{\ell\neq j}\left(v_\ell-z_\ell\beta\right),$$

and $\eta_j^2=3+(5-2n)/(n-2)$. In addition, each coordinate has to satisfy

$$\begin{cases} v_j^{(h)}<0 & \text{for } j\neq k_h \text{ if } j_h=k_h\\ v_j^{(h)}<v_{j_h}^{(h)}, \text{ and } 0<v_{j_h}^{(h)} & \text{for } j\neq k_h, \text{ if } j_h\neq k_h\end{cases},$$

which can be achieved by sampling from univariate truncated normal distributions. Our experience is that the Gibbs sampler need only be run for a few iterations.

The full conditional posterior of ρ has a particularly simple form. Define firstly

$$g_i(x,\alpha) = \{e^{\alpha} + \frac{s_{i1}(x)}{n-1}(e^{-\alpha} - e^{\alpha})\},$$

and note that the rate function is equal to $\lambda_i(x,\theta) = \rho g_i(x,\alpha)$, given state x, actor i and a fixed parameter α. Recall that $u_{m+1} = [t_1 - u_m]$, and the full conditional posterior of ρ, is proportional to the prior $\pi(\rho)$ times

$$\begin{aligned} & \exp\left(-\rho u_{m+1}\sum_i g_i(x(t_1),\alpha)\right) \\ & \times \prod_{h=1}^{m} \phi_{n-2}(\mathbf{v}_{h-1}; \mathbf{z}'_{h-1}\beta, \mathbf{I}_{n-2})\rho g_{i_h}(y_{h-1},\alpha) \exp\left(-u_h\rho\sum_i g_i(y_{h-1},\alpha)\right) \\ \propto \quad & \exp\left\{-\rho\left(\sum_{h=1}^{m+1}\sum_i u_h\, g_i(y_{h-1},\alpha)\right)\right\}\rho^m, \end{aligned}$$

which we recognise as the kernel of a Gamma distribution with parameters $m+1$ and

$$\frac{1}{\sum_{h=1}^{m+1}\sum_i u_h\, g_i(y_{h-1},\alpha)}.$$

The parameter α is updated using a Metropolis step. A suitable proposal density is a normal distribution centered over the present value. Note that in comparison with the logit specification where we had to set a candidate variance covariance matrix of size 6×6, the same model with the independent probit specification only requires that we determine a scalar. The full conditional posterior of α given all the rest is proportional to

$$\pi(\alpha)\exp\left\{-\rho\sum_{h=1}^{m+1} u_h \sum_i g_i(y_{h-1},\alpha)\right\}\prod_{h=1}^{m} g_{i_h}(y_{h-1},\alpha).$$

Finally there remains treating the move type (d). The main motivation behind using data augmentation in the frame work of the probit model (c.f. Albert and Chib, 1993; McCulloch and Rossi, 1994) is that regular Bayesian conjugacy theory applies once we have a realization of the latent variables. To simplify the analysis even further we employ a transformation suggested in McCulloch and Rossi (1994). Firstly, let C be the Cholesky root of the inverse variance covariance matrix, such that $\Sigma^{-1} = CC'$. Secondly, pre-multiply the regression equations by C', $C'\mathbf{v}_h = C'\mathbf{z}_h\beta + C'\varepsilon_h$ to obtain a system with standard normal errors $\mathbf{v}^*_h = \mathbf{v}^*_h\beta + \varepsilon^*_h$. Stacking the covariates in $\mathbf{Z}^* = (\mathbf{z}^{*\prime}_1, \ldots \mathbf{z}^{*\prime}_m)'$, the latent utility differences in $\mathbf{V}^* = (\mathbf{v}^{*\prime}_1, \ldots \mathbf{v}^{*\prime}_m)'$ and assuming that a priori $\beta \sim N(\mu, \mathbf{\Psi})$, the full conditional posterior of β given the rest is $N(\tilde{\mu}, \tilde{\mathbf{\Psi}})$, where

$$\tilde{\mathbf{\Psi}} = \left(\mathbf{Z}^{*\prime}\mathbf{Z}^* + \mathbf{\Psi}\right)^{-1}, \text{ and } \tilde{\mu} = \tilde{\mathbf{\Psi}}\left(\mathbf{Z}^{*\prime}\mathbf{V}^* + \mathbf{\Psi}\mu\right).$$

4.2.2. *Prior distributions.* As in CML specification, we have assumed that the parameters are mutually independent and vague a priori. In the previous section we noted however that there are conjugated prior distributions with respect to the full conditional posteriors, Gamma and Normal for ρ and β respectively, conditional on α. How to use this to set proper and informative priors is however not immediately clear, since the marginal posteriors are the marginalized joint posteriors marginalized with respect to the latent walks. There is the interesting possibility of training the prior on $\mathscr{W}_1(x(t_0), x)$ for all $x \in \bigcup_{i\in V} \mathscr{X}_i(x(t_0))$, that is an "imaginary " training sample, to obtain a proper prior. For convenience, the resulting normal mixture could then be approximated by a normal model.

Naturally, the multinomial probit model is not exempt from dangers of improper posteriors. Again, we use the ad-hoc criterion used for the multinomial logit model with the already mentioned caveats. We conjecture that $\max_{i\in V} |s_{i\ell}(x(t_0)) - s_{i\ell}(x(t_1))| > 0$ for all $\ell = 1, \ldots, p$, is a sufficient condition for proper posteriors.

4.3. **Posterior distributions.** The main objective of parameter inference is to asses what effects are supported by data. We do this primarily by inspecting the marginal posterior credibility intervals of the parameters. Because of the dependencies between networks statistics (in the model), some effects should be included regardless of whether they are close to zeros with high posterior probability. An example of this is the out-degree effect, which serves as a control for other statistics as well as an effect in its own right. This is further discussed in connection with Figures 6, 7,and 8.

The increased efficiency when sampling directly from the full conditional posterior of β is reflected in the sample auto correlations of the MCMC sample (see Figure 4). Whereas the autocorrelation decrease very slowly for the logit parameters, the autocorrelations drop rather more sharply for the probit parameters. These relatively high auto correlations for the logit model are not artifacts of too low an average jump distance, and thereby too high an acceptance probability, rather this is a remnant of the dependencies stemming from the latent paths. This is reflected in that "non–structural" parameters β_3 and β_4 display more well-behaved sample auto correlation functions than the structural parameters. It is important to maintain a good balance between the acceptance probability of the latent paths and parameters and in general it seems that the acceptance rate of the latent walks are negatively related to the acceptance rate of the parameters $\beta_1, \ldots, \beta_p$. The sample upon which Figure 4 is based had an acceptance rate of .48 for the latent walks and .23 for the parameters with the logit specification. For the probit sample the acceptance rates were .66 and .28 (i.e. for α). The scaling constant γ for the proposal variance was set to $.9/\sqrt{6}$ and .7 for the logit and probit parameters respectively.

The auto correlations reduce the value of many convergence statistics, since there is often an implicit independence assumption. Visual inspection of trace plots and the use of time series tools, in combination with over–dispersed starting points and multiple chains, are recommended for convergence assessment. Figure 2 displays the difference in the cumulative and batched means for two independent

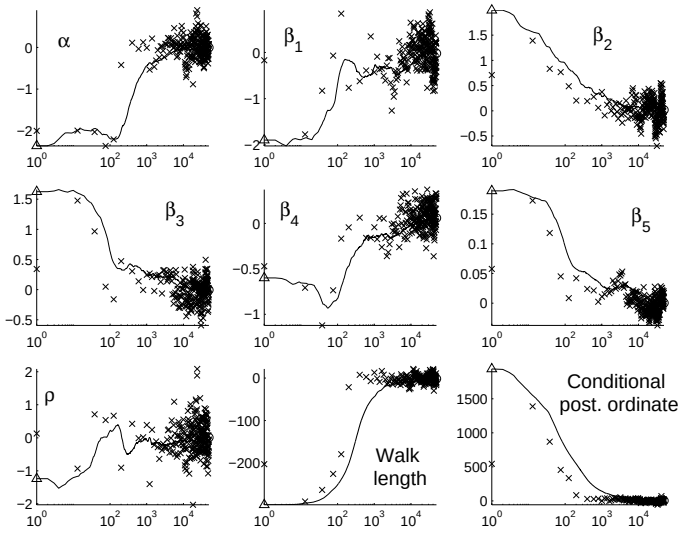

FIGURE 2. Differences in cumulative (–) and batched means (×) for two independent MCMC samples from logit specification. Initial values and grand means given by (Δ) and (∘) respectively

MCMC-samples for the logit parameters. One of the chains is started in a region of the parameter space that has relatively high posterior mass whereas the other is started in a region with low posterior mass. For all trace plots, the batched mean seems to have stabilized round 0 somewhere between iteration 100 and 1000. The differences in cumulative means lags behind because of the high initial differences but has caught up around the $10{,}000^{th}$ iteration. An alternative representation is to inspect the standardized output (Geweke, 1992) an example of which is given in Figure 3 for a typical MCMC-sample. The bounds should be treated very leniently since the variance estimates (spectral density estimators with Daniell windows, $M = 2T^{1/2}$) do not take the cross correlations into consideration nor the dependence on the latent walks.

Figure 5 gives the marginal posteriors[1] of the parameters for the logit and probit specification. For comparison, the Method of moments estimates (as obtained from the stochastic approximation algorithm described in Snijders and van Duijn, 1997), with approximate confidence intervals are included. It is interesting to note that the point estimates for the logit specification agree with that of the method of moments bar for α. The distribution of the statistic $\sum_{e \in \mathscr{N}} |X_e(t_0) - X_e(t_1)|$, the expected value of which is used in the moment equations, is relatively flat and not particularly sensitive to changes in α. This is not picked up by the stochastic approximation algorithm, and the 95% confidece interval just covers zero.

[1]All univariate density are estimated using normal kernel estimators with window width $.06an^{-1/5}$, where n is the number of points in the sample and a is the minimum of the sample variance and the interquartile range divided by 1.34. Multivariate densities are estimated with multivariate normal kernels and the pre-whitening method of Fukunaga (1972)

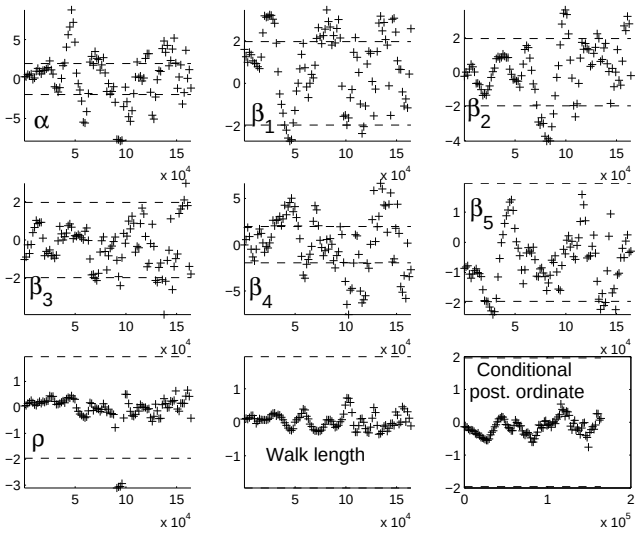

FIGURE 3. Geweke convergence diagnostics (+) with 95%- bounds (dashed lines)

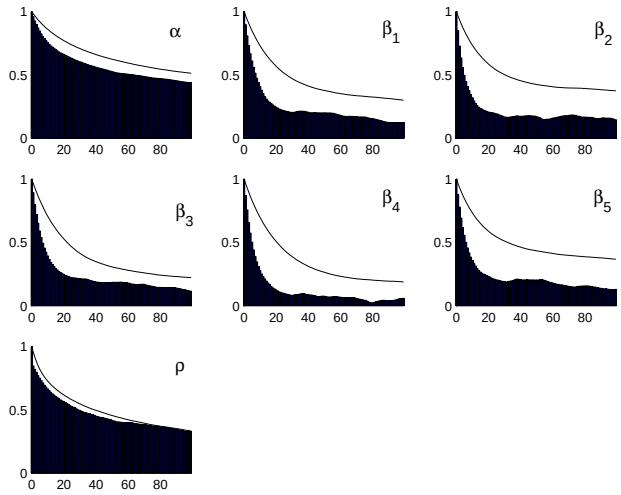

FIGURE 4. Comparing the sample auto correlation functions for MCMC samples from posteriors in logit (lines) and probit (bars)

In general the posteriors are fairly conclusive and easy to motivate: actors tend to form ties to people from whom they already receive ties; to people with low citation rate; and to people that are connected to someone they know beforehand. That the similarity with respect to citation rates is a factor in forming ties to others is not supported by data.

This brings us to the issue of interpretation of parameters and their relative magnitudes. From Figure 5 (logit specification), it is clear that β_1 is negative and β_5 is positive with high posterior probability. So, how should this be interpreted

seeing as in the transition from one network configuration to another, the transitivity count (s_{i5}) can not increase without the out-degree count (s_{i1}) increasing. To investigate this we employ a situation analysis framed by the model ingredients.

Consider the subgraph $x^{(0)}$, in Figure 6, consisting of vertices i, j, and v. Assume further that actor j may make a change to the composition of his out-going ties. The courses of action involving the other two actors in the sub graph, i and v, consists of removing the arc to v – in which case we would obtain the subgraph $y^{(0)}$ – or adding an arc to i – which would produce the subgraph $z^{(0)}$. Now, under the additional assumption that w_i, w_j, and w_v, all are zero, given the fitted model, the only relevant difference in the overall contribution to the statistics is a -1 and $+1$, associated with going from $x^{(0)}$ to $y^{(0)}$ and $z^{(0)}$, respectively. Because of IIA, the odds for actor j choosing to remove the arc (j, v) relative to adding the arc (j, i) is $e^{-2\beta_1}$. For the subgraph $x^{(1)}$, the situation is different, since there exists a two-path from j to i via ℓ_1, and forming the tie (j, i) would create a transitive triple. Hence, the odds for actor j choosing to remove the arc (j, v) relative to adding the arc (j, i) is $e^{-2\beta_1-\beta_5}$, given $x^{(1)}$ (under the assumption that $w_{\ell_1} = 0$). In general, given configuration $x^{(k)}$ as in Figure 6, the odds for removing the edge (j, v) (a transition to $y^{(k)}$) relative to adding (j, i) (a transition to $z^{(k)}$ with a k–fold triple) is $e^{-2\beta_1-k\beta_5}$. The posterior predictive odds for $k = 1, \ldots, 7$ are given in Figure 7. Naturally, there is a clear tendency for the support of $y^{(k)}$ to decrease with k, which is further illustrated in the left panel of Figure 8. Interesting to note is the differences when conditioning on the number of intermediate changes. It appears that the influence of transitivity relative to the influence of (the negative) out-degree decreases the more changes there have been.

Had there in Figure 6 also been an arc from i to j, adding the edge (j, i), would have created a reciprocated dyad. In that case, the odds for choosing $y^{(k)}$ relative to $z^{(k)}$ would have been $e^{-2\beta_1-\beta_2-k\beta_5}$. The posterior predictive expectancy of the odds of $y^{(k)}$ relative to $z^{(k)}$ in this case is illustrated in the right panel of Figure 8 for different k. When compared to adding an arc without producing a reciprocated relation, there is a similar ordering of the predictive probabilities with respect to the number of intermediate changes. The desirability of adding the arc increases considerably more quickly if it produces a reciprocated tie as well as a k–fold triple, seen as a function of k.

The results for the logit and probit specifications are to all intents and purposes the same with some differences in scale, with the exception for β_3 (see Figure 5 and Table 1). The marginal posterior probability that β_1 is negative is .99, to be compared with .003 for the logit specification, whereas the posterior probability that β_3 is greater than 0 is .94, compared to .9987 for the logit specification. Any suspicions that high posterior correlation between β_3 and β_4 might conceal the fact than one or other or both of the parameters actually plays a part are dispelled when seeing that the origin is contained in the HPD-ellipses in Figure 10. Differences might depend upon the more forgiving tails of the Gumbel distribution as compared to the Gaussian. As an example consider Figure 9. In the context of the network evolution process we see that the logit model is more willing to accept changes to dyads for which the expected utility is relatively lower than

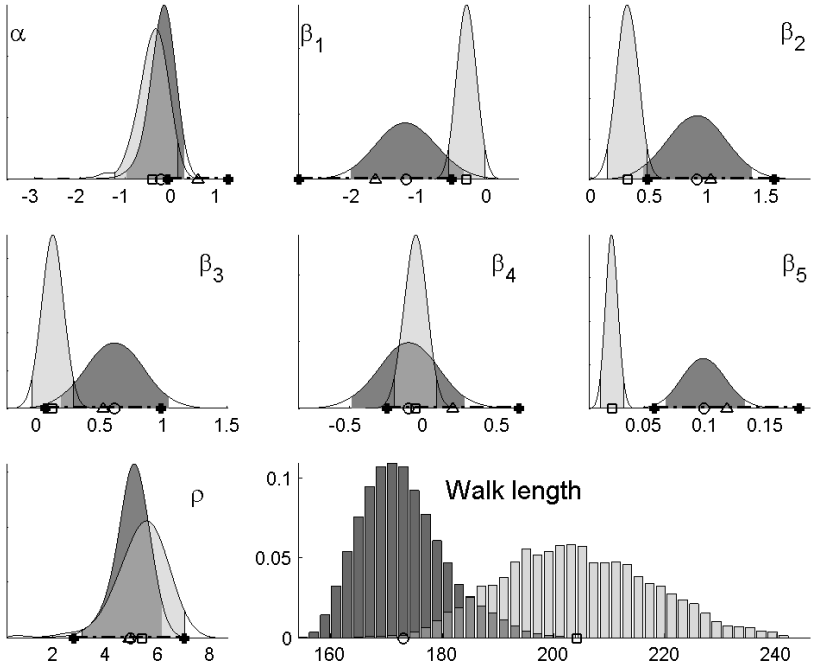

FIGURE 5. Marginal posterior densities for logit (probit) parameters, with 95% Credibility intervals, darker (lighter) shades, and posterior means ∘ (□); MM estimate (△) with approximate 95% Confidence intervals $(+ - \cdot - +)$

	MM	SE	Bayes logit	STD	Bayes probit	STD
α, activity in rate	.58	.333	−.23	.384	−.41	.341
β_1, activity	−1.70	.582	−1.23	.434	−.31	.139
β_2, mutuality	1.02	.270	.90	.238	.31	.086
β_3, low citation	.52	.228	.61	.214	.13	.082
β_4, citation-similarity	.20	.223	−.10	.190	−.06	.072
β_5, transitivity	.12	.029	.098	.0163	.024	.0047
ρ, constant in rate	4.89	1.055	4.94	.81	5.37	.97

TABLE 1. Point estimates for parameters in models fitted to dichotomized EIES data

in the case of the probit model. This to some extent explains why the probit posteriors (i.e. posteriors for β) are closer to zero with smaller posterior variance than the logit posteriors. This would also explain the higher posterior expected walk length for the probit model since the order of changes are more important in the probit model.

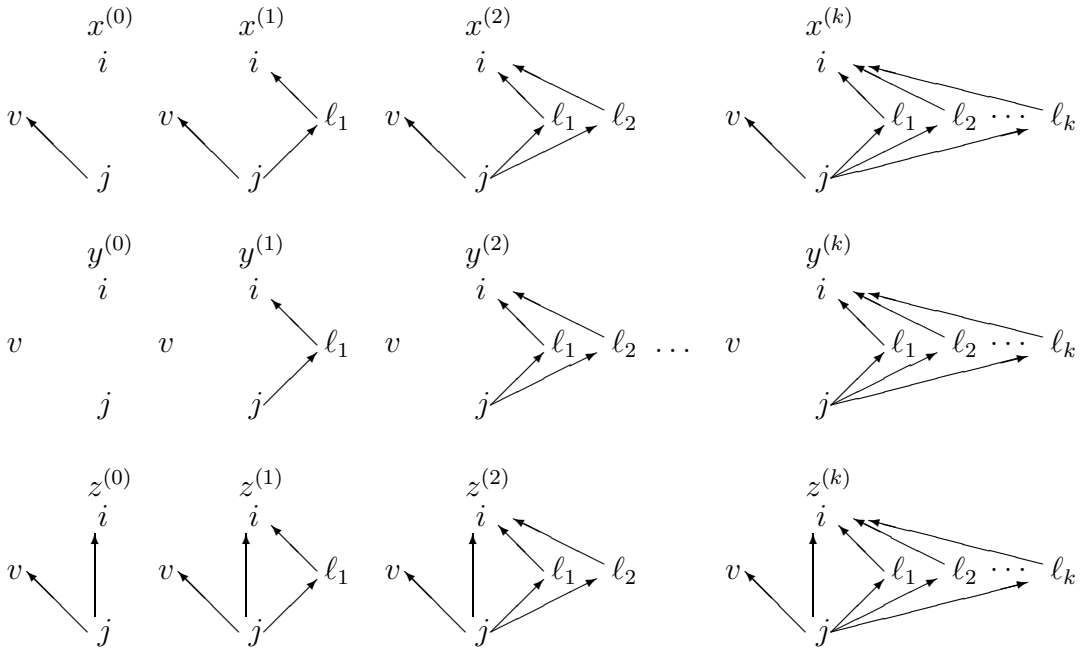

FIGURE 6. Changing $x^{(k)}$ by either removing the edge from j to v, $y^{(k)}$, or forming the edge from j to i, $z^{(k)}$

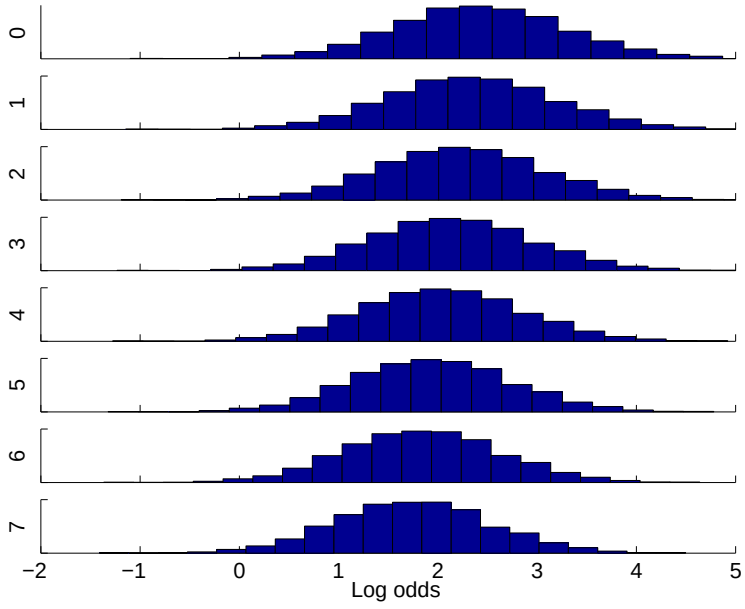

FIGURE 7. The posterior predictive log odds of a transition to $y^{(k)}$ relative to $z^{(k)}$, given $x^{(k)}$

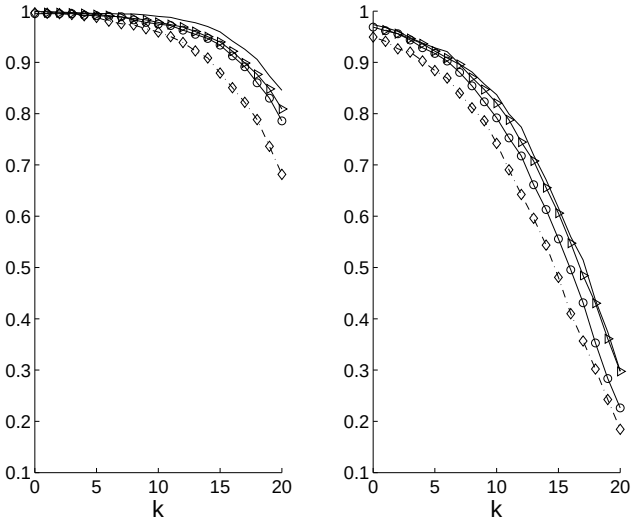

FIGURE 8. The posterior predictive probaility that the odds for $y^{(k)}$ relative to $z^{(k)}$, given $x^{(k)}$, is greater than 1 for $m < 171$ ($- \cdot \diamond - \cdot$), $m \geq 171$ and $m < 177$ ($-\circ-$), $m \geq 177$ (—-), overall ($-\triangleright-$). Case arc (j, i) is not reciprocated (left panel) and case arc (j, i) is reciprocated (right panel)

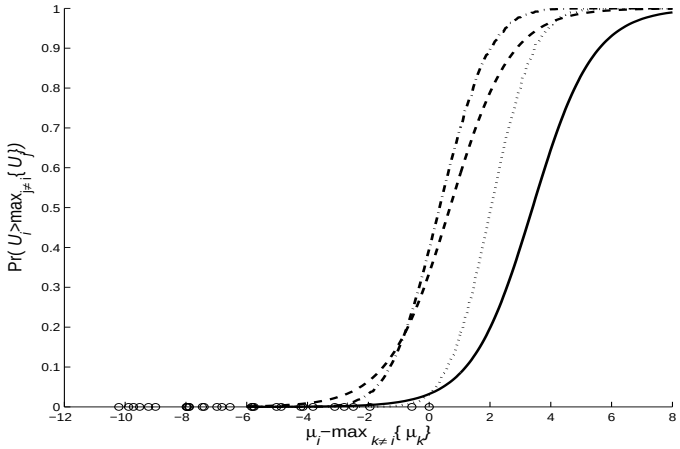

FIGURE 9. Multinomial probabilities. $U_j = \mu_j + \epsilon_j$ for $j = 1, \ldots, 32$, differences $\mu_j - \max_{k \neq i}\{\mu_k\}$ ($\circ$), ϵ_j are i.i.d. Gumbel ($- -$), and i.i.d. standard Normal ($-\cdot$). Homogeneous means, $\mu_j = \mu$ for $j \neq i$ where ϵ_j are i.i.d. Gumbel ($-$), and i.i.d. standard Normal ($\cdots$).

4.4. **Valued relations.** The original data had the ordinal labels in $\mathscr{R} = \{0, 1, 2, 3, 4\}$. To model the evolution with valued arcs, we use the model described in section 2.2.1. with a few minor modifications. Firstly, define the adjacency matrix $y = x(i \overset{a}{\rightsquigarrow} j)$ which differ from x in exactly $y_{ij} = x_{ij} + a$. Secondly, an actor i allowed to make a change given x, instead of only evaluating the graphs in $\mathscr{X}_i(x)$

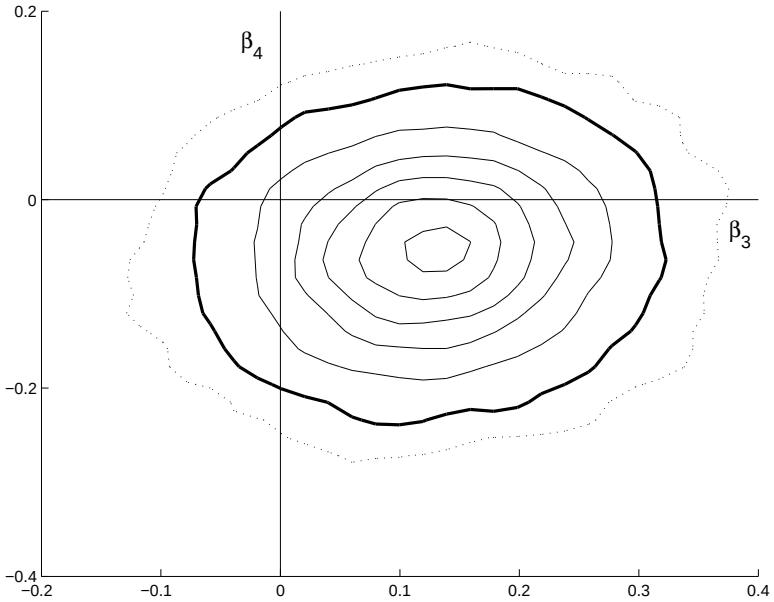

FIGURE 10. Contour plots of the bivariate posterior of β_3 and β_4 from the probit model, with HPD-elipses: 95% (thick line), 99% (dotted)

as defined in section 2.1.1., evaluates all graphs in the more generally defined $\mathscr{X}_i(x) = \{y \in \mathscr{X} : y = x(i \overset{a}{\rightsquigarrow} j) \text{ for some } j \in V \text{ and } a \in \{-1, 1\}\}$. The utility is then written

$$U_i(t, x, j, a) = r(\theta, i, j, a, x) + \epsilon_i(t, x, j, a).$$

In the empirical example we limit the analysis to the case when $\epsilon_i(t, x, j, a)$ are independently and identically distributed according to the extreme value type one distribution, for each t, x, and j and a such that $x(i \overset{a}{\rightsquigarrow} j) \in \mathscr{X}$. The form of the resulting conditional multinomial jump probabilities is then more or less identical to Eq. 2.6

$$\frac{e^{r(\theta,i,j,x,a)}}{\sum e^{r(\theta,i,k,x,a)}},$$

where the sum is taken over $\{k, a\} \in V \times \{-1, 1\} : x(i \overset{a}{\rightsquigarrow} k) \in \mathscr{X}$.

For the sake of comparison, we begin by assuming a similar linear form for the systematic component as in 4.1, using valued counterparts of the statistics used in the previous analysis. A natural extension of the degree count is the out-degree weighted by the strength of the relations

$$s_{i1}(x) = \sum_{j:(i,j)\in\mathscr{N}} x_{ij}.$$

Since reciprocity measures the distance between i and j with respect to x_{ij} and x_{ji}, valued reciprocity can be formulated in terms of distance

$$s_{i2}(x) = \sum_{j:(i,j)\in\mathscr{N}} (x_{ij} - x_{ji})^2 .$$

The interpretation of the distance is different than the interpretation of reciprocity, and the more unbalanced a dyad is, the more it contributes to the utility. Formulating transitivity with valued relations could be done in various ways. One interpretation is that an actor want his friend to have at least as strong ties as he has to them, leading to statistics based upon e.g. $(x_{ij} - x_{jh})^2 + (x_{ih} - x_{jh})^2$, for the triple (i, j, h). A convenient specification is simply saying that the strengths of the ties (i, j) and (i, h) are multiplied by a factor x_{jh}. The valued transitivity count then becomes

$$s_{i5}(x) = \sum_{j,h} x_{ij}x_{ih}x_{jh}.$$

We further use the same covariate counts as in the dichotomized example, $s_{i3}(x)$ and $s_{i4}(x)$.

In addition, to take further advantage of the increased detail, we introduce four statistics to be included in what Snijders (2001)(defined there only for binary data) calls the gratification function. To take into account the costs of breaking off already established relations, we define for $k = 1, \ldots, 4$,

$$g_{ik}(x, j, a) = \mathbf{1}\{a = -1\}\mathbf{1}\{x_{ij} = k\},$$

i.e. $g_{i1}(x, j, a)$ is equal to 1 if $x_{ij} = 1$ and $a = -1$, that is a transition from $x_{ij} = 1$ to $x_{ij} = 0$.

For valued data we fit two models, one where

$$r\left(\theta, i, j, a, x\right) = (s_{i1}(x(i \overset{a}{\rightsquigarrow} j)), \ldots, s_{i5}(x(i \overset{a}{\rightsquigarrow} j)))\beta,$$

and

$$r\left(\theta, i, j, a, x\right) = (s_{i1}(x(i \overset{a}{\rightsquigarrow} j)), \ldots, s_{i5}(x(i \overset{a}{\rightsquigarrow} j)), g_{i1}(x, j, a), \ldots, g_{i4}(x, j, a))\beta,$$

where β are 5×1 and 9×1 vectors respectively.

The MCMC algorithm as described for the dichotomous data in Section 4.1. applies with the addition of drawing $(a_h)_{h=1}^m$ according to Section 3.2.

4.5. **Posterior distributions for valued data.** The posterior means and standard deviations for the parameters in the models fitted to the valued data are given in Table 2, and in Figures 11, and 12. The first model (with 7 parameters) gives conclusions similar to those for dichotomized data except for the out-degree effect. When the strengths of relations are taken into account, there seems to be a tendency (marginally) for seeking strong ties to others. This conclusion is however somewhat week, as can be seen from the 96% Credibility interval in Figure 11, which covers the origin. There also seems to be a tendency for actors to strive for "equality" in their relations. The sign of β_2 suggests that it is unattractive to have discrepancies in the dyads, and that actors wants to match the strengths of their out-going ties to the strengths of their in-going ties. For the model with

	mean	STD	mean	STD
α, activity in rate	$-.23$	.24	$-.17$	.55
β_1, activity	.34	.17	-5.40	1.65
β_2, relation dissimilarity	$-.30$	.03	$-.29$	.03
β_3, low citation	.33	.10	.32	.10
β_4, citation dissimilarity	.14	.100	.08	.099
β_5, transitivity	.0024	.00097	.0033	.00105
β_6, loss of strength 1			-11.0	3.3
β_7, loss of strength 2			-12.5	3.3
β_8, loss of strength 3			-10.0	3.3
β_9, loss of strength 4			-9.9	3.4
ρ, constant in rate	3.26	0.21	3.41	0.60

TABLE 2. Posterior means and standard deviations for models fitted to original data values for the EIES data

effects added for breaking of ties with different strengths, the conclusions are again similar to the dichotomized case (Table 2, and Figure 12). As one would expect, if a change to an actors composition means that the actor has to decrease the strength of an established relation, this contributes negatively to the attractivity of that configuration. These effect has however to be interpreted in relation to the overall out-degree effect, which for this model has become negative.

The performance of the algorithm for the valued data is comparable to the performance for the dichotomized data. The acceptance rates for the 7-parameter (11-parameter) model were .61 (.49) and .22 (.13) for the latent walks and parameters respectively. The cardinality of the space of all possible walk is however considerably greater. Whereas the minimum number of changes for the dichotomized data was 154, for the original data values this becomes 371. Possibly because of high posterior correlation between the degree related parameters, β_1 and β_6 through β_9, the 11-parameter model takes a long time to settle. Interestingly, the stability of the algorithm is unaffected by the long excursions taken by the degree related parameters. It is instructive to study the trace plots in Figure 13. The Figure plots the deviations from the grand mean (as estimated for every 50'th iteration with the first 80,000 iterations discarded) in covariance norm (covariance estimated from same subsample as the mean). Comparing the top two panels with the bottom panel, we can state that the algorithm is remarkably stable for α, ρ and β_2 through β_5 and the conditional posterior ordinate considering that β_6 move as slowly in the parameter space as it does.

5. Concluding remarks

In this paper we have shown an alternative to the methods of moments estimators for longitudinal social network data. The main benefit is the more precise measure of uncertainty in the inference procedure that comes with the posterior distributions. We were also able to extend the class of models for longitudinal social networks both with respect to the mechanisms that drive network evolution

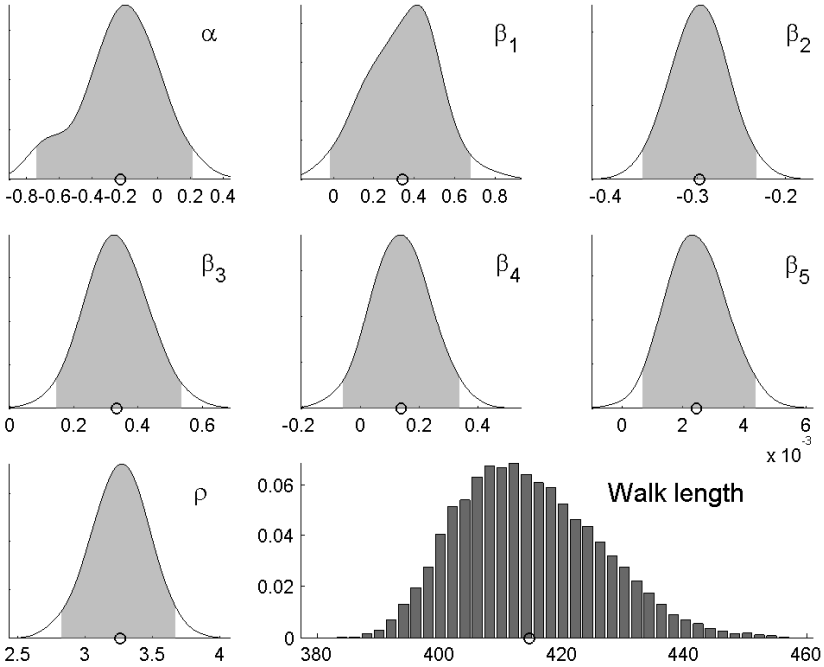

FIGURE 11. Marginal posterior densities for valued data with effects s_{i1} to s_{i5}, with 95% Credibility intervals, and posterior means ○

as well as the actual type of data. This was accomplished by focusing on the key components of the embedded chain rather than defining the model in terms of the generator matrix and the transition probabilities. The embedded chain also provides a natural way for defining latent variables with which observed data is augmented for inference purposes. Data augmentation ensures that flexibility of the model framework does not hamper the ability to conduct inference and enables exact Bayesian inference – with bonuses such as the analysis of intermediate paths; posterior odds; various conditionings; etc.

Although the inference procedure entails augmentation with data with a high dimensional support with many combinatorial quirks, the algorithms seems to work very well. Naturally, simultaneous sampling of parameters and latent variables has as a result that the chain moves slowly in the parameter space (and more so for the logit specification). This could however be considered a low price to pay. The number of evaluations of the complete data likelihood for the algorithm is in general of the same order of magnitude as for the MCMC implementation of the method of moments. The latter also relies on sampling from the sample space but the sampling is not made conditional on data in the same manner as is done here and as a consequence the precision is considerably smaller.

The advantages of analysing valued data rather than dichotomous data, when possible, were discussed. The analysis on valued data that was actually carried

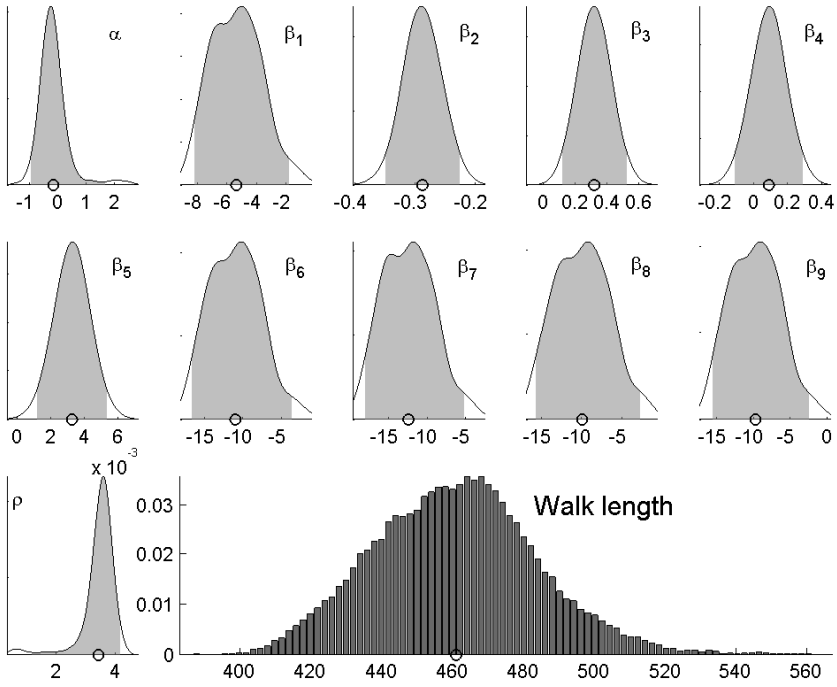

FIGURE 12. Marginal posterior densities for valued data with effects s_{i1} to s_{i9}, with 95% Credibility intervals, and posterior means ○

out mostly served the purpose of illustrating the possibilities. It turns out that the computational complexity of valued data is not much worse than for binary data. Much research is required in this area for assessing how the relative wealth of information contained in valued longitudinal social network data is most efficiently modeled.

There are some necessary future extensions to be done, notably the incorporation of changing composition (Huisman and Snijders, 2003) and changing behaviour. Bayesian inference is particularly suited to dealing with the former and in contrast to the ad-hoc imputation of Huisman and Snijders (2003), missing observations can be treated on a modeling basis. Missing observations enter the inference scheme as part of the set of latent variables. Especially interesting is the combination of missing values and correlated normal random components, and missing values and co-offending networks. Offenders not appearing in the police files can hardly be said to be missing at random (see Sarnecki, 1999, on the *dark figure* in criminal networks).

For the model framework as it has been presented here, more work is needed on interpreting effects, their magnitudes, analytical properties, etc (some work on these issues in the context of stochastic actor-oriented models can be found in Snijders 2001 and Snijders 2004). This is especially essential for setting prior distributions. In this paper we have only worked with vague, "non–informative",

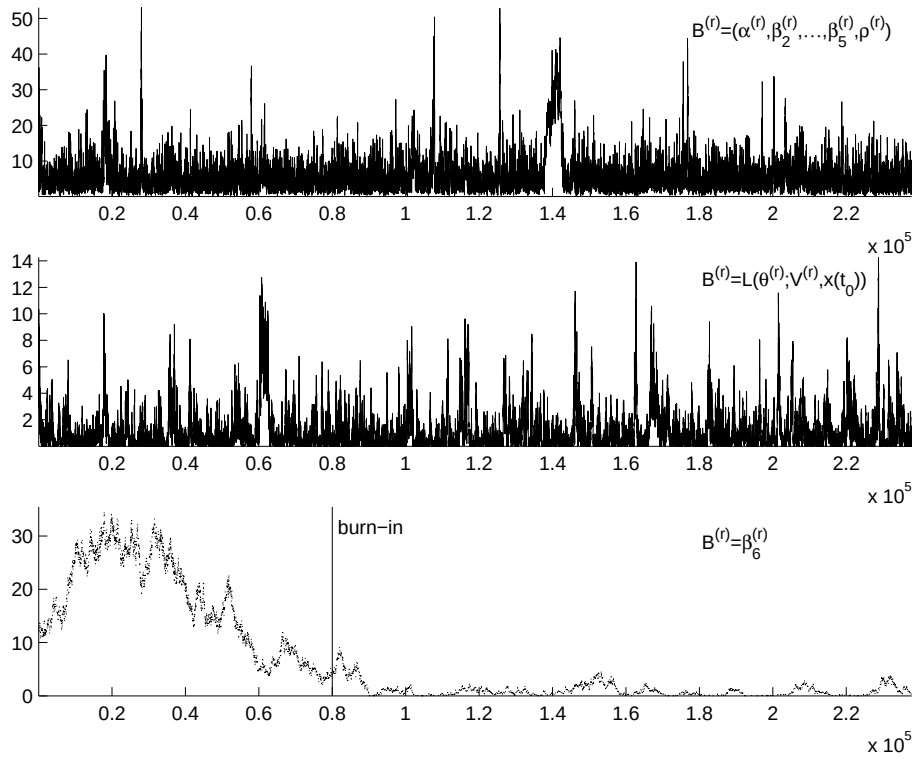

FIGURE 13. Trace plots of $||B^{(r)} - \bar{B}||_{\Sigma_B^{-1}}$ for model with s_{i1} through s_{i9}

prior distribution. There is a need for investigating prior distributions, both the a priori analysis - quantifying prior information in a useful way - and the posterior analysis - sensitivity analysis. Not only is this interesting because of the advantages that comes with a fully Bayesian analysis but also because of the necessity of using proper priors in the model selection process. Model selection is the next logical step after having put forward a scheme for parameter inference.

REFERENCES

Albert, J. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88, 669–679.

Auranen, K., Arjas, E., Leino, T., and Takala, A. K. (2000). "Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data." *Journal of the American Statistical Association*, 95, 1044–1053.

Banks, D. L. and Carley, K. M. (1996). "Models for network evolution." *Journal of Mathematical Sociology*, 21, 173–193. Also published in Doreian and Stockman (1997).

Bearman, P. S., Moody, J., and Stovel, K. (2004). "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology*. Forthcoming, July: Vol. 110.N1.

Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discusion)." *Journal of the Royal Statistical Society*, 36, 192–236.

— (2000). "Markov chain Monte Carlo for statistical inference." working paper, University of Washington, Center for Statistics and the Social Sciences.

Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis Theory and Practice*. Cambridge, Mass.: MIT Press.

Burt, R. S. (1976). "Positions in Networks." *Social Forces*, 55, 93–122.

— (1992). *Structural Holes: The social structure of competition*. Cambridge: Harvard University Press.

Chib, S. (1995). "Understanding the Metropolis–Hastings algorithm." *The American Statistician*, 49, 327–335.

Corander, J., Dahmström, K., and Dahmström, P. (1998). "Maximum likelihood estimation for Markov graphs." Research report 1998:8, Stockholm University, Department of Statistics.

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

Eerola, M., Gasbarra, D., Mäkelä, H. P., Linden, H., and Andreev, A. (2003). "Joint modelling of recurrent infections and antibody response by Bayesian data augmentation." *Scandinavian journal of statistics*, 30, 677–698.

Fienberg, S. E. and Wasserman, S. (1981). "Categorical data analysis of songle sociometric relations." *Sociological Methodology*, 156–192.

Frank, O. (2001). "Statistical estimation of co-offending youth networks." *Social Networks*, 23, 203–214.

— (2004). "Network sampling and model fitting." To appear in P. Carrington, J. Scott, & S. Wasserman (Eds.), Models and methods in social network analysis. New York: Cambridge University Press, 2004.

Frank, O. and Nowicki, K. (1993). "Exploratory statistical analysis of networks." In *Quo vadis, graph theory?*, vol. 55 of *Ann. Discrete Math.*, 349–365. Amsterdam: North-Holland.

Frank, O. and Strauss, D. (1986). "Markov Graphs." *Journal of the American Statistical Association*, 81, 832–842.

Fray, R. D. and Roselle, D. P. (1971). "Weighted lattice paths." *Pacific Journal of Mathematics*, 37, 85–96.

Freeman, S. C. and Freeman, L. C. (1979). "The networkers network: A study of the impact of a new communications medium on sociometric structure." Social Science Research Reports 46, University of California, Irvine CA.

Fukunaga, K., ed. (1972). *Introduction to Statistical Pattern Recognition*. New York: Academic Press.

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In *Bayesian statistics, 4 (Peñíscola, 1991)*, 169–193. New York: Oxford Univ. Press.

Gibson, G. J. (1997). "Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology." *Journal of the Royal Statistical Society*, 46, 215–233.

Gibson, G. J. and Renshaw, E. (1998). "Estimating parameters in stochastic compartmental models using Markov chain methods." *IMA Journal of Mathematics applied in Medicine and Biology*, 15, 19–40.

Gilks, W., Richardson, S., and Spiegelhalter, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82, 711–732.

Griffiths, W. E., Hill, R. C., and Pope, P. J. (1987). "Small sample properties of probit model estimators." *Journal of the American Statistical Association*, 82, 929–937.

Hancock, M. S. (2000). "Progress in statistical modeling of drug user and sexual networks."

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

Holland, P., Laskey, K. B., and Leinhardt, S. (1983). "Stochastic block-models: Some first steps." *Social Networks*, 5, 109–137.

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

— (1977b). "Social structure as a network process." *Zeitschrift für Soziologie*, 6, 386–402.

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

Kalbfleisch, J. D. and Lawless, J. F. (1985). "The analysis of panel data under a Markov assumption." *Journal of the American Statistical Association*, 80, 863–871.

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

Krackhardt, D. (1999). "The ties that torture: Simmelian tie analysis in organizations." *Research in the Sociology of Organisations*, 16, 183–210.

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

— (1995b). "Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure and their Interdependence." Ph.D. thesis, Amsterdam.

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

McCulloch, R. and Rossi, P. E. (1994). "An exact likelihood analysis of the multinomial probit model." *Journal of Econometrics*, 64, 207–240.

McFadden, D. (1974). "Conditional logit analysis of qualitative chooice behavior." In *Frontiers in Econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equations of state calculations by fast computing machine." *Journal*

of Chemical Physics, 21, 1087–1091.

Moreno, J. L. (1934). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Washington, D.C.: Nervous and Mental Disease Publishing Co. Reprinted in 1978 (Third Edition) by Becon House, inc., Beacon, NY.

Nowicki, K. and Snijders, T. A. B. (2001). "Estimation and prediction for stochastic blockstructures." *Journal of the American Statistical Association*, 96, 1077–1087.

O'Neill, P. and Roberts, G. (1999). "Bayesian inference for partially observed stochastic epidemics." *Journal of the Royal Statistical Society*, 162, 121–129.

Pattison, P. and Robins, G. (2002). "Neighbourhood-based models for social networks." *Sociological Methodology*, 32, 301–337.

Reis, A. J. and Farrington, D. P. (1991). "Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males." *Journal of Criminal Law and Criminology*, 82, 360–395.

Reiss, A. (1988). "Co-offending and criminal careers." In *Crime and Justice: An Annual Review of Research*, ed. N. Tonry, M. Morris. Chicago: University of Chicago Press.

Reiss, A. J. (1986). "Co-offender influences on criminal careers." In *Criminal Careers and "Career Criminals", vol. II*, eds. A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher. Washington, D.C.: National Academy Press.

Richardson, S. and Green, P. J. (1997). "On Bayesian analysis of mixtures with an unknown number of components (with discussion)." *Journal of the Royal Statistical Society*, 59, 731–792.

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

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms." *Ann. Appl. Probab.*, 7, 1, 110–120.

Robins, G. (1998). "Personal Attributes in Interpersonal Contexts: Statistical Models for Individual Characteristics and Social Relationships." Ph.d. dissertation, University of Melbourne, Department of Psychology.

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

— (2004). "Interdependencies and social processes: dependence graphs and generalized dependence structures." Forthcoming in Models and Methods in Social Network Analysis, P. J. Carrington and J. Scott and S. Wasserman (eds), New York: Cambridge University Press.

Robins, G., Pattison, P., and Wasserman, S. (1999). "Logit models and logistic regression for social networks: III. Values relations." *Psychometrika*, 64, 371–394.

Sanil, A., Banks, D., and Carley, K. (1995). "Models for evolving fixed node networks: Model fitting and model testing." *Social Networks*, 17, 65–81.

Sarnecki, J. (1999). "Co-offending youth networks in Stockholm." Working paper, Department of Criminology, Stockholm University.

Schweinberger, M. and Snijders, T. A. B. (2003). "Settings in social networks: A measurement model." *Sociological Methodology*, 33, 307–341.

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

— (1999). "The transition probabilities of the reciprocity model." *Journal of Mathematical Sociology*, 23, 241–253.

— (2001). "The statistical evaluation of social network dynamics." *Sociological Methodology*, 30, 361–395.

— (2002). "Markov chain Monte Carlo estimation of exponential random graph models." *Journal of Social Structure*, 3, 2.

— (2004). "Models for longitudinal network data." To appear as Chapter 11 in P. Carrington, J. Scott, and S. Wasserman (Eds.), Models and methods in social network analysis. New York: Cambridge University Press, 2004.

Snijders, T. A. B. and Nowicki, K. (1997). "Estimation and prediction for stochastic blockmodels for graphs with latent block structure." *Journal of Classification*, 14, 75–100.

Snijders, T. A. B. and van Duijn, M. A. J. (1997). "Simulation for statistical inference in dynamic network models." In *Simulating social phenomena*, eds. R. Conte, R. Hegselmann, and P. Terna, 493–512. Berlin: Springer.

Strauss, D. and Ikeda, M. (1990). "Pseudolikelihood estimation for social networks." *Journal of the American Statistical Association*, 85, 204–212.

Tallberg, C. (2004). "A Bayesian approach to modeling stochastic blockstructures with covariates." To appear in Journal of Mathematical Sociology.

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

Tierney, L. (1994). "Markov chains for exploring posterior distributions." *Ann. Statist.*, 22, 4, 1701–1762. With discussion and a rejoinder by the author.

Wasserman, S. (1977). "Stochastic Models for Directed Graphs." Ph.d. dissertation, University of Harvard, Department of Statistics.

— (1980a). "Analyzing social networks as stochastic processes." *Journal of the American Statistical Association*, 75, 280–294.

— (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.*. New York and Cambridge: Cambridge University Press.

Wasserman, S. and Iacobucci, D. (1988). "Sequential social network analysis." *Psychometrika*, 53, 261–282.

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

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

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden
E-mail address: johan.koskinen@stat.su.se