# Random Effects Models for Digraph Panel Data

Michael Schweinberger[†] and Tom A.B. Snijders[††]

## Abstract

Digraph panel data, corresponding to a given set of nodes and the directed graphs (digraphs) on the set of nodes which are observed at two or more discrete time points, are collected in the social sciences and other fields. Conventional models of digraph panel data assume that the data are discrete outcomes of a continuous-time Markov process on the set of possible digraphs defined on the set of nodes. Such models make the implicit assumption that all relevant knowledge with respect to nodes is observed in the form of covariates and correctly incorporated in the model, which may not be satisfied in applications. The present paper proposes Markov models which allow for unobserved heterogeneity across nodes by introducing random variables with unobserved outcomes, called random effects. To estimate parameters, maximum likelihood and Bayesian methods are proposed—using Markov chain Monte Carlo—and illustrated by an application to longitudinal social network data.

*Keywords:* longitudinal social network data, continuous-time Markov process, latent variables, hidden Markov models.

[†] Visiting Research Scholar, University of Washington; supported by grant 401-01-550 from the Netherlands Organisation for Scientific Research (NWO).

[††] University of Oxford

# 1. INTRODUCTION

In the social sciences and other fields, graphs have been exploited to represent data structures that correspond to links between entities (Wasserman and Faust 1994). An example is provided by sexual relationships between individuals (Jones and Handcock 2003), the structure of which is key to understanding the contagion of sexually transmitted diseases such as HIV / AIDS. Some other examples are friendships among university freshmen; old-boy networks; cooperation or transmission of information among employees of companies; transactions between companies; and trade relations between countries.

The present paper focuses on a given set of entities (nodes) and the directed links (arcs) between the entities, which can be represented by directed graphs (digraphs). It has long been argued in the social sciences (see, e.g., Holland and Leinhardt 1976) that the arcs, considered as random variables, tend to be dependent. To gain insight into the process that generates such dependent data, it is important to collect longitudinal data. Due to data collection constraints, panel data are the most common form of longitudinal data, that is, the digraph is observed at two or more discrete time points.

Conventional models of digraph panel data, dating back to Holland and Leinhardt (1977), assume that the digraph evolution is governed by a Markov process on the set of possible digraphs defined on the set of nodes, which operates in continuous time but is only observed at two or more discrete time points. Snijders (2001) considered a large family of continuous-time Markov models for digraph panel data. It allows to build models which capture the most important classes of dependencies in social networks, and which can readily be communicated to social scientists, because the models have an appealing interpretation in social science terms and can be regarded as directly substantive probability models in the sense of Cox (1990). The basic idea is to model the digraph evolution as a continuous-time Markov process, and let the nodes represent social actors who add and delete arcs with the purpose to obtain the best possible value of some node-specific objective function and random terms. In the model of Snijders (2001), the objective function is a weighted sum of statistics. The statistics depend on the digraph, and may include nodal covariates to account for the fact that nodes are heterogeneous. The weights, regarded as parameters, are assumed to be constant across nodes. The implicit assumption is that all there is to know with regard to nodes is observed in the form of covariates and correctly incorporated in the model. However, it is not unusual that some relevant nodal covariates are unobserved due to data collection constraints and limited prior knowledge of researchers as to what covariates are relevant, which casts doubt on the constant-weights assumption.

The present paper proposes to account for unobserved heterogeneity across nodes by introducing random variables with unobserved outcomes, called random effects. Random effects models are widely used in the social sciences (see, e.g., Longford 1993; Raudenbush and Bryk 2002; Skrondal and Rabe-Hesketh 2004); examples of random effects models for non-longitudinal social network data are Hoff (2005); Zijlstra, Van Duijn, and Snijders (2006). The random effects models considered here replace the constant-weights assumption by the assumption that the weights are unobserved outcomes of nodal random variables, governed by a probability law that is common to all nodes.

Maximum likelihood and Bayesian estimation of the parameters is proposed; for maximum likelihood estimation, the non-redundant elements of the random effects variance-covariance matrix are reparametrized so that estimates of the variance-covariance matrix are by construction symmetric and positive definite, and the estimation of very small variances is facilitated. Both maximum likelihood and Bayesian estimation use Markov chain Monte Carlo methods.

The paper is structured as follows. Section 2.1 describes a family of continuous-time Markov models with fixed effects, while Section 2.2 introduces random effects. Sections 3 and 4 discuss maximum likelihood and Bayesian estimation, respectively. Section 5 applies the model to longitudinal social network data.

## 2. MODEL

It is assumed that a binary, directed relation (digraph) on a given set of nodes $\mathcal{N} = \{1, \ldots, n\}$ has been observed at discrete time points $t_0 < t_1 < \cdots < t_H$. The observations are stored as binary $n \times n$ matrices $\mathbf{X}(t_0), \mathbf{X}(t_1), \ldots, \mathbf{X}(t_H)$, where $X_{ij}(t_h) = 1$ if there is an arc from node $i$ to node $j$ at time point $t_h$, and $X_{ij}(t_h) = 0$ otherwise; by convention, the diagonal elements are defined as $X_{ii}(t_h) \equiv 0$.

The assumption that $X_{ij}(t_h)$ is binary is made because such data are common and convenient; however, it is possible to extend the model to the case where $X_{ij}(t_h)$ takes on discrete, ordered values.

### 2.1 Fixed effects models

A simple model can be constructed by conditioning on the digraph $\mathbf{X}(t_0)$ and assuming that the digraphs $\mathbf{X}(t_1), \ldots, \mathbf{X}(t_H)$ are outcomes of a continuous-time Markov process on the set of possible digraphs defined on $\mathcal{N}$.

The continuous-time Markov process starts at time $t \equiv t_0$ with digraph $\mathbf{X} \equiv \mathbf{X}(t_0)$. A holding time $\Delta t$ is sampled from the negative exponential distribution with parameter $\phi$ (to be specified below), and at time $t = t + \Delta t$ the digraph $\mathbf{X}$ is allowed to change. Snijders (2001), following Holland and Leinhardt (1977), postulated that one, and only one, element $X_{ij}$ of $\mathbf{X}$ is allowed to change, and modeled the change process as driven by nodes $i$—representing social actors—as follows: the parameter $\phi \equiv \phi(\mathbf{X}, \boldsymbol{\theta})$ of the negative exponential distribution is decomposed into node-dependent rates of change $\phi_i(\mathbf{X}, \boldsymbol{\theta})$,

$$\phi(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \phi_i(\mathbf{X}, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is a parameter vector. Conditional on the event that $\mathbf{X}$ is allowed to change, a node $i \in \mathcal{N}$ is chosen with probability

$$\frac{\phi_i(\mathbf{X}, \boldsymbol{\theta})}{\phi(\mathbf{X}, \boldsymbol{\theta})}.$$

The chosen node $i$ is assumed to consider changing some element $X_{ij}$ by choosing the node $j \in \mathcal{N}$ which maximizes

$$f_i(\mathbf{X}, j, \boldsymbol{\theta}) + U_{ij}(t), \tag{1}$$

where $f_i(\mathbf{X}, j, \boldsymbol{\theta})$ is called the objective function of $i$ and $U_{ij}(t)$ is a random variable, representing the unknown determinants that influence the choice of $i$ and randomness. If (1) is maximized by choosing $j = i$, then $i$ is assumed to disregard all possible changes and to change nothing, otherwise $i$ is assumed to transform the element $X_{ij}$ into $1 - X_{ij}$. The process proceeds by updating $t$ and $\mathbf{X}$ in the described fashion.

Models can be specified by specifying the rate function $\phi_i(\mathbf{X}, \boldsymbol{\theta})$, the objective function $f_i(\mathbf{X}, j, \boldsymbol{\theta})$, and the distribution of $U_{ij}(t)$.

The rate function $\phi_i(\mathbf{X}, \boldsymbol{\theta})$ may be constant across nodes,

$$\phi_i(\mathbf{X}, \boldsymbol{\theta}) = \alpha,$$

where $\alpha > 0$ is a parameter, or non-constant,

$$\phi_i(\mathbf{X}, \boldsymbol{\theta}) = \alpha \exp\left[\boldsymbol{\epsilon}' e_i(\mathbf{X})\right],$$

where $\boldsymbol{\epsilon}$ is a parameter vector and $e_i(\mathbf{X})$ is a statistics vector depending on $\mathbf{X}$ and covariates; if $H \geq 2$, then $\alpha$ may depend on time interval $[t_{h-1}, t_h]$ $(h = 1, \ldots, H)$.

In the model of Snijders (2001), the objective function $f_i(\mathbf{X}, j, \boldsymbol{\theta})$ is given by

$$f_i(\mathbf{X}, j, \boldsymbol{\theta}) \quad = \quad \boldsymbol{\eta}' s_i(\mathbf{X}, j), \tag{2}$$

where $\boldsymbol{\eta}$ is a parameter vector and $s_i(\mathbf{X}, j)$ is a statistics vector; statistics can depend on $\mathbf{X}$, $j$, and covariates, and can be used to induce dependence among the arc processes $X_{ij}(t)$ (see Section 5).

It is convenient to assume that the $U_{ij}(t)$ are i.i.d. random variables with Gumbel$(0, 1)$ distribution (all $i, j, t$), because then the probability that a given node

$i$ chooses node $j \in \mathcal{N}$ can be written in closed form (McFadden 1974):

$$\psi_i(j \mid \mathbf{X}, \boldsymbol{\theta}) = \frac{\exp\left[f_i(\mathbf{X}, j, \boldsymbol{\theta})\right]}{\sum_{h=1}^{n} \exp\left[f_i(\mathbf{X}, h, \boldsymbol{\theta})\right]}. \tag{3}$$

## 2.2 Random effects models

The fixed effects models of Section 2.1 assume that the weight $\boldsymbol{\eta}$ is constant across nodes. To allow for node-specific weights, objective function (2) can be replaced by

$$f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta}) = \boldsymbol{\eta}_i' \, s_i(\mathbf{X}, j), \tag{4}$$

where the node-specific weight $\boldsymbol{\eta}_i$ is given by

$$\boldsymbol{\eta}_i = \boldsymbol{\beta} + \mathbf{A}\,\mathbf{V}_i,$$

where $\boldsymbol{\beta}$ is a $L \times 1$ parameter vector, $\mathbf{V}_i$ is a $K \times 1$ ($K \leq L$) random vector, and $\mathbf{A}$ is a $L \times K$ design matrix.

The $\mathbf{V}_i$—called random effects—are assumed to be independent of time and to be i.i.d. random variables with $K$-variate Gaussian distribution,

$$\mathbf{V}_i \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}), \; i = 1, \ldots, n,$$

where $E[\mathbf{V}_i] = \mathbf{0}$ can be assumed without loss of generality, and $\boldsymbol{\Sigma}$ is positive definite.

Such models can capture both observed nodal heterogeneity, in the form of nodal covariates contained in statistic $s_i(\mathbf{X}, j)$, and unobserved nodal heterogeneity, in the form of nodal random effects $\mathbf{V}_i$.

The random effects introduced above depend on the "sender" $i$ of the arc variable $X_{ij}$, but random effects may depend in addition on the "receiver" $j$. Suppose that, when node $i$ is allowed to change something, $i$ chooses the node $j \in \mathcal{N}$ which maximizes

$$f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta}) + T_j + U_{ij}(t),$$

where $T_j$ is a scalar-valued random effect. The random effect $T_j$ can be interpreted as the latent popularity of $j$, and is an alternative to other representations of popularity in terms of $f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta})$.

## 3. MAXIMUM LIKELIHOOD ESTIMATION

Let $H = 1$, so that the time interval under consideration is $[t_0, t_1]$ and the observations are the digraphs $\mathbf{X}(t_0)$ and $\mathbf{X}(t_1)$; by the Markov property, the extension to the case $H \geq 2$ is straightforward.

A complete observation of the continuous-time Markov process in time interval $[t_0, t_1]$ corresponds to digraphs $\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{M-1}, \mathbf{X}_M$, and holding times, where $M$ denotes the total number of opportunities for change in time interval $[t_0, t_1]$. The digraphs $\mathbf{X}_0 \equiv \mathbf{X}(t_0)$ and $\mathbf{X}_M \equiv \mathbf{X}(t_1)$ are observed and denoted by $\mathbf{y}$. The digraphs $\mathbf{X}_1, \dots, \mathbf{X}_{M-1}$ are unobserved and can be represented by $\mathbf{X}_0$ and the sequence $w = (i_m, j_m)_{m=1}^{M}$, where $m$ refers to the $m$-th opportunity for change, $i_m$ is the node that was allowed to change something, and $j_m$ is the node chosen by $i_m$. The holding times are unobserved, but cancel in the case of constant rate functions $\phi_i(\mathbf{X}, \boldsymbol{\theta}) = \alpha$ (see Section 3.2), which—for ease of presentation—is assumed henceforth; the case of non-constant rate functions can be obtained along the lines of Snijders, Koskinen, and Schweinberger (2007), who considered maximum likelihood estimation of fixed effects models. The random effects $\mathbf{V}_i$ are unobserved, and stored as rows of matrix $\mathbf{V}$. The unobserved data $\mathbf{V}$ and $w$ are referred to as $\mathbf{z}$. The parameters $\alpha$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}^{-1}$ are collected in parameter vector $\boldsymbol{\theta}$.

Under regularity conditions, the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ solves

$$\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}) = \mathbf{0}, \tag{5}$$

where $\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y})$ is the gradient of $\ln p_{\boldsymbol{\theta}}(\mathbf{y})$ with respect to $\boldsymbol{\theta}$, and $p_{\boldsymbol{\theta}}(\mathbf{y})$ is the probability density of $\mathbf{y}$. The problem is that $\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y})$ is, in general, intractable.

In the incomplete-data literature, a result of Fisher (1925) (cf. Efron 1977) turned out to be useful in dealing with such intractable estimation problems. Observe that

$$\nabla_{\boldsymbol{\theta}}\, p_{\boldsymbol{\theta}}(\mathbf{y}) = \nabla_{\boldsymbol{\theta}} \int p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})\ \mathrm{d}\,\mu(\mathbf{z}) = \int \nabla_{\boldsymbol{\theta}}\, p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})\ \mathrm{d}\,\mu(\mathbf{z}), \tag{6}$$

where interchanging the order of differentiation and integration is admissible by Theorem 2.7.1 of Lehmann and Romano (2005, p. 49) and the fact that $p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ is an exponential family density (see Section 2). By multiplying (6) under the integral sign by $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathbf{y})\, p_{\boldsymbol{\theta}}(\mathbf{y}) \,/\, p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$, one obtains the so-called Fisher identity:

$$\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}) \quad = \quad E_{\boldsymbol{\theta}}\left[\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}\right]. \tag{7}$$

Thus, solving (5) is equivalent to solving

$$E_{\boldsymbol{\theta}}\left[\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}\right] = \mathbf{0}. \tag{8}$$

Analytical evaluation of the expectation in (8) is infeasible, which rules out the use of standard root-finding methods such as Newton-Raphson to solve (8). However, if sampling from $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathbf{y})$ is possible, then the expectation in (8) can be approximated and root-finding algorithms based on stochastic approximation (Chen 2002) can be used to solve (8) (cf. Gu and Kong 1998). Here, Markov chain Monte Carlo (MCMC) sampling from $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathbf{y})$ is possible, thus stochastic approximation can be used to solve (8) by iterating the following steps:

I. `Data imputation step:` `sample` $\mathbf{z}_N \mid \mathbf{y}$.

II. `Parameter updating step:`

$$\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}}_{N-1} + a_{N-1}\, \mathbf{B}^{-1}\, \nabla_{\hat{\boldsymbol{\theta}}_{N-1}} \ln p_{\hat{\boldsymbol{\theta}}_{N-1}}(\mathbf{y}, \mathbf{z}_N),$$

where $\mathbf{z}_N$ has density $p_{\hat{\boldsymbol{\theta}}_{N-1}}(\mathbf{z}_N \mid \mathbf{y})$, $a_{N-1}$ is a sequence of positive numbers tending to 0, and $\mathbf{B}$ is a positive definite matrix of suitable order. A possible choice of $\mathbf{B}$ is given by $\mathbf{B} = -E_{\hat{\boldsymbol{\theta}}}\left[\nabla^2_{\hat{\boldsymbol{\theta}}} \ln p_{\hat{\boldsymbol{\theta}}}(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}\right]$, where $\nabla^2_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ is the Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\boldsymbol{\theta}$; $\mathbf{B}$ can be estimated by replacing the unknown MLE $\hat{\boldsymbol{\theta}}$ by an initial estimate $\tilde{\boldsymbol{\theta}}$, MCMC sampling from $p_{\tilde{\boldsymbol{\theta}}}(\mathbf{z} \mid \mathbf{y})$, and estimating the expectation by the corresponding MCMC sample average. A sensible modification of the algorithm is based on the averaging approach (cf. Yin 1991, and references therein), which is a multi-stage approach that generates in each stage interim estimates $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_N$ with $a_1 = a_2 = \cdots = a_N$ according to the iterative scheme above, and replaces $\hat{\boldsymbol{\theta}}_N$ by $(1/N)\sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_i$; stage $m \geq 2$ starts with the average of $\hat{\boldsymbol{\theta}}_i$ based on stage $m - 1$, with increased $N$ and decreased $a_N$. Under regularity conditions, such stochastic approximation estimators $\hat{\boldsymbol{\theta}}_N$ converge to the solution of (8) (cf. Yin 1991), which is the MLE $\hat{\boldsymbol{\theta}}$.

When having obtained an estimate of the MLE $\hat{\boldsymbol{\theta}}$ by stochastic approximation, the observed information matrix can be calculated as follows (Louis 1982):

$$
\begin{aligned}
&-\nabla^2_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}) = -E_{\boldsymbol{\theta}}[\nabla^2_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}] \\
&-\big\{E_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{Z})(\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{Z}))' \mid \mathbf{y}] - \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y})(\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y}))'\big\},
\end{aligned} \tag{9}
$$

where the expectations on the right-hand side of (9) can be estimated by MCMC sampling from $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathbf{y})$ and estimating the expectations by the corresponding MCMC sample averages.

To make the stochastic approximation algorithm operational, MCMC samples from $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathbf{y})$ are required, which is discussed in Section 3.1; in addition, three entities are needed in closed form, the probability density $p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ and the gradient and Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\boldsymbol{\theta}$; these are given in Section 3.2, along with a reparametrization of $\boldsymbol{\Sigma}^{-1}$.

### 3.1 Implementation: MCMC sampling

To sample $\mathbf{z} \mid \mathbf{y}$, where $\mathbf{z} = (\mathbf{V}, w)$, it is convenient to use MCMC cycle algorithms (Hastings 1970; Tierney 1994), which iterate, for a given value of $\boldsymbol{\theta}$, the following steps:

(a) Sample $w^{new} \mid \mathbf{y}, \mathbf{V}$.

(b) Sample $\mathbf{V}^{new} \mid \mathbf{y}, w^{new}$.

Steps (a) and (b) can be implemented by using the Metropolis-Hastings (M-H) method (Hastings 1970) as follows.

M-H step to sample $w^{\text{new}} \mid \mathbf{y}, \mathbf{V}$. A candidate $w^\star$ is generated from a distribution $a(w^\star \mid w)$, and $w^\star$ is accepted with probability

$$\pi = \min\left[1,\ \frac{p_{\boldsymbol{\theta}}(w^\star \mid \mathbf{y}, \mathbf{V})}{p_{\boldsymbol{\theta}}(w \mid \mathbf{y}, \mathbf{V})} \times \frac{a(w \mid w^\star)}{a(w^\star \mid w)}\right].$$

The candidate-generating distribution $a(w^\star \mid w)$ is identical to the candidate-generating distribution of Snijders et al. (2007), which is a probability distribution defined on a discrete set of simple proposals, corresponding to inserting and deleting $(i_m, j_m)$'s in the sequence $w = (i_m, j_m)_{m=1}^M$ or permuting sub-sequences of $w$, subject to the constraint that $\mathbf{X}_0 \equiv \mathbf{X}(t_0)$ and $\mathbf{X}_M \equiv \mathbf{X}(t_1)$.

M-H step to sample $\mathbf{V}^{\text{new}} \mid \mathbf{y}, w^{\text{new}}$. For $i = 1, \ldots, n$ independently, a candidate $\mathbf{V}_i^\star$ is generated from a distribution $b(\mathbf{V}_i^\star \mid \mathbf{V}_i)$, and $\mathbf{V}_i^\star$ is accepted with probability

$$\pi = \min\left[1,\ \frac{p_{\boldsymbol{\theta}}(\mathbf{V}_i^\star \mid \mathbf{y}, \mathbf{V}_{-i}, w^{\text{new}})}{p_{\boldsymbol{\theta}}(\mathbf{V}_i \mid \mathbf{y}, \mathbf{V}_{-i}, w^{\text{new}})} \times \frac{b(\mathbf{V}_i \mid \mathbf{V}_i^\star)}{b(\mathbf{V}_i^\star \mid \mathbf{V}_i)}\right],$$

where $\mathbf{V}_{-i}$ corresponds to the random effects matrix $\mathbf{V}$ without row $i$. Possible candidates are given by $\mathbf{V}_i^\star = \mathbf{a} + \mathbf{B}(\mathbf{V}_i - \mathbf{a}) + \mathbf{G}_i$, where $\mathbf{a}$ is a constant $K \times 1$ vector, $\mathbf{B}$ is a constant $K \times K$ matrix, and $\mathbf{G}_i \sim N_K(\mathbf{0}, \boldsymbol{\Omega})$. Let $\mathbf{I}_K$ be the $K \times K$ identity matrix. The choice $\mathbf{B} = \mathbf{I}_K$ corresponds to (i) random walk M-H algorithms; $\mathbf{B} = \mathbf{0}$ corresponds to (ii) independence samplers, in which case $\mathbf{a} = \mathbf{0}$ is convenient, the prior expectation of $\mathbf{V}_i$ (see Section 2.2); and $\mathbf{B} = -\mathbf{I}_K$ corresponds to (iii) first-order autoregressive M-H algorithms, which reflect $\mathbf{V}_i$ about $\mathbf{a}$ before adding increment $\mathbf{G}_i$ and can reduce MCMC autocorrelations; $\mathbf{a}$ can be calibrated during burn-in iterations. The scale matrix $\mathbf{\Omega}$ can be based on burn-in iterations or Bayesian point estimates of $\mathbf{\Sigma}$ (see Section 4).

### 3.2 Implementation: closed-form expressions

The probability density $p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ and the gradient and Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\boldsymbol{\theta}$ are derived below.

The model of Section 2—with constant rate functions $\phi_i(\mathbf{X}, \boldsymbol{\theta}) = \alpha$—implies that the probability density $p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ is given by

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) &= \alpha^M \exp[-n\alpha(t_1 - t_0)] \\ &\times \prod_{m=1}^{M} \psi_{i_m}(j_m \mid \mathbf{X}_{m-1}, \mathbf{V}_{i_m}, \boldsymbol{\theta}) \\ &\times \prod_{i=1}^{n} (2\pi)^{-K/2} \det[\mathbf{\Sigma}^{-1}]^{1/2} \exp\left[-\frac{1}{2} \mathbf{V}_i' \mathbf{\Sigma}^{-1} \mathbf{V}_i\right], \end{aligned} \tag{10}$$

where $\psi_{i_m}(j_m \mid \mathbf{X}_{m-1}, \mathbf{V}_{i_m}, \boldsymbol{\theta})$ is based on $f_{i_m}(\mathbf{X}_{m-1}, j_m, \mathbf{V}_{i_m}, \boldsymbol{\theta})$ as defined by (4)—as a natural extension of (3)—and $\det[\mathbf{\Sigma}^{-1}]$ is the determinant of $\mathbf{\Sigma}^{-1}$.

The gradient and Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\alpha$ and $\boldsymbol{\beta}$ can readily be obtained.

Concerning $\mathbf{\Sigma}^{-1}$, to impose symmetry and positive definiteness constraints on estimates of $\mathbf{\Sigma}^{-1}$, it is sensible to reparametrize the non-redundant elements of $\mathbf{\Sigma}^{-1}$ so that estimates of $\mathbf{\Sigma}^{-1}$ are by construction symmetric and positive definite.

Let $\mathbf{\Sigma}^{-1} = \mathbf{\Gamma}\mathbf{\Gamma}'$, where $\mathbf{\Gamma}$ is a $K \times K$ lower triangular matrix restricted by $\gamma_{ij} = 0$ (all $j > i$), called a Cholesky factor of $\mathbf{\Sigma}^{-1}$. Estimating $\mathbf{\Gamma}$ itself is associated with the drawback that $\mathbf{\Gamma}$ is not globally identifiable due to the fact that any column of $\mathbf{\Gamma}$ can be multiplied by $-1$ without changing the value of the likelihood.

Therefore, let $\mathbf{\Gamma} = \mathbf{\Delta}\mathbf{\Lambda}^{1/2}$, where $\mathbf{\Delta}$ is a $K \times K$ lower triangular matrix constrained by $\delta_{ii} = 1$ (all $i$) and $\delta_{ij} = 0$ (all $j > i$), $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix with elements $\lambda_{ii} = \exp[\xi_{ii}]$ on the main diagonal, and the elements $\xi_{ii}$ are stored on the main diagonal of a $K \times K$ diagonal matrix $\mathbf{\Xi}$; for related parametrizations, see Hedeker and Gibbons (1996); Pinheiro and Bates (1996). It is evident that estimating $\xi_{ii} = \ln \lambda_{ii}$ produces positive estimates of $\lambda_{ii}$ by construction, and facilitates the estimation of very small variances. To disregard the constant elements 0, 1 of $\mathbf{\Delta}$ and $\mathbf{\Xi}$, let $\mathrm{v}(\mathbf{\Delta})$ and $\mathrm{v}(\mathbf{\Xi})$ be the vectors obtained from $\mathrm{vec}(\mathbf{\Delta})$ and $\mathrm{vec}(\mathbf{\Xi})$ by eliminating all constant elements, respectively, where vec is the vec operator which transforms its matrix argument into a column vector by stacking the columns of the matrix one underneath the other; and let $\mathbf{D}_{(\mathbf{\Delta})}$ and $\mathbf{D}_{(\mathbf{\Xi})}$ be the unique matrices satisfying $\mathbf{D}_{(\mathbf{\Delta})}\,\mathrm{v}(\mathbf{\Delta}) + \mathrm{vec}(\mathbf{I}_K) = \mathrm{vec}(\mathbf{\Delta})$ and $\mathbf{D}_{(\mathbf{\Xi})}\,\mathrm{v}(\mathbf{\Xi}) = \mathrm{vec}(\mathbf{\Xi})$, respectively, where $\mathbf{I}_K$ is the $K \times K$ identity matrix; note that $\mathrm{vec}(\mathbf{\Delta})$ is an affine rather than a linear function of $\mathrm{v}(\mathbf{\Delta})$ because of the constraint $\delta_{ii} = 1$ (all $i$).

The vectors $\mathrm{v}(\mathbf{\Delta})$ and $\mathrm{v}(\mathbf{\Xi})$ are the parameters to be estimated. The gradient and Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\mathrm{v}(\mathbf{\Delta})$ and $\mathrm{v}(\mathbf{\Xi})$ are derived in the

Appendix and are given by

$$\nabla_{\mathrm{v}(\boldsymbol{\Delta})} \ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = \mathbf{D}'_{(\boldsymbol{\Delta})} \operatorname{vec}(A(\boldsymbol{\Sigma})\, \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{1/2}), \tag{11}$$

$$\nabla_{\mathrm{v}(\boldsymbol{\Xi})} \ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = \frac{1}{2}\mathbf{D}'_{(\boldsymbol{\Xi})} \operatorname{vec}(\boldsymbol{\Gamma}' A(\boldsymbol{\Sigma})\, \boldsymbol{\Gamma}), \tag{12}$$

$$\nabla^2_{\mathrm{v}(\boldsymbol{\Delta})\,\mathrm{v}(\boldsymbol{\Xi})} \ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = \begin{pmatrix} \nabla^2_{\boldsymbol{\Delta}\boldsymbol{\Delta}} & \nabla^2_{\boldsymbol{\Delta}\boldsymbol{\Xi}} \\ \nabla^2_{\boldsymbol{\Xi}\boldsymbol{\Delta}} & \nabla^2_{\boldsymbol{\Xi}\boldsymbol{\Xi}} \end{pmatrix}, \tag{13}$$

where

$$\nabla^2_{\boldsymbol{\Delta}\boldsymbol{\Delta}} = -\mathbf{D}'_{(\boldsymbol{\Delta})} \left[n\mathbf{E}_K((\boldsymbol{\Delta}^{-1})' \otimes \boldsymbol{\Delta}^{-1}) + (\boldsymbol{\Lambda} \otimes \mathbf{H})\right] \mathbf{D}_{(\boldsymbol{\Delta})},$$

$$\nabla^2_{\boldsymbol{\Delta}\boldsymbol{\Xi}} = (\nabla^2_{\boldsymbol{\Xi}\boldsymbol{\Delta}})' = -2\mathbf{D}'_{(\boldsymbol{\Delta})}(\boldsymbol{\Lambda} \otimes \mathbf{H}\boldsymbol{\Delta})\mathbf{D}_{(\boldsymbol{\Xi})},$$

$$\nabla^2_{\boldsymbol{\Xi}\boldsymbol{\Xi}} = -\frac{1}{2}\mathbf{D}'_{(\boldsymbol{\Xi})}(\boldsymbol{\Lambda} \otimes \boldsymbol{\Delta}'\mathbf{H}\boldsymbol{\Delta})\mathbf{D}_{(\boldsymbol{\Xi})},$$

where $A(\boldsymbol{\Sigma}) = n\boldsymbol{\Sigma} - \mathbf{H}$ and $\mathbf{H} = \sum_{i=1}^{n} \mathbf{V}_i\mathbf{V}_i'$, while $\mathbf{E}_K$ is the unique permutation matrix satisfying $\mathbf{E}_K \operatorname{vec}(\mathbf{C}) = \operatorname{vec}(\mathbf{C}')$ for every $K \times K$ matrix $\mathbf{C}$ (cf. Magnus 1988, p. 35), and $\otimes$ is the Kronecker product.

The blocks of the Hessian matrix $\nabla^2_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ involving $\alpha$, $\boldsymbol{\beta}$ on one hand and $\mathrm{v}(\boldsymbol{\Delta})$, $\mathrm{v}(\boldsymbol{\Xi})$ on the other hand vanish.

When having obtained an estimate of the MLE $\hat{\boldsymbol{\theta}}$ by stochastic approximation, the observed information matrix (9) can be estimated. To estimate the observed information matrix for interesting functions of $\mathrm{v}(\boldsymbol{\Delta})$ and $\mathrm{v}(\boldsymbol{\Xi})$ such as $\mathrm{v}(\boldsymbol{\Sigma})$, where $\mathrm{v}(\boldsymbol{\Sigma})$ is obtained from $\operatorname{vec}(\boldsymbol{\Sigma})$ by eliminating the elements $\sigma_{ij}$ (all $j > i$), the gradient and Hessian matrix of $\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Sigma})$ are required, which are derived in the Appendix and are given by

$$\nabla_{\mathrm{v}(\boldsymbol{\Sigma})} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) \;=\; -\frac{1}{2}\mathbf{D}'_{(\boldsymbol{\Sigma})} \operatorname{vec}(\boldsymbol{\Sigma}^{-1} A(\boldsymbol{\Sigma})\, \boldsymbol{\Sigma}^{-1}), \tag{14}$$

$$\nabla^2_{\mathrm{v}(\boldsymbol{\Sigma})} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) \;=\; -\mathbf{D}'_{(\boldsymbol{\Sigma})} \left[(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{H}\boldsymbol{\Sigma}^{-1}) - \frac{n}{2}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right] \mathbf{D}_{(\boldsymbol{\Sigma})}, \tag{15}$$

where $\mathbf{D}_{(\boldsymbol{\Sigma})}$ is the unique matrix satisfying $\mathbf{D}_{(\boldsymbol{\Sigma})}\, \mathrm{v}(\boldsymbol{\Sigma}) = \operatorname{vec}(\boldsymbol{\Sigma})$ for every symmetric $K \times K$ matrix $\boldsymbol{\Sigma}$ (cf. Magnus 1988, p. 55).

## 4. BAYESIAN ESTIMATION

Bayesian inference concerning $\boldsymbol{\theta}$ is based on the posterior probability density $p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p_{\boldsymbol{\theta}}(\mathbf{y})\, p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$; in most applications, it is reasonable to start with the assumption of prior independence of $\alpha$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}^{-1}$; convenient families of prior distributions are

$$\begin{aligned} \alpha &\sim \mathrm{Gamma}(\gamma, \delta), \\ \boldsymbol{\beta} &\sim N_L(\mathbf{0}, \boldsymbol{\Psi}), \\ \boldsymbol{\Sigma}^{-1} &\sim \mathrm{Wishart}(\varphi, \boldsymbol{\Omega}). \end{aligned}$$

The posterior distribution is intractable, but samples from the posterior distribution can be obtained by combining the following Gibbs and M-H steps by means of cycling or mixing (cf. Hastings 1970; Tierney 1994):

M-H step to sample $\mathbf{z} \mid \alpha, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, \mathbf{y}$. One iteration of the cycle algorithm sketched in Section 3.1 is sufficient.

Gibbs sampling of $\alpha \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, \mathbf{y}, \mathbf{z}$. If the prior of $\alpha$ is Gamma$(\gamma, \delta)$ parametrized such that $E_{\gamma,\delta}[\alpha] = \gamma/\delta$, then the full conditional posterior of $\alpha$ is Gamma$(\gamma + M, \delta + n(t_1 - t_0))$, which can be sampled.

M-H step to sample $\boldsymbol{\beta} \mid \alpha, \boldsymbol{\Sigma}^{-1}, \mathbf{y}, \mathbf{z}$. A candidate $\boldsymbol{\beta}^\star$ is generated from a distribution $b(\boldsymbol{\beta}^\star \mid \boldsymbol{\beta})$, and $\boldsymbol{\beta}^\star$ is accepted with probability

$$\pi = \min\left[1, \frac{p(\boldsymbol{\beta}^\star \mid \alpha, \boldsymbol{\Sigma}^{-1}, \mathbf{y}, \mathbf{z})}{p(\boldsymbol{\beta} \mid \alpha, \boldsymbol{\Sigma}^{-1}, \mathbf{y}, \mathbf{z})} \times \frac{b(\boldsymbol{\beta} \mid \boldsymbol{\beta}^\star)}{b(\boldsymbol{\beta}^\star \mid \boldsymbol{\beta})}\right].$$

Convenient candidates are given by $\boldsymbol{\beta}^\star = \mathbf{a} + \mathbf{B}(\boldsymbol{\beta} - \mathbf{a}) + \mathbf{G}$, where $\mathbf{a}$ is a constant $L \times 1$ vector, $\mathbf{B}$ is a constant $L \times L$ matrix, and $\mathbf{G} \sim N_L(\mathbf{0}, \boldsymbol{\Omega})$. Let $\mathbf{I}_L$ be the $L \times L$ identity matrix. Possible choices of $\mathbf{B}$ are (i) $\mathbf{B} = \mathbf{I}_L$ (random walk M-H algorithms); (ii) $\mathbf{B} = \mathbf{0}$ (independence samplers); and (iii) $\mathbf{B} = -\mathbf{I}_L$ (first-order autoregressive M-H algorithms), which reflects $\boldsymbol{\beta}$ about $\mathbf{a}$ before adding increment $\mathbf{G}$ and can reduce MCMC autocorrelations. In case of (ii) and (iii), $\mathbf{a}$ can be based on point estimates of $\boldsymbol{\beta}$, such as method of moments estimates (see Snijders 2001, assuming $\boldsymbol{\Sigma} \equiv \mathbf{0}$) or the MLE (see Section 3). The scale matrix $\boldsymbol{\Omega}$ can be based on burn-in iterations or the inverse observed information matrix of $\boldsymbol{\beta}$ at the MLE of $\boldsymbol{\beta}$ (see Section 3).

Gibbs sampling of $\boldsymbol{\Sigma}^{-1} \mid \alpha, \boldsymbol{\beta}, \mathbf{y}, \mathbf{z}$. If the prior of $\boldsymbol{\Sigma}^{-1}$ is Wishart$(\varphi, \boldsymbol{\Omega})$ parametrized such that $E_{\varphi,\boldsymbol{\Omega}}[\boldsymbol{\Sigma}^{-1}] = \varphi\,\boldsymbol{\Omega}$, then the full conditional posterior is Wishart$(\varphi + n, (\mathbf{H} + \boldsymbol{\Omega}^{-1})^{-1})$ ($\mathbf{H} = \sum_{i=1}^{n} \mathbf{V}_i \mathbf{V}_i'$), which can be sampled.

## 5. APPLICATION

The model is applied to data collected as part of the Teenage Friends and Lifestyle Study (Pearson and West 2003), corresponding to friendships among 160 students of a Scottish school cohort observed at three time points $t_0 < t_1 < t_2$ between 1995 and 1997. 129 students were present at all three time points, among which 56 girls. Here, the friendships among the $n = 56$ girls are studied, corresponding to $56 \times 56$ matrices $\mathbf{X}(t_0)$, $\mathbf{X}(t_1)$, and $\mathbf{X}(t_2)$, where $X_{ij}(t_h) = 1$ if girl $i$ called girl $j$ a friend at time point $t_h$, and $X_{ij}(t_h) = 0$ otherwise. At time point $t_0$, 156 of $56(56-1) = 3{,}080$ possible arcs were observed; in time interval $[t_{h-1}, t_h]$ $(h = 1, 2)$, 79 and 65 arcs were added while 75 and 68 arcs were deleted, respectively.

Let $\mathbf{X}$ be the digraph at time $t \in [t_0, t_2]$ and suppose that $\mathbf{X}$ is allowed to change, let $i$ be the girl that is allowed to change something, let $j$ be the girl chosen by $i$, let $\mathbf{X}^\star$ be the digraph with elements $X_{ij}^\star = X_{ij}$ if $i = j$ and $X_{ij}^\star = 1 - X_{ij}$ otherwise, while $X_{kl}^\star = X_{kl}$ (all $(k,l) \neq (i,j)$), and let $\mathbf{I}_K$ be the $K \times K$ identity matrix. A simple model is constructed by using rate functions $\phi_i(\mathbf{X}, \boldsymbol{\theta}) = \alpha_h$ in time interval $[t_{h-1}, t_h]$ $(h = 1, 2)$ and assuming that the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are constant across time intervals. Interesting components $s_{ik}(\mathbf{X}, j)$ of statistics vector $s_i(\mathbf{X}, j)$ in objective function $f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta})$ are given by

$s_{i1}(\mathbf{X}, j) = \sum_{l=1}^{n} X_{il}^\star$ (number of arcs),

$s_{i2}(\mathbf{X}, j) = \sum_{l=1}^{n} X_{il}^\star X_{li}^\star$ (number of reciprocated arcs),

$s_{i3}(\mathbf{X}, j) = \sum_{h=1}^{n} \sum_{l=1, l \neq h}^{n} X_{ih}^\star X_{hl}^\star X_{il}^\star$ (number of transitive triplets),

$s_{i4}(\mathbf{X}, j) = \sum_{l=1}^{n} X_{il}^{\star}\, c_l(t_{h-1})$ (number of arcs weighted by covariate $c_l(t_{h-1})$),

which—by multiplying (3) by $g_i/g_i$, where $g_i = \exp[-f_i(\mathbf{X}, i, \mathbf{V}_i, \boldsymbol{\theta})]$—is equivalent to specifying $f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta})$ as

$$f_i(\mathbf{X}, j, \mathbf{V}_i, \boldsymbol{\theta}) \;\; = \;\; \gamma_{ij}\, (X_{ij}^{\star} - X_{ij}), \tag{16}$$

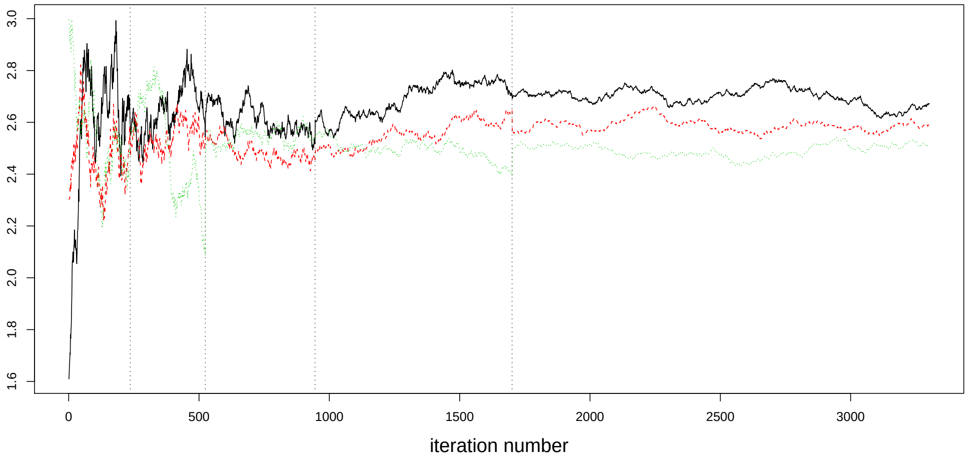
where the weight $\gamma_{ij}$ is given by

$$\gamma_{ij} \;\; = \;\; \eta_{i1} + \eta_{i2}\, X_{ji} + \eta_{i3}\, (\textstyle\sum_{h=1, h \neq j}^{n} X_{ih} X_{hj} + \sum_{l=1, l \neq j}^{n} X_{il} X_{jl}) + \eta_{i4}\, c_j(t_{h-1}).$$

The covariate $c_j(t_{h-1})$ refers to the amount of financial resources of girl $j$ at time point $t_{h-1}$ ($h = 1, 2$), which may be regarded as an indicator of social-economic status; in time interval $[t_{h-1}, t_h]$, $s_{i4}(\mathbf{X}, j)$ was computed by using $c_j(t_{h-1})$, and $c_j(t_{h-1})$ was centered at 0 and rescaled to variance 1. The most important candidate for girl-dependent weights $\eta_{ik}$ is $\eta_{i1}$, because $\eta_{i1}$ captures one of the most fundamental features of the data—the distribution of the number of arcs $\sum_{l=1}^{n} X_{il}(t_h)$ at time point $t_h$ ($h = 1, 2$)—and almost all models used in applications include $\eta_{i1}$; to keep the model as simple and parsimonious as possible, the remaining weights $\eta_{ik}$ are assumed to be constant across girls, $\eta_{ik} = \beta_k$ ($k = 2, 3, 4$). In the Bayesian framework, the priors are (a) Gamma$(1, 10^{-10})$ for $\alpha_h$ ($h = 1, 2$); (b) $N_4(\mathbf{0}, 10^{10}\, \mathbf{I}_4)$ for $\boldsymbol{\beta}$; and (c) Wishart$(1, 10\, \mathbf{I}_1)$ for $\boldsymbol{\Sigma}^{-1}$, implying that large variances are a priori unlikely, motivated by the fact that the weights $\boldsymbol{\eta}_i = \boldsymbol{\beta} + \mathbf{A}\, \mathbf{V}_i$ enter the model through exponential functions (see (3)) and therefore large weights—and thus large variances—are implausible.

The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ was estimated by a five-stage stochastic approximation algorithm with 3,300 iterations in total, as sketched in Section 3 for random effects models and more detailed in Snijders et al. (2007) for fixed effects models. As starting values of $\boldsymbol{\Sigma} \equiv \sigma^2$, .20, .10, and .05 were used, which were thought to be reasonable values. A trace plot of the interim estimates of the MLE $\hat{\xi}$ of $\xi = -\ln(\sigma^2)$ is shown in Figure 1; note that at the end of each stage (indicated by a vertical line), the present interim estimate is replaced by the average of interim estimates of the preceding stage, and the final stochastic approximation estimate $\hat{\xi}_N$ is the average of interim estimates of the fifth stage. Figure 1 shows that, from the second stage on, most interim

**Figure 1: Trace plot of interim estimates of MLE $\hat{\xi}$ using five-stage stochastic approximation algorithms with multiple starting values**

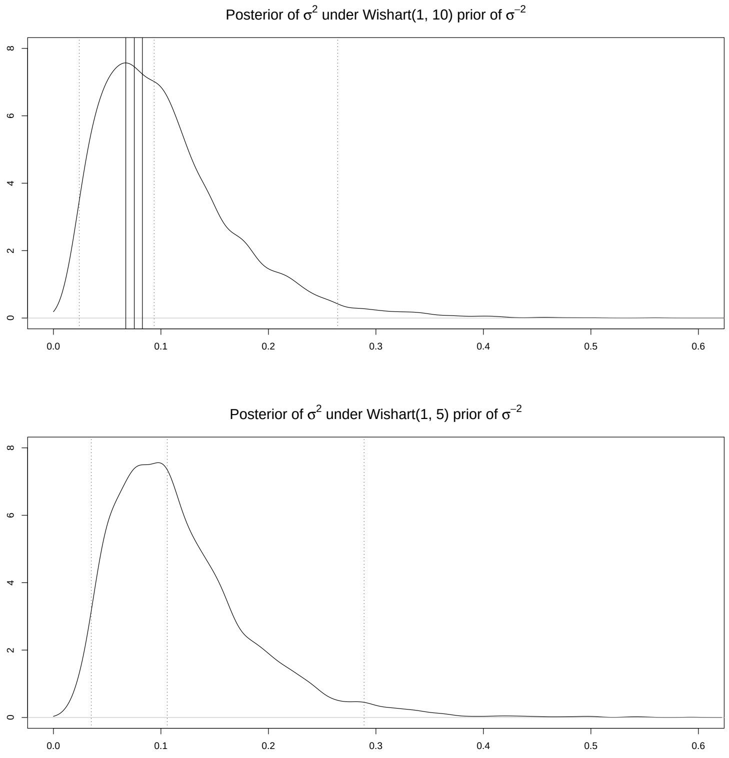


A vertical line indicates the end of a stage, at which point the present interim estimate is replaced by the average of interim estimates of the preceding stage.

estimates are in a neighborhood of 2.6. To detect non-convergence of the stochastic approximation estimates $\hat{\xi}_N$, an alternative to using multiple starting values is based on the Fisher identity (7): the so-called $t$-ratio of the average of the complete-data score $\nabla_{\hat{\xi}_N} \ln p_{\hat{\boldsymbol{\theta}}_N}(\mathbf{y}, \mathbf{z})$ across augmented sets of data $(\mathbf{y}, \mathbf{z})$ to its standard deviation can be used to diagnose non-convergence; for each of the three estimation runs, 2,000 augmented sets of data $(\mathbf{y}, \mathbf{z})$ were generated conditional on $\mathbf{y}$ and the final stochastic approximation estimate $\hat{\boldsymbol{\theta}}_N$ of $\hat{\boldsymbol{\theta}}$, and the $t$-ratio was computed. The estimation runs with starting values .20, .10, and .05 produced $t$-ratios .002, .012, and .020, respectively, which are small enough for practical purposes and do not suggest non-convergence.

It is interesting to see in what region of the marginal posterior density of $\sigma^2$ the three resulting estimates of the MLE $\hat{\sigma}^2$ of $\sigma^2$ are located. The marginal posterior density of $\sigma^2$ was estimated by the Bayesian methods of Section 4 with 220,000 iterations, where the first 20,000 iterations were discarded as burn-in iterations and every 40th sampled value of the last 200,000 iterations was recorded; non-convergence was checked by using trace plots and convergence checks of Raftery and Lewis (1996). The marginal posterior density of $\sigma^2$ is shown in Figure 2, and it is evident that the three estimates of $\hat{\sigma}^2$ are close to the mode of the marginal posterior density of $\sigma^2$. However, a relevant question is how sensitive the posterior is to the prior. An alternative to the chosen Wishart$(1, 10\, \mathbf{I}_1)$ prior of $\sigma^{-2}$ is Wishart$(1, 5\, \mathbf{I}_1)$: the resulting marginal posterior of $\sigma^2$ is depicted in Figure 2, and a tentative conclusion is that the posterior seems to be not too sensitive to the prior (within the Wishart

## Figure 2: Marginal posterior density of $\sigma^2$ under two priors



——— estimates of MLE $\hat{\sigma}^2$; $\cdots\cdots$ .025, .500, and .975 posterior quantiles of $\sigma^2$.

family of priors).

To interpret the parameters, 95% posterior intervals, the posterior median, and the MLE of $\boldsymbol{\theta}$ are shown in Table 1. The evidence with respect to $\boldsymbol{\beta}$ may be interpreted as follows: if an average girl $i$—average in the sense that $V_{i1} = E[V_{i1}] = 0$ (the prior expectation of $V_{i1}$, see Section 2.2) and thus $\eta_{i1} = \beta_1$—does not consider girl $j$ to be a friend ($X_{ij} = 0$), then $i$ will not tend to establish a friendship with $j$ (negative values of $\beta_1$) unless there are good reasons (cf. (16)): such as $j$ considering $i$ to be a friend ($X_{ji} = 1$), $j$ being the friend of at least one friend of $i$ ($X_{ih}X_{hj} = 1$), $j$ sharing

**Table 1: Estimates of $\boldsymbol{\theta}$**

| | 95% posterior interval | posterior median | MLE | (s.e.) |
|---|---|---|---|---|
| $\alpha_1$ | $[5.000, 7.861]$ | 6.294 | 6.421 | (.725) |
| $\alpha_2$ | $[4.193, 6.994]$ | 5.370 | 5.389 | (.653) |
| $\beta_1$ | $[-2.831, -2.402]$ | $-2.600$ | $-2.560$ | (.104) |
| $\beta_2$ | $[1.776, 2.419]$ | 2.089 | 2.042 | (.171) |
| $\beta_3$ | $[.351, .503]$ | .424 | .410 | (.036) |
| $\beta_4$ | $[.003, .233]$ | .121 | .118 | (.056) |
| $\sigma^2$ | $[.024, .264]$ | .094 | .067 | (.055) |

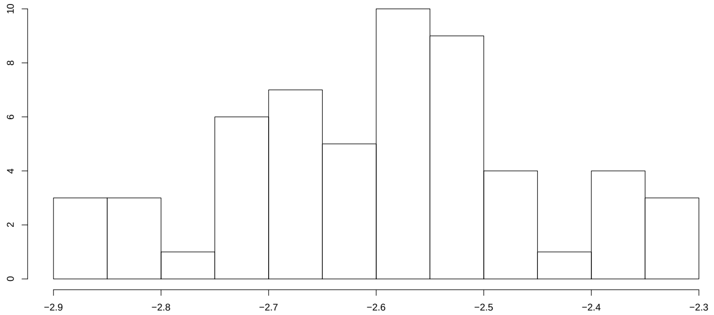
Bayesian estimates are based on the Wishart$(1, 10\,\mathbf{I}_1)$ prior of $\sigma^{-2}$; the MLE of $\boldsymbol{\theta}$ is based on the estimation run with .20 as starting value of $\sigma^2$.

at least one friend with $i$ ($X_{il}X_{jl} = 1$), or $j$ having high social-economic status—all increasing the likelihood that $i$ establishes a friendship to $j$; if $i$ considers $j$ to be a friend ($X_{ij} = 1$), then the same mechanisms decrease the likelihood that $i$ cancels the friendship to $j$. The magnitude of $\sigma^2$ hints that there is non-negligible variation among girls with respect to the weights $\eta_{i1} = \beta_1 + V_{i1}$, but at the same time the weights $\eta_{i1}$ still appear to be negative, which is confirmed by the posterior medians of $\eta_{i1}$ ($i = 1, \ldots, n$) shown in Figure 3.

## 6. DISCUSSION

A family of models for digraph panel data was proposed, which represents unobserved heterogeneity across nodes by random variables with unobserved outcomes

**Figure 3: Histogram of posterior medians of $\eta_{i1}$ ($i = 1, \ldots, n$) under Wishart$(1, 10\,\mathbf{I}_1)$ prior of $\sigma^{-2}$**



(random effects), and parameter estimation was considered in a maximum likelihood and Bayesian framework.

Concerning maximum likelihood estimation, the use of the inverse observed information matrix as an approximation of the variance-covariance matrix of the MLE makes implicit use of large-sample theory. The asymptotic properties of the MLE are unknown, and deriving them is beyond the scope of the present paper. However, for many social network models properties of estimators are still unknown, and much more work needs to be done to clarify theoretical issues arising from the application of statistical models to social network data; see, for instance, the discussion of Hunter and Handcock (2006) in the framework of curved exponential random graph models for social networks.

Another important area of future research is the development of model selection tools: while posterior predictive checks (under fixed effects models) can be used (a) to assess how well (posterior) predicted data match the observed data, using suitable summary statistics, and (b) to suggest what random effects should be included, the choice of summary statistics is arbitrary, and, more importantly, the increase in model complexity by adding random effects is not taken into account.

## APPENDIX: GRADIENTS AND HESSIAN MATRICES

Using the notation of Section 3.2, the gradients and Hessian matrices of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Delta})$, $\mathrm{v}(\boldsymbol{\Xi})$, and $\mathrm{v}(\boldsymbol{\Sigma})$ are derived based on the theory of matrix differential calculus of Magnus and Neudecker (1988), abbreviated as "MN". Twice differentiability is assumed throughout.

Observe that $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ can be written as

$$\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) = \frac{n}{2} \ln \det[\boldsymbol{\Sigma}^{-1}] - \frac{1}{2} \operatorname{tr}(\mathbf{H}\boldsymbol{\Sigma}^{-1}) + c,$$

where $c$ is a scalar not depending on $\boldsymbol{\Sigma}^{-1}$, and $\operatorname{tr}(\mathbf{H}\boldsymbol{\Sigma}^{-1})$ is the trace of $\mathbf{H}\boldsymbol{\Sigma}^{-1}$.

*Lemma 1.* The first differential of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Delta})$ and $\mathrm{v}(\boldsymbol{\Xi})$ is given by

$$\begin{aligned} \mathrm{d} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) &= (\operatorname{vec}(A(\boldsymbol{\Sigma})\, \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{1/2}))'\, \mathbf{D}_{(\boldsymbol{\Delta})}\, \mathrm{d}\, \mathrm{v}(\boldsymbol{\Delta}) \\ &+ \frac{1}{2} (\operatorname{vec}(\boldsymbol{\Gamma}' A(\boldsymbol{\Sigma})\, \boldsymbol{\Gamma}))'\, \mathbf{D}_{(\boldsymbol{\Xi})}\, \mathrm{d}\, \mathrm{v}(\boldsymbol{\Xi}). \end{aligned}$$

*Proof of Lemma 1.* Using $\mathrm{d} \ln \det[\boldsymbol{\Sigma}^{-1}] = \operatorname{tr}(\boldsymbol{\Sigma}\, \mathrm{d}\, \boldsymbol{\Sigma}^{-1})$ (MN, Theorem 2, pp. 150–151),

$$\mathrm{d} \ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z}) = \frac{n}{2} \mathrm{d} \ln \det[\boldsymbol{\Sigma}^{-1}] - \frac{1}{2} \operatorname{tr}(\mathbf{H}\, \mathrm{d}\, \boldsymbol{\Sigma}^{-1}) = \frac{1}{2} \operatorname{tr}(A(\boldsymbol{\Sigma})\ \mathrm{d}\, \boldsymbol{\Sigma}^{-1}).$$

Using $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Delta}'$, applying Cauchy's rule of invariance (MN, Theorem 13, p. 96), and using $\mathrm{tr}(\mathbf{B}\mathbf{C}) = (\mathrm{vec}(\mathbf{B}'))' \,\mathrm{vec}(\mathbf{C})$ for $K \times K$ matrices $\mathbf{B}$, $\mathbf{C}$,

$$\begin{aligned} \mathrm{d}\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) &= \frac{1}{2}\,\mathrm{tr}(A(\boldsymbol{\Sigma})\ \mathrm{d}(\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Delta}')) \\ &= (\mathrm{vec}(A(\boldsymbol{\Sigma})\,\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{1/2}))'\,\mathrm{d}\,\mathrm{vec}(\boldsymbol{\Delta}) + \frac{1}{2}(\mathrm{vec}(\boldsymbol{\Gamma}'A(\boldsymbol{\Sigma})\,\boldsymbol{\Gamma}))'\,\mathrm{d}\,\mathrm{vec}(\boldsymbol{\Xi}). \end{aligned}$$

Using $\mathrm{vec}(\boldsymbol{\Delta}) = \mathbf{D}_{(\boldsymbol{\Delta})}\,\mathrm{v}(\boldsymbol{\Delta}) + \mathrm{vec}(\mathbf{I}_K)$ and $\mathrm{vec}(\boldsymbol{\Xi}) = \mathbf{D}_{(\boldsymbol{\Xi})}\,\mathrm{v}(\boldsymbol{\Xi})$ and applying Cauchy's rule of invariance completes the proof. □

*Lemma 2.* The first differential of $\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Sigma})$ is given by

$$\mathrm{d}\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = -\frac{1}{2}(\mathrm{vec}(\boldsymbol{\Sigma}^{-1}A(\boldsymbol{\Sigma})\,\boldsymbol{\Sigma}^{-1}))'\,\mathbf{D}_{(\boldsymbol{\Sigma})}\,\mathrm{d}\,\mathrm{v}(\boldsymbol{\Sigma}).$$

*Proof of Lemma 2.* Using $\mathrm{d}\,\boldsymbol{\Sigma}^{-1} = -\boldsymbol{\Sigma}^{-1}(\mathrm{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}$ (MN, Theorem 3, p. 151),

$$\begin{aligned} \mathrm{d}\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) &= -\frac{n}{2}\,\mathrm{d}\ln\det[\boldsymbol{\Sigma}] - \frac{1}{2}\,\mathrm{tr}(\mathbf{H}\,\mathrm{d}\,\boldsymbol{\Sigma}^{-1}) \\ &= -\frac{1}{2}(\mathrm{vec}(\boldsymbol{\Sigma}^{-1}A(\boldsymbol{\Sigma})\,\boldsymbol{\Sigma}^{-1}))'\,\mathrm{d}\,\mathrm{vec}(\boldsymbol{\Sigma}). \end{aligned}$$

Using $\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{D}_{(\boldsymbol{\Sigma})}\,\mathrm{v}(\boldsymbol{\Sigma})$ and applying Cauchy's rule of invariance completes the proof. □

*Lemma 3.* The second differential of $\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Delta})$ and $\mathrm{v}(\boldsymbol{\Xi})$ is given by

$$\begin{aligned} \mathrm{d}^2\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) &= -(\mathrm{d}\,\mathrm{v}(\boldsymbol{\Delta}))'\,\mathbf{D}'_{(\boldsymbol{\Delta})}\,[n\mathbf{E}_K((\boldsymbol{\Delta}^{-1})' \otimes \boldsymbol{\Delta}^{-1}) + (\boldsymbol{\Lambda} \otimes \mathbf{H})]\,\mathbf{D}_{(\boldsymbol{\Delta})}\,\mathrm{d}\,\mathrm{v}(\boldsymbol{\Delta}) \\ &\quad -2(\mathrm{d}\,\mathrm{v}(\boldsymbol{\Delta}))'\,\mathbf{D}'_{(\boldsymbol{\Delta})}\,(\boldsymbol{\Lambda} \otimes \mathbf{H}\boldsymbol{\Delta})\,\mathbf{D}_{(\boldsymbol{\Xi})}\,\mathrm{d}\,\mathrm{v}(\boldsymbol{\Xi}) \\ &\quad -\frac{1}{2}(\mathrm{d}\,\mathrm{v}(\boldsymbol{\Xi}))'\,\mathbf{D}'_{(\boldsymbol{\Xi})}\,(\boldsymbol{\Lambda} \otimes \boldsymbol{\Delta}'\mathbf{H}\boldsymbol{\Delta})\,\mathbf{D}_{(\boldsymbol{\Xi})}\,\mathrm{d}\,\mathrm{v}(\boldsymbol{\Xi}). \end{aligned}$$

*Proof of Lemma 3.* Note that

$$\mathrm{d}^2\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = \frac{n}{2}\,\mathrm{d}(\mathrm{d}\ln\det[\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Delta}']) - \frac{1}{2}\,\mathrm{d}(\mathrm{d}\,\mathrm{tr}(\boldsymbol{\Delta}'\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Lambda})).$$

One can show that

$$
\begin{aligned}
\mathrm{d}(\mathrm{d}\ln\det[\mathbf{\Delta\Lambda\Delta}']) &= -2\,\mathrm{tr}(\mathbf{\Delta}^{-1}(\mathrm{d}\,\mathbf{\Delta})\mathbf{\Delta}^{-1}\,\mathrm{d}\,\mathbf{\Delta}),\\
\mathrm{d}(\mathrm{d}\,\mathrm{tr}(\mathbf{\Delta}'\mathbf{H}\mathbf{\Delta\Lambda})) &= 2\,\mathrm{tr}(\mathbf{\Lambda}(\mathrm{d}\,\mathbf{\Delta})'\mathbf{H}\,\mathrm{d}\,\mathbf{\Delta}) + 4\,\mathrm{tr}(\mathbf{\Lambda}(\mathrm{d}\,\mathbf{\Delta})'\mathbf{H}\mathbf{\Delta}\,\mathrm{d}\,\mathbf{\Xi})\\
&\quad + \mathrm{tr}(\mathbf{\Lambda}(\mathrm{d}\,\mathbf{\Xi})\mathbf{\Delta}'\mathbf{H}\mathbf{\Delta}\,\mathrm{d}\,\mathbf{\Xi}).
\end{aligned}
$$

Collecting terms and using $\mathrm{vec}(\mathbf{\Delta}') = \mathbf{E}_K\,\mathrm{vec}(\mathbf{\Delta})$ and $\mathbf{E}_K = \mathbf{E}_K'$ as well as $\mathrm{vec}(\mathbf{BCD}) = (\mathbf{D}'\otimes\mathbf{B})\,\mathrm{vec}(\mathbf{C})$ for $K\times K$ matrices $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, one obtains

$$
\begin{aligned}
\mathrm{d}^2\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) &= -(\mathrm{d}\,\mathrm{vec}(\mathbf{\Delta}))'\,[n\mathbf{E}_K((\mathbf{\Delta}^{-1})'\otimes\mathbf{\Delta}^{-1}) + (\mathbf{\Lambda}\otimes\mathbf{H})]\,\mathrm{d}\,\mathrm{vec}(\mathbf{\Delta})\\
&\quad -2(\mathrm{d}\,\mathrm{vec}(\mathbf{\Delta}))'\,(\mathbf{\Lambda}\otimes\mathbf{H}\mathbf{\Delta})\,\mathrm{d}\,\mathrm{vec}(\mathbf{\Xi})\\
&\quad -\frac{1}{2}(\mathrm{d}\,\mathrm{vec}(\mathbf{\Xi}))'\,(\mathbf{\Lambda}\otimes\mathbf{\Delta}'\mathbf{H}\mathbf{\Delta})\,\mathrm{d}\,\mathrm{vec}(\mathbf{\Xi}).
\end{aligned}
$$

Since $\mathrm{vec}(\mathbf{\Delta})$ and $\mathrm{vec}(\mathbf{\Xi})$ are affine functions of $\mathrm{v}(\mathbf{\Delta})$ and $\mathrm{v}(\mathbf{\Xi})$, respectively, Theorem 11 of MN (p. 112) can be invoked, which completes the proof.□

*Lemma 4.* The second differential of $\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ with respect to $\mathrm{v}(\mathbf{\Sigma})$ is given by

$$
\mathrm{d}^2\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) = -(\mathrm{d}\,\mathrm{v}(\mathbf{\Sigma}))'\mathbf{D}_{(\mathbf{\Sigma})}'\left[(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{\Sigma}^{-1}) - \frac{n}{2}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\right]\mathbf{D}_{(\mathbf{\Sigma})}\,\mathrm{d}\,\mathrm{v}(\mathbf{\Sigma}).
$$

*Proof of Lemma 4.*

$$
\begin{aligned}
\mathrm{d}^2\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z}) &= \frac{n}{2}\,\mathrm{tr}(\mathbf{\Sigma}^{-1}(\mathrm{d}\,\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\,\mathrm{d}\,\mathbf{\Sigma}) - \mathrm{tr}(\mathbf{\Sigma}^{-1}(\mathrm{d}\,\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{\Sigma}^{-1}\,\mathrm{d}\,\mathbf{\Sigma})\\
&= -(\mathrm{d}\,\mathrm{vec}(\mathbf{\Sigma}))'\left[(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{\Sigma}^{-1}) - \frac{n}{2}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\right]\mathrm{d}\,\mathrm{vec}(\mathbf{\Sigma}).
\end{aligned}
$$

Since $\mathrm{vec}(\mathbf{\Sigma})$ is a linear function of $\mathrm{v}(\mathbf{\Sigma})$, Theorem 11 of MN (p. 112) applies, completing the proof.□

The gradients of $\ln p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{z})$ with respect to $\mathrm{v}(\mathbf{\Delta})$, $\mathrm{v}(\mathbf{\Xi})$, and $\mathrm{v}(\mathbf{\Sigma})$ follow from the first identification theorem of MN (Theorem 6, p. 87) and Lemmas 1 and 2, and are given by (11), (12), and (14), respectively. The Hessian matrices of $\ln p_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{z})$ with respect to $\mathrm{v}(\boldsymbol{\Delta})$, $\mathrm{v}(\boldsymbol{\Xi})$, and $\mathrm{v}(\boldsymbol{\Sigma})$ follow from the second identification theorem of MN (Theorem 6, p. 107) and Lemmas 3 and 4, and are given by (13) and (15), respectively.

## REFERENCES

Chen, H. F. (2002), *Stochastic Approximation and its Applications*, Dordrecht: Kluwer Academic.

Cox, D. R. (1990), "Role of Models in Statistical Analysis," *Statistical Science*, 5, 169–174.

Efron, B. (1977), "Discussion on the Paper by Professor Dempster et al.," *Journal of the Royal Statistical Society, Ser. B*, 39, 29.

Fisher, R. A. (1925), "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, 22, 700–725.

Gu, M. G., and Kong, F. H. (1998), "A Stochastic Approximation Algorithm with Markov Chain Monte Carlo Method for Incomplete Data Estimation Problems," in *Proceedings of the National Academy of Sciences USA*, Vol. 95, pp. 7270–7274.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and their Applications," *Biometrika*, 57, 97–109.

Hedeker, D., and Gibbons, R. D. (1996), "MIXREG: A Computer Program for Mixed-Effects Regression Analysis with Autocorrelated Errors," *Computer Methods and Programs in Biomedicine*, 49, 229–252.

Hoff, P. D. (2005), "Bilinear Mixed-Effects Models for Dyadic Data," *Journal of the American Statistical Association*, 100, 286–295.

Holland, P. W., and Leinhardt, S. (1976), "Local Structure in Social Networks," *Sociological Methodology*, 1–45.

— (1977), "A Dynamic Model for Social Networks," *Journal of Mathematical Sociology*, 5, 5–20.

Hunter, D. R., and Handcock, M. S. (2006), "Inference in Curved Exponential Family Models for

Networks," *Journal of Computational and Graphical Statistics*, 15, 565–583.

Jones, J. H., and Handcock, M. S. (2003), "Social Networks: Sexual Contacts and Epidemic Thresholds," *Nature*, 423, 605–606.

Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer, 3rd ed.

Longford, N. T. (1993), *Random Coefficient Models*, Oxford: Oxford University Press.

Louis, T. A. (1982), "Finding Observed Information Using the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 44, 98–130.

Magnus, J. R. (1988), *Linear Structures*, New York: Oxford University Press.

Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. Zarembka, P., New York: Academic Press, pp. 105–142.

Pearson, M., and West, P. (2003), "Drifting Smoke Rings: Social Network Analysis and Markov Processes in a Longitudinal Study of Friendship Groups and Risk-Taking," *Connections*, 25, 59–76.

Pinheiro, J. C., and Bates, D. M. (1996), "Unconstrained Parametrizations for Variance-Covariance Matrices," *Statistics and Computing*, 6, 289–296.

Raftery, A. E., and Lewis, S. M. (1996), "Implementing MCMC," in *Markov chain Monte Carlo in Practice*, eds. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., London: Chapman & Hall, Chap. 7, pp. 115–130.

Raudenbush, S. W., and Bryk, A. S. (2002), *Hierarchical Linear Models*, Thousand Oaks: Sage Publications.

Skrondal, A., and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling*, Boca Raton: Chapman & Hall.

Snijders, T. A. B. (2001), "The Statistical Evaluation of Social Network Dynamics," in *Sociological*

*Methodology*, eds. Sobel, M., and Becker, M., Boston and London: Basil Blackwell, pp. 361–395.

Snijders, T. A. B., Koskinen, J., and Schweinberger, M. (2007), "Maximum Likelihood Estimation for Social Network Dynamics," unpublished manuscript.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728.

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

Yin, G. (1991), "On Extensions of Polyak's Averaging Approach to Stochastic Approximation," *Stochastics and Stochastics Reports*, 36, 245–264.

Zijlstra, B. J. H., Van Duijn, M. A. J., and Snijders, T. A. B. (2006), "The Multilevel p2 Model: A Random Effect Model for the Analysis of Multiple Social Networks," *Methodology*, 2, 42–47.