Bayesian inference for dynamic social network data

Johan H. Koskinen[a,b,*,1], Tom A.B. Snijders[c,d]

[a]*Department of Statistics, Stockholm University, Stockholm, Sweden*
[b]*Department of Psychology, University of Melbourne, Parkville VIC 3010, Australia*
[c]*University of Oxford, Nuffield College, New Road, Oxford OX1 1NF, UK*
[d]*Department of Sociology, University of Groningen, The Netherlands*

Abstract

We consider a continuous-time model for the evolution of social networks. A social network is here conceived as a (di-) graph on a set of vertices, representing actors, and the changes of interest are creation and disappearance over time of (arcs) edges in the graph. Hence we model a collection of random edge indicators that are not, in general, independent. We explicitly model the interdependencies between edge indicators that arise from interaction between social entities. A Markov chain is defined in terms of an embedded chain with holding times and transition probabilities. Data are observed at fixed points in time and hence we are not able to observe the embedded chain directly. Introducing a prior distribution for the parameters we may implement an MCMC algorithm for exploring the posterior distribution of the parameters by simulating the evolution of the embedded process between observations.
© 2007 Elsevier B.V. All rights reserved.

1. Introduction

The methods of social network analysis (e.g., Wasserman and Faust, 1994; Carrington et al., 2005) have long been employed in studying social phenomena ranging from friendship patterns of children in school classes (Moreno, 1934) to the spread of HIV (Klovdahl, 1985). The importance of studying the dynamics of social networks is generally stressed and there are some obvious shortcomings of cross-sectional data. Snijders (2001), proposed a statistical model for the analysis of longitudinal social network data (for empirical applications see e.g. van de Bunt et al., 1999, and van Duijn et al., 2003). Inference has so far been limited to an MCMC implementation of the method of moments, the properties of which (i.e. the estimators) are not fully understood. Here we propose a Bayesian inference scheme for these models as well as a model formulation that makes provisions for relaxing some of the assumptions and that allows for a wider scope for the type of data that may be analysed.

---

* Corresponding author. Department of Psychology, University of Melbourne, Parkville VIC 3010, Australia. Tel.: +61 03 834 44768; fax: +61 03 93 47 6618.

E-mail addresses: johank@unimelb.edu.au (J.H. Koskinen), tom.snijders@nuffield.ox.ac.uk (T.A.B. Snijders).

2. Preliminaries

In this paper, we consider social networks construed as graphs or directed graphs, where the vertices represent actors in a network and edges and arcs represent how the actors are relationally tied to each other. The set of actors is assumed to be fixed and we are interested in how the set of edges or arcs changes over time.

Here follows a brief characterization of the general class of models. We restrict attention to binary social network data. Denote by $V = \{1, \ldots, n\}$ a fixed set of actors and denote by $\mathcal{N}$ the subset of pairs of actors that constitute the set of possible relational ties between actors. For a non-directed relation (network) such as friendship $\mathcal{N} = \binom{V}{2}$, and the elements in $\mathcal{N}$ are called edges. For a directed network such as "give advice to", $\mathcal{N} = V^{(2)} = \{(i, j) \in V \times V : i \neq j\}$, the elements in which are called arcs. There may also be other types of restrictions on $\mathcal{N}$ such as for example in the case of a bipartite network on two classes of actors A and B, in which case we set $\mathcal{N} = A \times B$. A (di-) graph on V with (arc) edge set $E \subseteq \mathcal{N}$, can be described by a collection $x = (x_e : e \in \mathcal{N})$ of $N = |\mathcal{N}|$, (arc) edge indicators, $x_e = \mathbf{1}\{e \in E\}$. We refer to x as the adjacency matrix of the implied graph or digraph since the conventional matrix form is easily recovered from x and $\mathcal{N}$. Let $\mathcal{X} = \{0, 1\}^{\mathcal{N}}$, be the space of all binary adjacency matrices.

We consider a continuous-time Markov chain $\{x(t)\}_{t \in \mathbb{R}}$, with outcome space $\mathcal{X}$, the intensity matrix $Q = (q(x, y))_{x, y \in \mathcal{X}}$, and transition matrix $P^t = (P^t_{xy})_{x, y \in \mathcal{X}}$. To capture the desired property that the edges change one at a time we proceed by first defining the Hamming metric for the elements of $\mathcal{X}$ as

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

and define $x(e)$ as the matrix that differs from x in exactly the element $e \in \mathcal{N}$. Note that $x(e)(e) = x$. Secondly, in the following we will only consider processes such that

$$q(x, y) = \begin{cases} q_e(x) & \text{if } y = x(e), \\ 0 & \text{if } |x - y| > 1, \\ -\sum_{e \in \mathcal{N}} q_e(x) & \text{if } x = y. \end{cases}$$

This can be interpreted as a process on the vertices of the binary N-cube, i.e. a graph $\mathcal{G} = (\mathcal{X}, \mathcal{E})$, $\mathcal{E} = \{(x, y) \in \mathcal{X} \times \mathcal{X} : |x - y| = 1\}$, that jumps between adjacent vertices. We refer to such jumps as mini-steps.

If we set $q_e(x) = N^{-1}$, we obtain the random walk on $\mathcal{X}$, and we have (see e.g. Aldous, 1983) that

$$P^t_{xy} = 2^{-N}(1 + e^{-2t/N})^{N-|x-y|}(1 - e^{-2t/N})^{|x-y|}, \tag{1}$$

for all $x, y \in \mathcal{X}$.

Consider now discriminating between changing an element of x from zero to one and changing an element from one to zero, the first corresponding to adding an edge and the second to removing an edge. To obtain the so called independent arcs model (Wasserman, 1977), we set $q_e(x)$ equal to λ_0 if $x_e = 0$, and λ_1 if $x_e = 1$. For $x, y \in \mathcal{X}$, define

$$N_{hk} = N_{hk}(x, y) = \sharp\{e | x_e = h, y_e = k\},$$

and the transition probabilities are given by

$$P^t_{xy} = \binom{N_{00} + N_{01}}{N_{01}} \binom{N_{11} + N_{10}}{N_{11}} \xi_0(t)^{N_{01}}(1 - \xi_0(t))^{N_{00}} \xi_1(t)^{N_{11}}(1 - \xi_1(t))^{N_{10}},$$

where

$$\xi_a(t) = \frac{\lambda_a}{\lambda_a + \lambda_{1-a}} + \frac{(-1)^{1-a}}{\lambda_{1-a}} \exp\{-(\lambda_a + \lambda_{1-a})t\},$$

for $a = 0, 1$.

A third model with an explicit form for P^t_{xy} is the reciprocity model (Wasserman, 1977, 1980; Leenders, 1995), defined for digraphs, $\mathcal{N} = V^{(2)}$, by

$$q_{ij}(x) = \lambda_0 + \mu_0 x_{ji} \quad (x_{ij} = 0), \tag{2}$$

and

$$q_{ij}(x) = \lambda_1 + \mu_1 x_{ji} \quad (x_{ij} = 1), \tag{3}$$

subject to the constraint that q_{ij} is positive for all $i, j \in V, j \neq i$ (since $e = (i, j)$). The formula for P_{xy}^t requires a little more space than is available here (we refer to Leenders, 1995, or Snijders, 1999, for the exact expression). Note, in Eqs. (2) and (3), the distinction that is made between reciprocated ties, $x_{ij} = x_{ji} = 1$, and non-reciprocated ties $x_{ij} \neq x_{ji}$.

3. Model specification

We now proceed to express the transition intensities as the product of two more easily interpretable factors. We define the class of models considered in this paper from the point of view of the embedded chain of a continuous-time process on $\mathscr{X}$. The transition probabilities in the embedded chain are

$$\eta(\theta, x, y), \tag{4}$$

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

$$\lambda(\theta, x). \tag{5}$$

It is assumed that $\eta(\theta, x, y) > 0$ for $x, y \in \mathscr{X}$ such that $|x - y| = 1$ and 0 otherwise. The $p \times 1$ parameter vector $\theta \in \Theta$ includes all unknown parameters. The two functions (4) and (5) determine the rate functions

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

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

It is worth pointing out that the formulation in terms of the embedded chain lends itself to easily interpreted change mechanisms. The transition probabilities (4) in the stochastic actor-oriented model (Snijders, 2001) for directed graphs are typically modelled using random utility models (Maddala, 1983).

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. In the sequel we refer by t_0 and t_1 to two generic consecutive observation points, $t_0 < t_1$, and $T = (t_1 - t_0)$.

For t_1 and t_0, denote the distance between two observations by

$$H = |x(t_1) - x(t_0)|.$$

From the construction of the mini-step, the number of mini-steps used to transform $x(t_0)$ into $x(t_1)$ must equal $m = H + 2k$, for some $k \in \{0, 1, 2, \ldots\}$.

For given m let $\mathscr{W}_m(x, y)$ be the set of all sequences $v_0, v_1, \ldots, v_m$, such that $v_0 = x$ and $v_m = y$, and $|v_{i-1} - v_i| = 1$, $v_i \in \mathscr{X}$ for $i = 1, \ldots, m$. Equivalently, $\mathscr{W}_m(x, y)$ is the set of all m-walks in the N-cube $\mathscr{G} = (\mathscr{X}, \mathscr{E})$, from x to y. Unless otherwise stated $\mathscr{W}_m = \mathscr{W}_m(x(t_0), x(t_1))$. Define also $\mathscr{S}_m = \{(u_1, \ldots, u_m) \in (0, T)^m : u_1 + \cdots + u_m < T\}$. For a fixed m, let the latent variable $w = (y, u)$, comprise $y = (y_h)_{h=0}^m \in \mathscr{W}_m$, and $u = (u_h)_{h=1}^m \in \mathscr{S}_m$, with the interpretation that y constitutes a walk from the observation $x(t_0)$ to the observation $x(t_1)$, for which the time between consecutive changes y_h and y_{h+1} is u_h, for $h = 1, \ldots, m$. For a given $w \in \mathscr{C}_m = \mathscr{W}_m \times \mathscr{S}_m$, we define the augmented likelihood

$$L(\theta; w, x(t_0)) = \exp\left\{ -\sum_{h=1}^{m+1} u_h \lambda(\theta, y_{h-1}) \right\} \prod_{h=1}^{m} \eta(\theta, y_{h-1}, y_h)\lambda(\theta, y_{h-1}). \tag{6}$$

Given observations $x(t_0)$ and $x(t_1)$ the data likelihood

$$P_{x(t_0), x(t_1)}^T = L_D(\theta; x(t_1), x(t_0))$$

is obtained from (6) by marginalizing with respect to $w \in \mathscr{C} = \cup_{k=0}^\infty \mathscr{C}_{H+2k}$.

4. Parameter inference

Assuming that we have a model in the form specified in the previous section, and that we have specified a prior distribution $\pi(\theta)$, the main objective of the proposed inference scheme is to obtain the posterior distribution

$$\pi(\theta|x(t_0), \ldots, x(t_M)) \propto \pi(\theta) \prod_{\ell=0}^{M-1} L_D(\theta; x(t_{\ell+1}), x(t_\ell)). \tag{7}$$

In accordance with the account of the model formulation above, since the likelihood factors, we may treat each pair of consecutive observations separately.

Under the assumption that the augmented likelihood $L(\theta; w, x(t_0))$ is cheap to evaluate for all θ and $w \in \mathscr{C}$, a sample from the joint posterior of w and θ can be obtained using the adaption of the reversible jump MCMC (Green, 1995) proposed in Koskinen (2004; due to the nature of the model, the inference scheme has many features in common with inference in models for disease transmission, e.g. Auranen et al., 2000). This sampling scheme rests on the fact that in the full conditional posterior:

$$\pi(w|\theta, x(t_1), x(t_0)) \propto L(\theta; w, x(t_0)), \tag{8}$$

the data likelihood only enters as a multiplicative constant when data are augmented (Tanner and Wong, 1987) by the latent walk. The full conditional posterior of the parameter vector θ given data and an observation w on the latent evolution is

$$\pi(\theta|w, x(t_1), x(t_0)) \propto L(\theta; w, x(t_0))\pi(\theta).$$

To obtain a sample from the marginal posterior of θ given data (7), we sample from the joint distribution of θ and w given data giving a sequence $(\theta^{(g)}, w^{(g)})_{g=1}^G$, where $(\theta^{(g)})_{g=1}^G$ is an approximate sample from the marginal distribution of θ given data. In each iteration g we successively draw

$$w^{(g)} \sim w|\theta^{(g-1)}, x(t_1), x(t_0),$$

and

$$\theta^{(g)} \sim \theta|w^{(g)}, x(t_1), x(t_0).$$

For $w^{(g-1)} \in \mathscr{C}_m$, we perform draws from (8) using a Metropolis–Hastings updating step, with the necessary modifications for variable dimensions (Green, 1995), accepting or rejecting a proposed move to $w^* \in \mathscr{C}_{m+2k}$, for a $k = -1, 0, 1$, chosen with probability a_m^-, a_m^0, and a_m^+, respectively. The Hastings ratio for the acceptance probability is given by

$$\frac{L(\theta^{(g-1)}; w^*, x(t_0))}{L(\theta^{(g-1)}; w^{(g-1)}, x(t_0))} \frac{q(y^{(g-1)}|y^*)q(u^{(g-1)}|y^{(g-1)}, w^*)}{q(y^*|y^{(g-1)})q(u^*|y^*, w^{(g-1)})} J,$$

where we use q as a generic form for the proposal density and J is the Jacobian of the transformation of going from $u^{(g-1)}$ to u^*.

4.1. Move type 1: propose longer walk

Given a latent walk $w = (y_h, u_h)_{h=1}^m$, we propose a move to $w^* \in \mathscr{C}_{m+2}$, by choosing an element $e \in \mathscr{N}$, uniformly at random, and selecting two insertion points h' and h'' where we insert two superfluous changes. For a chosen $e \in \mathscr{N}$, the proposed latent walk is set to

$$y_0, y_1, \ldots, y_{h'}, y_{h'}(e), y_{h'+1}(e), \ldots, y_{h''}(e), y_{h''}(e)(e), y_{h''+1}, \ldots, y_m.$$

In order to write the proposal probability $q(y^*|y)$ we need to introduce some extra notation. Firstly, if we let e_h denote the element that is changed in the hth mini-step, i.e. $y_h = y_{h-1}(e_h)$, the sequence $(y_h)_{h=1}^m$ is completely determined by y_0 and the sequence $(e_h)_{h=1}^m$. Second, let m_{ev} be the number of uninterrupted e-runs of length v in $(e_h)_{h=1}^m$, and denote by m_e the total number of times the element e is changed. Note that we have $m_e = \sum_v v m_{ev}$ and that for element

$e \in \mathcal{N}$, m_e is odd if and only if $|x_e(t_0) - x_e(t_1)| = 1$. Since there are $m - m_e - 2$ distinct insertion points once we have chosen e, the implied proposal probability is given by

$$q(y^*|y) = a_m^+ \frac{1}{N} \binom{m - m_e - 2}{2}^{-1}.$$

4.2. Move type 2: propose shorter walk

If $m > H$ we may propose a move to $w^* \in \mathscr{C}_{m-2}$ by removing a superfluous change to an element $e \in \mathcal{N}$, from among the N^+ elements e for which $m_e > 1$. In total there are $m_{e+} = \sum_{v \geqslant 1} m_{ev}$ e-runs and given that we have chosen the element $e \in \mathcal{N}$, we can either remove a pair of changes to e from one of the $m_{e+} - m_{e1}$ e-runs of length greater than 1 or a pair from two distinct e-runs. The proposal probability of each walk proposed in this way is given by

$$q(y^*|y) = a_m^- \frac{1}{N^+} \left[\binom{m_{e+}}{2} + m_{e+} - m_{e1} \right]^{-1}.$$

4.3. Move type 3 and updating holding times

Moves to $w^* \in \mathscr{C}_m$ simply consists of changing the order of mini-steps in w. Rearranging the order can be done in many ways and a convenient way of doing this is by choosing h' and h'' such that $e_{h'} \neq e_{h''}$, setting $e_{h'}^* = e_{h''}$ and $e_{h''}^* = e_{h'}$. Since this move is symmetric the proposal ratio is unity. This move is not required for the ergodicity of the Markov chain but is included to improve mixing (for the details for drawing w see Koskinen, 2004).

If we propose a move from $w \in \mathscr{C}_m$ to $w^* \in \mathscr{C}_{m+2}$, we need to propose two new holding times $u_{h'}^*$ and $u_{h''}^*$ in addition to the extra changes. These are drawn from a rescaled Dirichlet distribution with parameters $(1,1,m+1)$, and to make room for $u_{h'}^*$ and $u_{h''}^*$, the current holding times are scaled by $(T - u_{h'}^* - u_{h''}^*)/T$. The part in the Hastings ratio that depends on the holding times becomes

$$\frac{q(u; y, w^*)}{q(u^*; y^*, w)} J = \frac{T^2}{(m+2)(m+1)}. \tag{9}$$

For moves from $w \in \mathscr{C}_{m+2}$ to $w^* \in \mathscr{C}_m$, two holding times are removed and the remaining rescaled in an analogous fashion. Instead of the Hastings ratio in (9), we have the reciprocal of this.

4.4. Updating the parameter vector

Although it may sometimes be possible to perform draws from the full conditional posterior $\cdot|w^{(g)}, x(t_1), x(t_0)$ using standard distributions, typically the model specification only permits that we use a Metropolis–Hastings updating step. If we propose a move to θ^* sampled from a distribution $q(\theta^*|\theta^{(g-1)})$, given that we are in the current state $\theta^{(g-1)}$, the Hastings ratio is of the standard form

$$\frac{L(\theta^*; w, x(t_0))\pi(\theta^*)}{L(\theta^{(g-1)}; w, x(t_0))\pi(\theta^{(g-1)})} \frac{q(\theta^{(g-1)}|\theta^*)}{q(\theta^*|\theta^{(g-1)})}.$$

For $q(\cdot|\theta)$ in the empirical examples below we have used a Gamma distribution for the non-negative parameter and independently thereof a multivariate normal distribution, both centred on the current parameter point. Although the mixing of the MCMC sampler is contingent on the choice of proposal density to a large extent (Tierney, 1994), standard distributions are satisfactory provided the scaling is chosen appropriately. The variance for a multivariate normal proposal distribution is conveniently set using a quasi-Newton algorithm that outputs the Hessian when maximizing the complete data likelihood for a fixed $w \in \mathscr{C}_H$.

5. Example: the electronic information exchange system

In this section we analyse the electronic information exchange system (EIES) data collected by Freeman and Freeman (1979), using one of the models fitted with the method of moments in Snijders and van Duijn (1997). The relation

Table 1
Point estimates for parameters in model fitted to EIES data using proposed Bayesian analysis and method of moments (MM), with posterior standard deviations and approximate standard errors

	MM	SE	Bayes	STD
α, activity on rate	.58	.333	−.23	.384
β_1, activity	−1.70	.582	−1.23	.434
β_2, reciprocity	1.02	.270	.90	.238
β_3, citation attraction	−.52	.228	−.61	.214
β_4, citation dissimilarity	−.20	.223	−.10	.190
β_5, transitivity	.12	.029	.098	.0163
ρ, constant in rate	4.89	1.055	4.94	.81

(originally valued) is directed and was measured at two points in time for a network consisting of $n = 32$ actors ($N = 992$). The data set is described in Wasserman and Faust (1994). We consider the relation defined by the respondent claiming to have met the other person or being a friend. Arguments for the proposed model specification and details of the dichotomization of the relation can be found in Snijders and van Duijn (1997).

The intention is to model the evolution of the network as driven by the actors' efforts to seek "better" network configurations. A binary variable z_i is available, reflecting the citation rate of actor $i \in V$. First, we introduce the parameters $\rho > 0$, α, and $\beta_1, \ldots, \beta_5$, denoting these collectively by the parameter vector θ. Define for each $i \in V$ and $x \in \mathcal{X}$:

- out-degree: $s_{i1}(x) = \sum_{j \in V \setminus \{i\}} x_{ij}$,
- reciprocity: $s_{i2}(x) = \sum_{j \in V \setminus \{i\}} x_{ij} x_{ji}$,
- citation attraction: $s_{i3}(x) = \sum_{j \in V \setminus \{i\}} x_{ij} z_j$,
- citation dissimilarity: $s_{i4}(x) = \sum_{j \in V \setminus \{i\}} x_{ij} |z_i - z_j|$,
- transitivity: $s_{i5}(x) = \sum_{j, h \in V \setminus \{i\}} x_{ij} x_{ih} x_{jh}$.

It is assumed that all actors i independently make changes to the vector $(x_{ij})_{j \neq i}$ at rate

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

and thus (5) is equal to $\lambda(\theta, x) = \sum_{i \in V} \lambda_i(x, \theta)$. Assume further that once actor i makes a change at time t, i changes his/her out-relation to exactly one actor j, choosing j such that the "utility"

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

of the newly obtained x is maximized, where the systematic part $r(\theta, i, j, x) = \sum_{k=1}^5 \beta_k s_{ik}[x((i, j))]$, and the random components $\varepsilon_i(t, x, j)$ are iid type 1 extreme value distribution for all t, x, and j. It is straightforward to show (Maddala, 1983) that the transition probability (4) for going from x to $x((i, j))$ in this example is given by

$$\eta(\theta, x, x((i, j))) = \frac{\lambda_i(x, \theta)}{\lambda(\theta, x)} \frac{e^{r(\theta, i, j, x)}}{\sum_{j \in V \setminus \{i\}} e^{r(\theta, i, j, x)}},$$

which together with (5) provides us with all we need for the augmented likelihood function (6).

The model is fitted using "non-informative" prior distributions for the parameters (a burn-in of between 5 and 10 thousand iterations seems sufficient in this case and a thinning constant of 50–100 iterations; tables and figures are based on a total sample length of $G = 50,000$, burn-in of 4000, and thinning constant of 100). The point estimates for the Bayesian procedure (posterior expected value) and the method of moments (MM) in Table 1 largely agree in the sense that all MM-estimates are included in the corresponding .95 highest posterior density (HPD) intervals (Fig. 1 (a)), with the exception of α, the influence of activity on rate. From Fig. 1 (a) we may conclude that there is no evidence for effects in the evolution of the EIES network of activity on rate (α) or of a differential preference for ties depending on dissimilarity in the number of citations (β_4). One has to be a little cautious when interpreting marginal posteriors
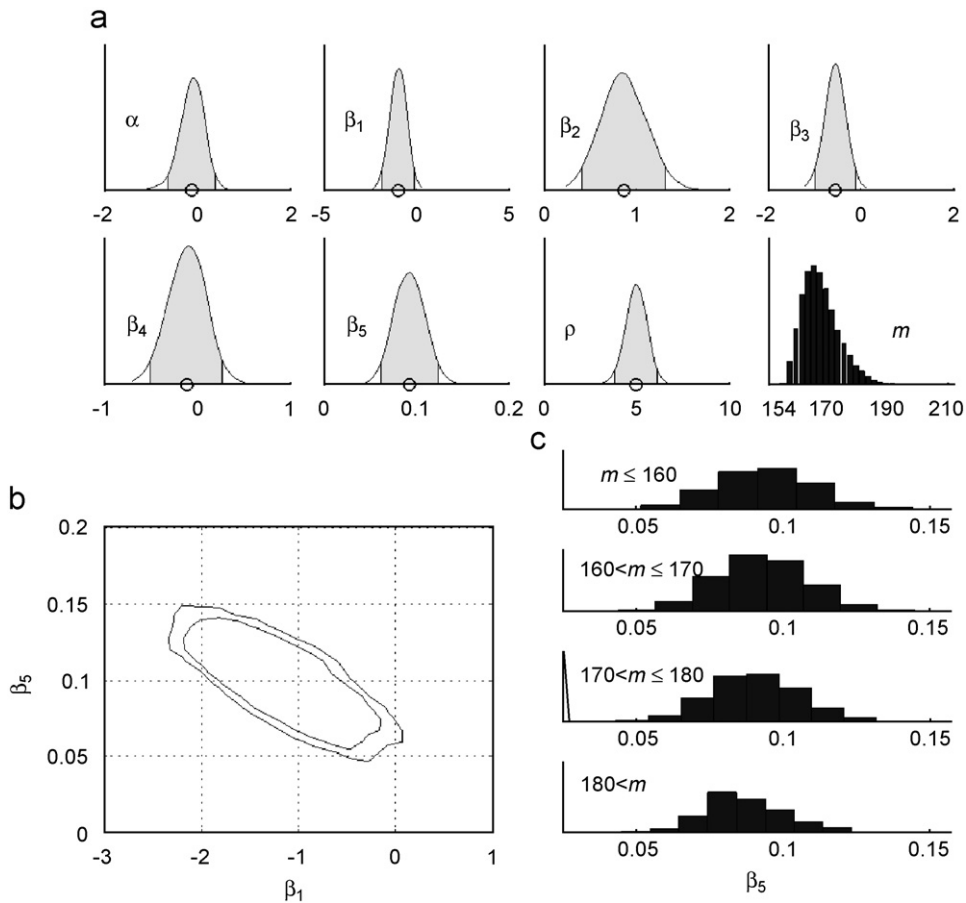
Fig. 1. Posterior analysis for the EIES data: (a) marginal posterior distributions with .95 HPD intervals, (b) bivariate posterior with .95 and .99 HPD region, and (c) posterior conditional on walk length.

since there is usually a relatively high degree of posterior correlation between parameters, such as for example between β_1 and β_5, Fig. 1 (b). Given data, the number of unobserved changes, m, between observations appears to be somewhat higher (Fig. 1 (a)) than the 154 changes required to transform $x(t_0)$ into $x(t_1)$. The dependence of the parameters on the walk length is moderate as exemplified in Fig. 1 (c).

6. Example: university freshmen

We now turn to look at friendship nominations for a group of 32 university freshmen in a Dutch university. There are observations for three points in time, t_0, t_1, and t_2, with three weeks between each observation (the observations used here were labelled t_2, t_3, and t_4, respectively, in the original study, van de Bunt et al., 1999). We fit a reciprocity model (as described in Section 2) as well as an extended model to $x(t_0)$ and $x(t_1)$. The extended model is similar to the model in the preceding section. It comprises the effect of activity on rate α; and the systematic part of r of the "utility" function U_i is a weighted sum of the activity, reciprocity, and transitivity statistics, a number of covariate effects, and moreover the statistic counting the number of other actors at geodesic distance 2, defined as $s_i(x) = \sharp\{j : x_{ij} = 0, \max_k(x_{ik}x_{kj}) > 0\}$. The marginal posteriors for the reciprocity model suggests that there may not be a strong tendency towards mutuality and when the extended model is fitted, only higher-order effects, the number of distance-2 alters and transitivity, appear to be of importance. (van de Bunt et al., 1999, conclude that reciprocity only has an effect after t_2.)

Consider now evaluating the two model specifications using the observation $x(t_2)$ (that was not used in the estimation process). While it is generally meaningless to predict a specific network configuration in social network analysis,
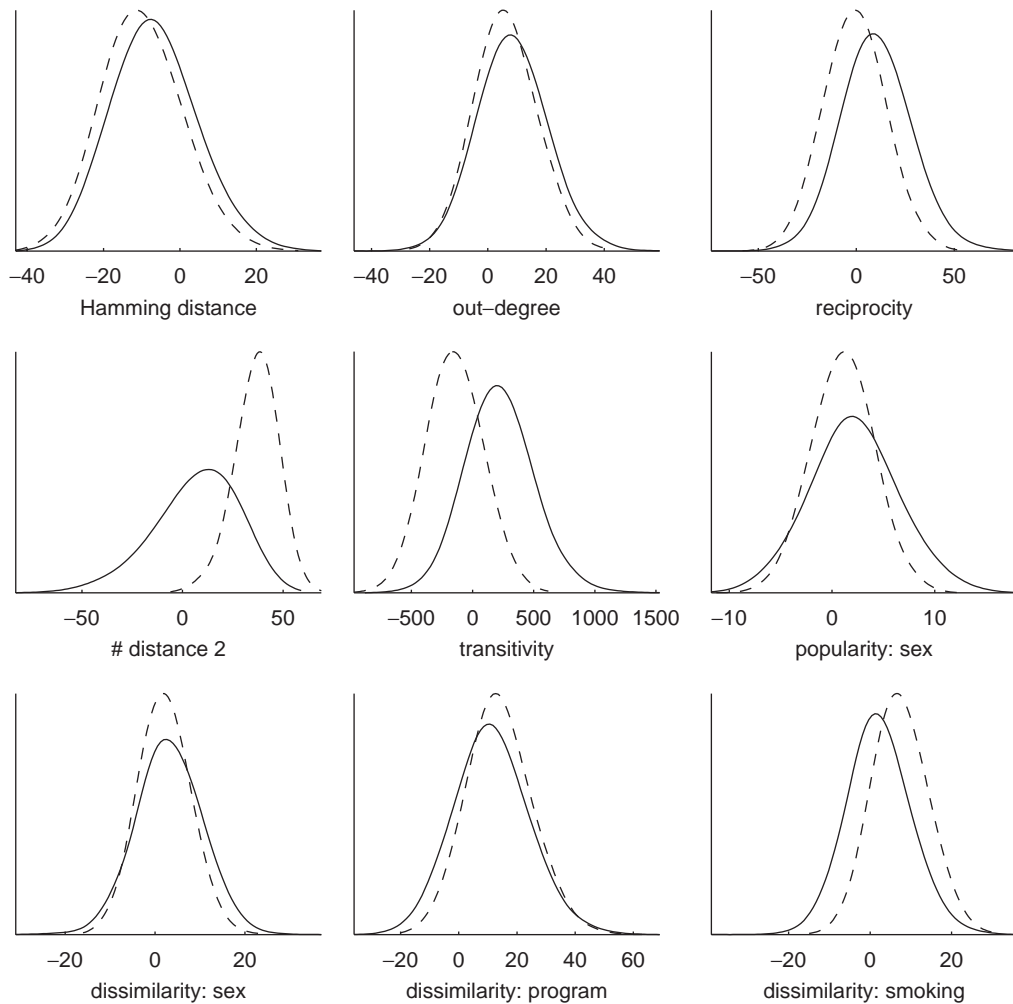
Fig. 2. Predictions for van de Bunts freshmen students. Differences $Z(y(t_2)) - Z(x(t_2))$, for $y(t_2)$ from the predictive distribution $p(y(t_2)|x(t_1), x(t_0))$ (dashed, reciprocity model, and solid lines extended model).

summaries of structural properties are of interest. Frequently in the literature structural properties of networks are invoked as determinants of behaviour and performance, or of the capability and capacity of the networks as transmission paths of diseases or information.

Given a sample $(\theta^{(g)})_{g=1}^{G}$ from the posterior distribution of θ given $x(t_0)$ and $x(t_1)$, we may obtain a sample from the predictive distribution

$$p(y(t_2)|x(t_0), x(t_1)) = \int P_{x(t_1), y(t_2)}^{t_2 - t_1} \pi(\theta|x(t_0), x(t_1))\, d\theta,$$

by simulating adjacency matrices $y(t_2)$ for each $\theta^{(g)}$, according to the process defined by $\eta(\theta^{(g)}, x, y)$ and $\lambda(\theta^{(g)}, x)$. A few predictive distributions (of functions of the network) for the university freshmen at time t_2, obtained in this way, for the two fitted models are displayed in Fig. 2. For ease of comparison we have subtracted the observed value of the corresponding statistic. While the reciprocity model captures all of the dyad-level statistics it is not sophisticated enough to capture adequately the tendency towards network closure (a negative parameter corresponding to the distance-2 statistic indicates a "dislike" of having indirect friends, given that everything else is constant).

7. Discussion

In this paper we have proposed an alternative estimation procedure to the method of moments estimator for longitudinal social network data. In place of the approximate standard errors from the MM-analysis as a measure of uncertainty, we are provided with an entire sample from the posterior distribution of the parameters. We may explore the posterior distribution thus obtained to investigate marginal as well as joint distributions of the parameters and in addition the distribution of functions of the parameters such as log-odds. The model formulation in combination with the Bayesian approach allows for natural extensions to treat for example missing data and networks with vertex set changing over time. For the empirical illustrations we have settled for non-informative priors. Although a fully Bayesian analysis may seem a difficult task in light of the relative complexity of the model, a compromise, when proper prior distributions are required, is to employ prior distributions trained (O'Hagan, 1995) on some of the initial observations. Parts of the procedure (programmed by Michael Schweinberger) are available in the PC program SIENA, which is part of the StOCNET package (available free of charge from, http://stat.gamma.rug.nl/stocnet).

References

Aldous, D., 1983. Minimization algorithms and random walk on the d-cube. Ann. Probab. 11, 403–413.

Auranen, K., Arjas, E., Leino, T., Takala, A.K., 2000. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. J. Amer. Statist. Assoc. 95, 1044–1053.

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

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

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

Klovdahl, A.S., 1985. Social networks and the spread of infectious diseases: the AIDS example. Soc. Sci. Med. 21, 1203–1216.

Koskinen, J.H., 2004. Bayesian inference for longitudinal social networks. Research Report, Number 2004:4, Department of Statistics, Stockholm University.

Leenders, R.Th.A.J., 1995. Models for network dynamics: a Markovian framework. J. Math. Sociol. 20, 1–21.

Maddala, G.S., 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.

Moreno, J.L., 1934. Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama. Nervous and Mental Disease Publishing Co, Washington, DC.

O'Hagan, A., 1995. Fractional Bayes factors for model comparison. J. Roy. Statist. Soc. Ser. B 57, 99–138 (With discussion and a reply by the author).

Snijders, T.A.B., 1999. The transition probabilities of the reciprocity model. J. Math. Sociol. 23, 241–253.

Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. Sociol. Methodol. 30, 361–395.

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

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation (with discussion). J. Amer. Statist. Assoc. 82, 528–550.

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Ann. Statist. 22, 1701–1762.

van de Bunt, G.G., van Duijn, M.A.J., Snijders, T.A.B., 1999. Friendship networks through time: an actor-oriented statistical network model. Comput. Math. Organ. Theory 5, 167–192.

van Duijn, M.A.J., Zeggelink, E.P.H., Huisman, M., Stokman, F.M., Wasseur, F.W., 2003. Evolution of sociology freshmen into a friendship network. J. Math. Sociol. 27, 153–191.

Wasserman, S., 1977. Stochastic Models for Directed Graphs. Department of Statistics, University of Harvard.

Wasserman, S., 1980. Analyzing social networks as stochastic processes. J. Amer. Statist. Assoc. 75, 280–294.

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