Statistical Inference for Continuous-Time Markov Processes With Block Structure Based On Discrete-Time Network Data

Michael Schweinberger
Department of Statistics, Rice University
6100 Main St, Houston, TX 77005, U.S.A.
E-mail: m.s@rice.edu

Abstract

A widely used approach to modeling discrete-time network data assumes that discrete-time network data were generated by an unobserved continuous-time Markov process. While such models can capture a wide range of network phenomena and are popular in social network analysis, the models are based on the homogeneity assumption that all nodes share the same parameters. We remove the homogeneity assumption by allowing nodes to belong to unobserved subsets of nodes, called blocks, and assuming that nodes in the same block have the same parameters while nodes in distinct blocks have distinct parameters. The resulting models capture unobserved heterogeneity across nodes and admit model-based clustering of nodes based on network properties chosen by researchers. We develop Bayesian data-augmentation methods and apply them to discrete-time observations of an ownership network of non-financial companies in Slovenia in its critical transition from a socialist economy to a market economy. We detect a small subset of shadow-financial companies that outpaces others in terms of the rate of change and the desire to accumulate stock of other companies.

Keywords: finite mixture models, model-based clustering, random graphs, social networks

1 Introduction

Network data help understand a connected world by shedding light on how connections are created and change over time, and how connections affect outcomes of interest, such as public health or national security. As a consequence, the statistical analysis of network data has garnered considerable attention (Kolaczyk 2009).

We focus here on longitudinal network data, consisting of observations of a population network at discrete time points. A widely used approach to modeling discrete-time network data assumes that discrete-time network data were generated by an unobserved continuous-time Markov process. Continuous-time Markov processes of network data were pioneered by Holland and Leinhardt (1977a,b) and Wasserman (1980), but did not become popular until Snijders (2001, 2017) proposed actor-driven parameterizations of continuous-time Markov processes and elaborated statistical methods for estimating them (followed by Koskinen and Snijders 2007; Schweinberger and Snijders 2007; Snijders et al. 2010, and others). Some more recent developments can be found in, e.g., Snijders et al. (2007), Steglich et al. (2010), Niezink and Snijders (2017), Block et al. (2018), Stadtfeld et al. (2018), and Krause et al. (2018). These models are known as stochastic actor-oriented models in the social networks literature (Snijders 2017), and are widely used to study how connections are created and change over time, how connections affect the behavior of actors (social influence), and how the behavior of actors affects connections (social selection) (see, e.g., Snijders et al. 2007; Steglich et al. 2010). But, while popular in social network analysis, these models are based on the homogeneity assumption that all nodes share the same parameters, which may be violated in practice.

We remove the homogeneity assumption by allowing nodes to belong to unobserved subsets of nodes, called blocks, and assuming that nodes in the same block have the same parameters while nodes in distinct blocks have distinct parameters. The resulting models can capture unobserved heterogeneity across nodes and admit model-based clustering of nodes based on network properties chosen by researchers. To infer the parameters of the unobserved continuous-time Markov process along with the block structure from discrete-time network data, we develop Bayesian data-augmentation methods. The issue of non-identifiable parameters, arising from the invariance of the likelihood function to permutations of the labels of blocks, is solved in a Bayesian decision-theoretic framework. We demonstrate the usefulness of these models by applying them to discrete-time observations of an ownership network of non-financial companies in Slovenia in its critical transition from a socialist economy to a market economy (Pahor 2003; Pahor, Prasnikar, and Ferligoj 2004). We are able to detect a small subset of companies that outpaces a large subset of companies in terms of the rate of change as well as the desire to accumulate stock of other companies. These results lend support to the conjecture of Pahor (2003) that the ownership network consists of a large subset of non-financial companies and a small subset of shadow-financial companies, i.e., companies that are not known as financial companies but behave as financial companies.

The remainder of the paper is structured as follows. Section 2 introduces continuous-time Markov processes with block structure. Section 3 proposes Bayesian data-augmentation methods to estimate the parameters of continuous-time Markov processes with block structure from discrete-time network data. We demonstrate the usefulness of these models by an application to an ownership network in Section 4.

Relation to stochastic block models The assumption underlying the proposed continuous-time Markov processes with block structure, that nodes in the same block have the same parameters, is reminiscent of the assumption of stochastic block models (Nowicki and Snijders 2001), that nodes in the same block have the same parameters. Stochastic block models (Fienberg and Wasserman 1981; Holland et al. 1983; Wasserman and Anderson 1987; Nowicki and Snijders 2001) build on the notion of structural equivalence introduced by Lorrain and White (1971). According to Lorrain and White (1971), blocks are subsets of nodes that are connected to the same nodes in the network and hence have equivalent positions in the network. Stochastic models with block structure extend the deterministic notion of structural equivalence to a stochastic notion of structural equivalence. According to Wasserman and Anderson (1987), blocks are subsets of nodes that have the same connection probabilities, although nodes belonging to the same block may not have the same connections to other nodes in the network. Stochastic block models may be the simplest stochastic models with block structure, but there are many other stochastic models with block structure. Most of them are based on relaxations of the notion of structural equivalence. For example, degree-corrected stochastic block models (Zhao et al. 2012) assume that connection probabilities depend on blocks, but capture unobserved heterogeneity in the propensities of nodes to form connections; and mixed membership block models (Airoldi et al. 2008) assume that the block memberships of nodes depend on who interacts with whom. The proposed models can likewise be viewed as stochastic models of structural equivalence: While stochastic block models assume that the edges of all nodes in the same block are governed by the same parameters, the proposed models assume that the changes of edges of all nodes in the same block are governed by the same parameters. That said, there are notable differences: the proposed models are models of longitudinal network data rather than cross-sectional network data, and changes of edges may be affected by transitivity and other structural network features (Wasserman and Faust 1994).

Other, related models Other, related models are temporal stochastic block and latent space models (e.g., Fu et al. 2009; Sewell and Chen 2015, 2016; Sewell et al. 2016) and temporal exponential-family random graph models (Robins and Pattison 2001; Hanneke et al. 2010; Ouzienko et al. 2011; Krivitsky and Handcock 2014), among others (e.g., Katz and Proctor 1959; Durante and Dunson 2014; Sewell 2017). However, the first class of models does not allow to model a wide range of network phenomena (although some of them do capture a stochastic tendency towards transitivity), while the second class of models cannot capture unobserved heterogeneity (although it can capture observed heterogeneity through covariates). An additional class of related models are relational event models (Butts 2008), but relational event models focus on edges without duration (e.g., emails), whereas we focus on edges with duration (e.g., friendships, ownerships of stock).

2 Model

We consider discrete-time network data, in the form of a population of nodes $\mathcal{N} = \{1, \ldots, n\}$ with a population graph observed at two or more discrete time points in some time interval $\mathcal{T} = [t_0, t_1] \subset \mathbb{R}$, where $t_0 < t_1$.

To capture unobserved heterogeneity in discrete-time network data, we assume that the population $\mathcal{N}$ is partitioned into $K \geq 2$ subpopulations $1, \ldots, K$, called blocks. Denote by $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ vectors of block memberships, where element $Z_{i,k}$ of vector $\mathbf{Z}_i$ is 1 if node $i \in \mathcal{N}$ is member of block k and 0 otherwise. We assume that the block membership vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are generated by

$$\mathbf{Z}_i \mid \alpha_1, \ldots, \alpha_K \overset{\text{iid}}{\sim} \text{Multinomial}(1; \alpha_1, \ldots, \alpha_K), \ i \in \mathcal{N},$$

where $\alpha_1, \ldots, \alpha_K$ are the parameters of the multinomial distribution satisfying $0 < \alpha_k < 1$ $(k = 1, \ldots, K)$ and $\sum_{k=1}^{K} \alpha_k = 1$. We write henceforth $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$.

Conditional on the partition of the population $\mathcal{N}$ into $K \geq 2$ subpopulations, the population graph $\mathbf{Y}(t) = (Y_{i,j}(t))_{(i,j) \in \mathcal{N} \times \mathcal{N}, t \in \mathcal{T}}$ in time interval $\mathcal{T} = [t_0, t_1] \subset \mathbb{R}$ is governed by a continuous-time Markov process. Here, $Y_{i,j}(t) = 1$ indicates that there is a directed edge from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$ at time $t \in \mathcal{T}$ and $Y_{i,j}(t) = 0$ otherwise. By convention, self-relationships are discarded by constraining $Y_{i,i}(t) = 0$ for all nodes $i \in \mathcal{N}$.

In the following, we develop the proposed continuous-time Markov modeling framework from first principles and clarify the underlying assumptions and limitations of the framework. Suppose that the Markov process $\mathbf{Y}(t)$ is at graph $\mathbf{Y} \in \mathcal{Y}$ at time $t \in \mathcal{T}$. Then the transition probability of moving from graph $\mathbf{Y}$ to graph $\mathbf{Y}^\star \neq \mathbf{Y} \in \mathcal{Y}$ in a time interval $(t, t+h)$ of length $h > 0$ is assumed to be of the form

$$\mathbb{P}[\mathbf{Y}(t+h) = \mathbf{Y}^\star \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}] = \prod_{i,j=1}^{n} \mathbb{P}[Y_{i,j}(t+h) = Y_{i,j}^\star \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}] + o(h). \qquad (1)$$

Here,

$$\mathbb{P}[Y_{i,j}(t+h) = Y_{i,j}^\star \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}] = q_{i,j}(\mathbf{Y}, \mathbf{Z})\, h + o(h)$$

denotes the transition probability of going from graph $\mathbf{Y}(t) = \mathbf{Y}$ to graph $\mathbf{Y}(t+h) = \mathbf{Y}^\star \neq \mathbf{Y}$ in time interval $(t, t+h)$ by changing $Y_{i,j}$ to $Y_{i,j}^\star = 1 - Y_{i,j}$ while leaving all other edges unchanged,

$$q_{i,j}(\mathbf{Y}, \mathbf{Z}) = \lim_{h \longrightarrow 0} \frac{\mathbb{P}[Y_{i,j}(t+h) = Y_{i,j}^\star \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}]}{h}$$

denotes the rate of change of $Y_{i,j}$ given $\mathbf{Y}(t) = \mathbf{Y}$, and $o(h)$ denotes a term that is of a smaller order of magnitude than the length $h > 0$ of time interval $(t, t+h)$.

Equation (1) shows that these continuous-time Markov processes make two important, related assumptions:

1. Changes of edges in short time intervals $(t, t+h)$ are independent conditional on $\mathbf{Y}(t) = \mathbf{Y}$ and $\mathbf{Z}$.

2. Changes of the population graph are local in the sense that the probability that more than one edge in time interval $(t, t+h)$ changes is $o(h)$ (Holland and Leinhardt 1977a; Wasserman 1977, 1980; Snijders 2001).

While these assumptions restrict the range of possible model specifications, continuous-time Markov processes with these assumptions have turned out to be useful in practice, because changes of edges can depend on other edges in the population graph at time t, allowing to model transitivity and many other interesting forms of network dependence (Wasserman and Faust 1994).

The Markov process $\mathbf{Y}(t)$ is fully specified by specifying the rates of change $q_{i,j}(\mathbf{Y}, \mathbf{Z})$. We consider an attractive specification along the lines of Snijders (2001), given by

$$q_{i,j}(\mathbf{Y}, \mathbf{Z}) = \lambda_i(\mathbf{Y}, \mathbf{Z})\; p_i(j \mid \mathbf{Y}, \mathbf{Z}), \tag{2}$$

where $\lambda_i(\mathbf{Y}, \mathbf{Z})$ satisfies $\lambda_i(\mathbf{Y}, \mathbf{Z}) > 0$ for all $i \in \mathcal{N}$ and $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ satisfies $0 < p_i(j \mid \mathbf{Y}, \mathbf{Z}) < 1$ for all $(i, j) \in \mathcal{N} \times \mathcal{N}$ and $\sum_{j \neq i}^{n} p_i(j \mid \mathbf{Y}, \mathbf{Z}) = 1$ for all $i \in \mathcal{N}$. Here, $\lambda_i(\mathbf{Y}, \mathbf{Z})$ can be interpreted as the rate of change of actor i, whereas $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ can be interpreted as the conditional probability that actor i chooses to update her relationship to actor j, given that actor i changes one of her relationships.

The rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z})$ and conditional probabilities $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ can depend on the population graph $\mathbf{Y}$ and the block structure $\mathbf{Z}$ as follows:

$$\begin{aligned}
\lambda_i(\mathbf{Y}, \mathbf{Z}) \quad &\equiv \quad \lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) \quad &= \quad \exp[\boldsymbol{\eta}_{i,1}^T(\mathbf{Z}, \boldsymbol{\theta}_1)\, s_{i1}(\mathbf{Y})] \\
p_i(j \mid \mathbf{Y}, \mathbf{Z}) \quad &\equiv \quad p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2) \quad &= \quad \exp[\boldsymbol{\eta}_{i,2}^T(\mathbf{Z}, \boldsymbol{\theta}_2)\, s_{i2}(j, \mathbf{Y}) - \psi_i(\mathbf{Z}, \boldsymbol{\theta}_2)], \quad j \in \mathcal{N}_i,
\end{aligned}$$

where $\mathcal{N}_i = \mathcal{N} \setminus \{i\}$ and

$$\psi_i(\mathbf{Z}, \boldsymbol{\theta}_2) = \log \sum_{k \in \mathcal{N}_i} \exp[\boldsymbol{\eta}_{i,2}^T(\mathbf{Z}, \boldsymbol{\theta}_2)\, s_{i2}(k, \mathbf{Y})].$$

Here, $\boldsymbol{\eta}_{i,1}(\mathbf{Z}, \boldsymbol{\theta}_1)$ and $\boldsymbol{\eta}_{i,2}(\mathbf{Z}, \boldsymbol{\theta}_2)$ are vectors of parameters and $s_{i1}(\mathbf{Y})$ and $s_{i2}(j, \mathbf{Y})$ are vectors of statistics. The rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ and conditional probabilities $p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2)$ of nodes $i \in \mathcal{N}$ depend the block memberships of nodes $i \in \mathcal{N}$ via the parameters

$$\begin{aligned}
\boldsymbol{\eta}_{i,1}(\mathbf{Z}, \boldsymbol{\theta}_1) &= \boldsymbol{\theta}_1^T \mathbf{Z}_i,\ i = 1, \ldots, n \\
\boldsymbol{\eta}_{i,2}(\mathbf{Z}, \boldsymbol{\theta}_2) &= \boldsymbol{\theta}_2^T \mathbf{Z}_i,\ i = 1, \ldots, n,
\end{aligned}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are matrices of parameters. The element (j, k) of the matrix of parameters $\boldsymbol{\theta}_1$ can be interpreted as the strength of effect j on the rate of change of nodes in block k,

whereas the element (j, k) of the matrix of parameters $\boldsymbol{\theta}_2$ can be interpreted as the strength of effect j on changes of edges initiated by nodes in block k.

Remark 1. Model-based clustering based on network properties chosen by researchers. Models can be specified by choosing statistics $s_{i1}(\mathbf{Y})$ and $s_{i2}(j, \mathbf{Y})$, i.e., by choosing functions of the network of interest. The resulting models can capture unobserved heterogeneity across nodes and admit model-based clustering of nodes based on network properties chosen by researchers. We give examples of specifications of $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ and $p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2)$ in Section 4, where we cluster nodes based on the rate of change and the number of edges.

3 Bayesian inference

To infer the parameters of the unobserved continuous-time Markov process along with the unobserved blocks from discrete-time network data, we develop Bayesian data-augmentation methods.

We first state the likelihood function and priors in Sections 3.1 and 3.2, respectively, and then develop Bayesian Markov chain Monte Carlo data-augmentation methods in Section 3.3. Solutions of the label-switching problem of Bayesian Markov chain Monte Carlo algorithms, which is rooted in the invariance of the likelihood function to the labeling of the blocks, are discussed in Section 3.4. Throughout, we focus on a time interval $[t_0, t_1]$ and assume that the continuous-time Markov process is observed at t_0 and t_1 because, by the Markov property, the extension to multiple, non-overlapping time intervals is straightforward. In addition, we condition on the population graph $\mathbf{Y}(t_0)$ at time t_0, which has the advantage that we do not need to make assumptions about the process that generated $\mathbf{Y}(t_0)$.

3.1 Likelihood function

We start with the likelihood function of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ based on an observation of the continuous-time Markov process $\mathbf{Y}(t)$ and block structure $\mathbf{Z}$. An observation of the continuous-time Markov process $\mathbf{Y}(t)$ corresponds to the number of changes M in time interval $[t_0, t_1]$ and the sequence $W_M = (h_m, i_m, j_m)_{m=1}^M$ of holding times h_m and ordered pairs of nodes (i_m, j_m) that make changes at times $t_0 + \sum_{k=1}^m h_k$ $(m = 1, \ldots, M)$.

The likelihood function of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ given W_M and $\mathbf{Z}$ factorizes as follows:

$$L(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, \mathbf{Z}) \;\propto\; L(\boldsymbol{\alpha}; \mathbf{Z}) \times L(\boldsymbol{\theta}_1; W_M, \mathbf{Z}) \times L(\boldsymbol{\theta}_2; W_M, \mathbf{Z}). \tag{3}$$

The likelihood function of $\boldsymbol{\alpha}$ given $\mathbf{Z}$ is proportional to

$$L(\boldsymbol{\alpha}; \mathbf{Z}) \;\propto\; \prod_{i=1}^{n} \prod_{k=1}^{K} \alpha_k^{Z_{i,k}}.$$

According to the theory of continuous-time Markov processes (Karlin and Taylor 1975) along with parameterization (2), the likelihood function of $\boldsymbol{\theta}_1$ given W_M and $\mathbf{Z}$ is proportional to

$$\begin{aligned} L(\boldsymbol{\theta}_1; W_M, \mathbf{Z}) &\propto \left\{ \prod_{m=1}^{M} \lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1) \exp\left[-\lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\, h_m\right] \frac{\lambda_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)}{\lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)} \right\} \\ &\times \exp\left[-\lambda(\mathbf{Y}_M, \mathbf{Z}, \boldsymbol{\theta}_1) \left(t_1 - t_0 - \sum_{m=1}^{M} h_m \right) \right] \end{aligned}$$

and the likelihood function of $\boldsymbol{\theta}_2$ given W_M and $\mathbf{Z}$ is proportional to

$$L(\boldsymbol{\theta}_2; W_M, \mathbf{Z}) \propto \prod_{m=1}^{M} p_{i_m}(j_m \mid \mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_2),$$

where

$$\lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1) = \sum_{k=1}^{n} \lambda_k(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1).$$

3.2 Priors

We consider non-parametric stick-breaking priors (Ferguson 1973; Ishwaran and James 2001; Teh 2010), which help sidestep the selection of the number of blocks K. The advantage of using stick-breaking priors is that one does not have to specify the number of non-empty blocks, because the number of non-empty blocks is random (Teh 2010).

A stick-breaking construction of $\boldsymbol{\alpha}$ is given by

$$\begin{aligned} \alpha_1 &= V_1 \\ \alpha_k &= V_k \prod_{j=1}^{k-1} (1 - V_j),\ k = 2, 3, \ldots, \end{aligned}$$

where

$$V_k \mid A_k, B_k \overset{\text{ind}}{\sim} \text{Beta}(A_k, B_k),\ k = 1, 2, \ldots.$$

The process can be thought of as starting with a stick of length 1, partition the stick into two pieces of length proportional to V_k and $1 - V_k$, assigning the length of the first segment to α_k and continuing to partition the second segment, $k = 1, 2, \ldots$ Stick-breaking priors can be approximated by truncated stick-breaking priors (Ishwaran and James 2001): by choosing a large number K considered to be an upper bound to the number of blocks needed to obtain good goodness-of-fit, and truncating the stick-breaking prior by setting $V_K = 1$ (which corresponds to assigning the entire length of the remaining stick to α_K), so that $\sum_{k=1}^{K} \alpha_k = 1$. We use truncated stick-breaking priors, which implies that $\boldsymbol{\alpha}$ is generalized Dirichlet distributed (Connor and Mosiman 1969; Ishwaran and James 2001), and note that the Dirichlet prior is a special case of the generalized Dirichlet prior (Connor and Mosiman 1969).

If the rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \theta_1) = \theta_1 > 0$ are constant, then it is convenient to use the conjugate prior given by

$$\theta_1 \,|C, D \sim \text{Gamma}(C, D).$$

Otherwise, the prior of the unique elements of $\boldsymbol{\theta}_1$, stored in the vector $\text{v}(\boldsymbol{\theta}_1)$, is assumed to be Gaussian, where

$$\text{v}(\boldsymbol{\theta}_1) \sim N(0, \text{diag}(\boldsymbol{\Sigma}_1)),$$

where $\text{diag}(\boldsymbol{\Sigma}_1)$ is a diagonal variance-covariance matrix.

The prior of the unique elements of $\boldsymbol{\theta}_2$, stored in the vector $\text{v}(\boldsymbol{\theta}_2)$, is assumed to be

$$\text{v}(\boldsymbol{\theta}_2) \sim N(0, \text{diag}(\boldsymbol{\Sigma}_2)),$$

where $\text{diag}(\boldsymbol{\Sigma}_2)$ is a diagonal variance-covariance matrix.

3.3 Bayesian data-augmentation methods

We approximate the posterior by using Bayesian Markov chain Monte Carlo data-augmentation methods.

To reduce the Markov chain Monte Carlo error, we integrate out the holding times $(h_1, \ldots, h_M)$, as suggested by Snijders et al. (2010). Note that, without eliminating the holding times, we would need Markov chain Monte Carlo algorithms with dimension-changing moves (e.g., reversible-jump Metropolis-Hastings algorithms), because the dimension M of the vector of holding times $(h_1, \ldots, h_M)$ is unknown.

To eliminate the holding times, note that in the special case where the rates of change $\lambda_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \theta_1) = \theta_1 > 0$ are constant, the likelihood function of θ_1 given W_M and $\mathbf{Z}$ is proportional to

$$L(\theta_1; W_M, \mathbf{Z}) \;\; \propto \;\; \exp\left[-n\,(t_1 - t_0)\,\theta_1\right] \; \left[n\,(t_1 - t_0)\,\theta_1\right]^M, \tag{4}$$

which implies that we do not need the holding times $(h_1, \ldots, h_M)$ in order to evaluate $L(\theta_1; W_M, \mathbf{Z})$.

In general, when the rates of changes $\lambda_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)$ are not constant, the likelihood function of $\boldsymbol{\theta}_1$ given W_M and $\mathbf{Z}$ can be approximated by

$$L(\boldsymbol{\theta}_1; W_M, \mathbf{Z}) \;\; \approx \;\; \frac{p_{\mu_T, \sigma_T^2}(t_1)}{\lambda(\mathbf{Y}_M, \mathbf{Z}, \boldsymbol{\theta}_1)}, \tag{5}$$

where $p_{\mu_T, \sigma_T^2}(.)$ denotes the probability density function of $N(\mu_T, \sigma_T^2)$, with mean $\mu_T = \sum_{m=1}^{M} 1 \,/\, \lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)$ and variance $\sigma_T^2 = \sum_{m=1}^{M} 1 \,/\, \lambda^2(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)$. The approximation (5) shares with (4) the advantage that the holding times $(h_1, \ldots, h_M)$ are not needed to approximate $L(\boldsymbol{\theta}_1; W_M, \mathbf{Z})$. The approximation (5) takes advantage of the fact that the holding times $h_1, \ldots, h_M$ are independent Exponential random variables with parameters

$\lambda(\mathbf{Y}_0, \mathbf{Z}, \boldsymbol{\theta}_1), \ldots, \lambda(\mathbf{Y}_{M-1}, \mathbf{Z}, \boldsymbol{\theta}_1)$, respectively. Thus, the Lindeberg-Feller Central Limit Theorem for independent (but not identically distributed) random variables implies that the distribution of $\sum_{m=1}^{M} h_m$ can be approximated by $N(\mu_T, \sigma_T^2)$, provided the number of changes M in time interval $[t_0, t_1]$ is large. Mathematical details can be found in Snijders et al. (2010). A small simulation study in Snijders et al. (2010) suggests that the approximation of the likelihood function works well in scenarios with 20–32 nodes, 2 time intervals, and 50–112 expected changes in each time interval, implying that the total number of expected changes is 100–224. In the application in Section 4, the number of nodes is 165, the number of time intervals is 4, and the observed numbers of changes in the 4 time intervals are 52, 60, 35, and 90. So the total number of observed changes (237) exceeds the total number of expected changes in the simulation study (100–224) and hence the approximation of the likelihood function can be expected to work well.

We describe Markov chain Monte Carlo methods for sampling from the posterior in Appendix A, taking advantage of (4) and (5).

Remark 2. It is worth noting that Koskinen and Snijders (2007) first developed Bayesian inference for continuous-time Markov models of discrete-time network data, albeit without unobserved block structure. The Bayesian procedure described above differs from the Bayesian procedure of Koskinen and Snijders (2007) as follows:

1. We infer unobserved block structure, whereas Koskinen and Snijders (2007) do not consider unobserved block structure.

2. Koskinen and Snijders (2007) infer the unobserved holding times of the continuous-time Markov process, whereas we do not infer them but integrate them out. To infer the unobserved holding times and address the issue that the dimension M of the vector of unobserved holding times $h_1, \ldots, h_M$ depends on the unobserved number of changes M, Koskinen and Snijders (2007) use a reversible jump Metropolis-Hastings algorithm (Green 1995). The idea of integrating out unobserved holding times, rather than inferring them, is due to Snijders et al. (2010), and is motivated by the desire to simplify the Markov chain Monte Carlo algorithm and reduce its simulation variance (see pages 574 and 586 of Snijders et al. 2010).

3.4 Label-switching problem

The likelihood function (3) stated in Section 3.1 is invariant to the labeling of blocks, which implies that Bayesian Markov chain Monte Carlo algorithms may exhibit label-switching (Stephens 2000). While the stick-breaking prior described in Section 3.2 is not invariant to the labeling of blocks, we have observed that Bayesian Markov chain Monte Carlo algorithms nonetheless experience label-switching, because the likelihood function dominates the prior when there are enough data.

To solve the label-switching problem of Markov chain Monte Carlo algorithms, we follow the Bayesian decision-theoretic approach of Stephens (2000). In other words, we choose a loss function and minimize the posterior expected loss. To introduce the basic idea in its simplest form, consider the following toy example with $n = 4$ nodes and $K = 2$ blocks labeled 1 and 2, and the following sample of size $N = 4$ from the posterior:

```
1 1 2 2
1 1 2 2
2 2 1 1
2 2 1 1
```

Here, the first row shows the first sample of block memberships of nodes 1, 2, 3, 4, the second row shows the second sample, etc. The sample of size $N = 4$ reveals at least three interesting facts:

- Nodes 1 and 2 are assigned to the same block in all samples.
- Nodes 3 and 4 are assigned to the same block in all samples.
- The block of nodes 3 and 4 is different from the block of nodes 1 and 2.

However, naive summaries of the posterior are problematic, because the labels of the two blocks switched between the first two samples and the last two samples. For example, if we wanted to report estimates of the posterior probabilities that nodes 1 and 2 belong to blocks 1 and 2 and reported the proportions of samples that assign them to blocks 1 and 2 as estimates (which are 1/2 and 1/2, respectively), then the estimates would conceal the fact that nodes 1 and 2 are assigned to the same block in all samples.

To undo the label-switching and obtain estimates of the posterior classification probabilities along the way, consider the following thought experiment.

First, suppose that we want to report estimates of the posterior classification probabilities, and assume that the true block memberships $\mathbf{Z}^\star$ of nodes are known to be:

- Nodes 1 and 2 belong to block 1, so $Z^\star_{1,1} = Z^\star_{2,1} = 1$ and $Z^\star_{1,2} = Z^\star_{2,2} = 0$.
- Nodes 3 and 4 belong to block 2, so $Z^\star_{3,1} = Z^\star_{4,1} = 0$ and $Z^\star_{3,2} = Z^\star_{4,2} = 1$.

Let $\mathbf{Q} = (q_{i,k})$ be the matrix of posterior classification probabilities, where $q_{i,k}$ is the posterior probability that node i belongs to block k. To estimate $\mathbf{Q}$, consider the objective function

$$g(\mathbf{Q};\, \mathbf{Z}^\star) \;=\; \prod_{i=1}^{n} q_{i,\sum_{k=1}^{K} k\, Z^\star_{i,k}}.$$

It is not too hard to see that the maximizer $\mathbf{Q}^\star = (q^\star_{i,k})$ of $g(\mathbf{Q};\, \mathbf{Z}^\star)$,

$$\mathbf{Q}^\star \;=\; \arg\max_{\mathbf{Q}} g(\mathbf{Q};\, \mathbf{Z}^\star),$$

is given by $q^\star_{i,1} = 1$ and $q^\star_{i,2} = 0$ $(i = 1, 2)$ and $q^\star_{i,1} = 0$ and $q^\star_{i,2} = 1$ $(i = 3, 4)$, where the maximization is over all matrices $\mathbf{Q}$ such that $q_{i,k} \geq 0$ $(k = 1, \ldots, K)$ and $\sum_{k=1}^K q_{i,k} = 1$ $(i = 1, \ldots, n)$. In other words, $\mathbf{Q}^\star$ suggests that, with high posterior probability, nodes 1 and 2 belong to block 1 and nodes 3 and 4 belong to block 2, which are indeed the true block memberships. We could thus report $\mathbf{Q}^\star$ as an educated guess of the posterior classification probabilities, provided $\mathbf{Z}^\star$ is known.

In practice, $\mathbf{Z}^\star$ is unknown, but suppose that $\mathbf{Q}^\star$ is known. Then we could relabel the sample of block memberships $\mathbf{Z}_l$ $(l = 1, \ldots, N)$ by choosing permutations ν_l that maximize $g(\mathbf{Q}^\star;\, \nu_l(\mathbf{Z}_l))$ $(l = 1, \ldots, N)$:

$$\nu^\star_l \;\; = \;\; \arg\max_{\nu_l} g(\mathbf{Q}^\star;\, \nu_l(\mathbf{Z}_l)), \quad l = 1, \ldots, N.$$

The maximizers $\nu^\star_1, \ldots, \nu^\star_N$ are not unique, but the lack of uniqueness is not a concern: Any sequence of permutations that undoes the label-switching is useful. In the toy example, it is not hard to see that the permutations $\nu^\star_l(1) = 1$ and $\nu^\star_l(2) = 2$ $(l = 1, 2)$ and $\nu^\star_l(1) = 2$ and $\nu^\star_l(2) = 1$ $(l = 3, 4)$ are maximizers of $g(\mathbf{Q}^\star;\, \nu_l(\mathbf{Z}_l))$ $(l = 1, \ldots, N)$. Using permutations $\nu^\star_1, \ldots, \nu^\star_N$, we can permute the sample of block memberships $\mathbf{Z}_l$ $(l = 1, \ldots, N)$ as follows:

```
1 1 2 2
1 1 2 2
1 1 2 2
1 1 2 2
```

In other words, we have undone the label-switching.

It goes without saying that in practice neither $\mathbf{Q}^\star$ nor $\nu^\star_1, \ldots, \nu^\star_N$ are known, but it is natural to devise an iterative optimization algorithm for undoing the label-switching and obtaining estimates of posterior classification probabilities as follows. First, notice that maximizing the objective function $g(\mathbf{Q};\, \nu_l(\mathbf{Z}_l))$ is equivalent to minimizing the loss function

$$f(\mathbf{Q};\, \nu_l(\mathbf{Z}_l)) \;\; = \;\; -\log g(\mathbf{Q};\, \nu_l(\mathbf{Z}_l)) \;\; = \;\; -\sum_{i=1}^n \log q_{i,\sum_{k=1}^K k\, Z_{i,k}}, \quad l = 1, \ldots, N.$$

Suppose that initial permutations $\nu^{(0)}_1, \ldots, \nu^{(0)}_N$ are available, e.g., $\nu^{(0)}_l(k) = k$ $(k = 1, \ldots, K,$ $l = 1, \ldots, N)$. A natural minimization algorithm iterates the following two steps until a local minimum of the loss function has been found:

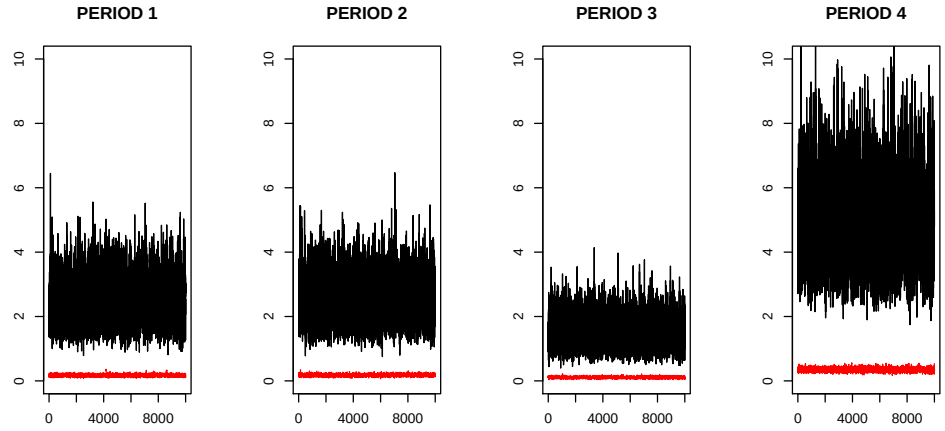
At iteration $m = 1, 2, \ldots,$ compute:

1. Given $\nu^{(m-1)}_1, \ldots, \nu^{(m-1)}_N$, compute

$$\mathbf{Q}^{(m)} \;\; = \;\; \arg\min_{\mathbf{Q}} \sum_{l=1}^N f(\mathbf{Q};\, \nu^{(m-1)}_l(\mathbf{Z}_l)),$$

where the minimization is over all matrices $\mathbf{Q}$ such that $q_{i,k} \geq 0$ $(k = 1, \ldots, K)$ and $\sum_{k=1}^K q_{i,k} = 1$ $(i = 1, \ldots, n)$.

Figure 1: Trace plots of the rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ of companies i in blocks 1 and 2, as defined in (6). The black-colored lines refer to the rates of change of companies in block 1 in periods 1, 2, 3, and 4, whereas the red-colored lines refer to the rate of change of companies in block 2 in periods 1, 2, 3, and 4. These trace plots do not show signs of non-convergence. Summaries of the posterior of the rates of change are shown in Table 1.



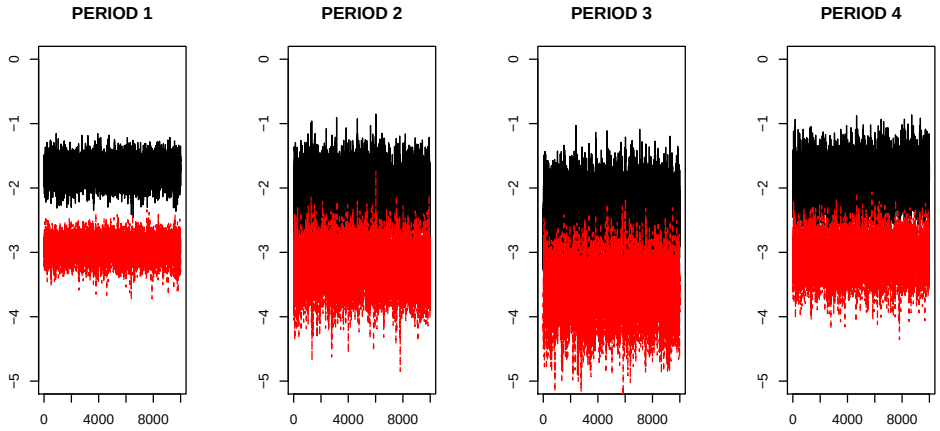
2. Given $\mathbf{Q}^{(m)}$, compute

$$\nu_l^{(m)} \quad = \quad \arg\min_{\nu_l} f(\mathbf{Q}^{(m)};\, \nu_l(\mathbf{Z}_l)), \quad l = 1, \ldots, N.$$

Upon convergence, the Markov chain Monte Carlo sample of block memberships and parameters can be relabeled by using the optimal permutations obtained at the last iteration, and the optimal classification probabilities obtained at the last iteration can be reported as estimates of the posterior classification probabilities.

Remark 3. Implementation. The minimization algorithm described above converges to a local minimum of the loss function. It is therefore advisable to run the minimization algorithm multiple times, with starting values chosen at random. In addition, it is worth noting that Step 2 involves minimization over all $K!$ possible permutations of the block labels $1, \ldots, K$. Unless K is small, Step 2 is time-consuming. A time-saving alternative is Simulated Annealing (Schweinberger and Handcock 2015, Supplement C). Both exact versions of Step 2 (based on minimizing over all $K!$ permutations) and approximate versions of Step 2 (based on Simulated Annealing) are implemented in `R` package `hergm` (Schweinberger and Luna 2018). The sample in the toy example can be relabeled by using the `R` script in Appendix B.

Figure 2: Trace plots of the outdegree parameters of companies in blocks 1 and 2. The black-colored lines refer to the outdegree parameters of companies in block 1 in periods 1, 2, 3, and 4, whereas the red-colored lines refer to the outdegree parameters of companies in block 2 in periods 1, 2, 3, and 4. These trace plots do not show signs of non-convergence. Summaries of the posterior of the outdegree parameters are shown in Table 1.



4 Application

We demonstrate the usefulness of the model-based clustering framework by applying it to an ownership network of non-financial companies, of which some companies are suspected to be shadow-financial companies.

Pahor (2003) studied ownership of stock holdings among non-financial companies in Slovenia observed at 5 time points between 2000 and 2002, where $Y_{i,j}(t) = 1$ means that company i holds stock of company j at time t and $Y_{i,j}(t) = 0$ otherwise. The observations fall into a period in which Slovenia transitioned from a socialist economy to a market economy. Pahor (personal communication) conjectured that the ownership network has unobserved heterogeneity—not captured by the covariates used in Pahor (2003)—in that the network consists of a large subset of non-financial companies and a small subset of shadow-financial companies: companies that used to produce non-financial goods but shifted the focus from the production of non-financial goods to trading stock of other companies. Shadow-financial companies are thought to buy and sell stock more frequently and accumulate more stock through time than non-financial companies. We focus here on the most prosperous region of Slovenia, which is known as Central Slovenia and includes Ljubljana, the capital of Slovenia (see Table 3.4 of Pahor 2003, p. 123). The data set consists of ownerships of stock among $n = 165$ companies in Central Slovenia. The observed number of changes between the 5 observations of the ownership network are given by 52, 60, 35, and 90, respectively, and the observed number of relationships at the 5 time points is given by 148, 168, 174, 175, and 191, respectively. A more detailed description of the ownership network can be found in Pahor

Figure 3: Trace plots of reciprocity and transitivity parameter.

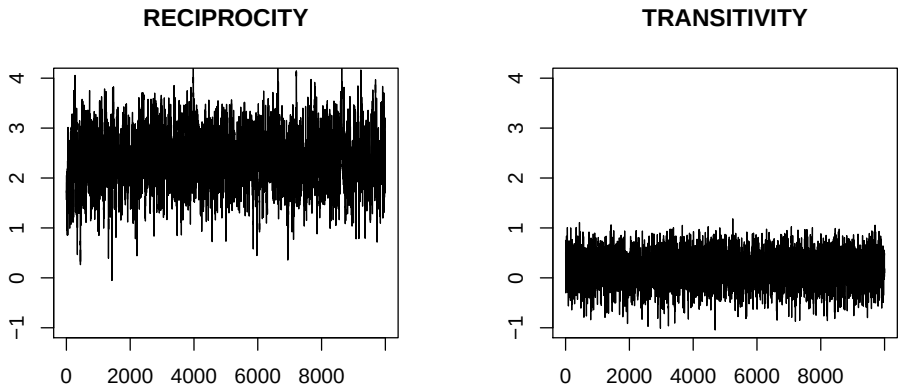
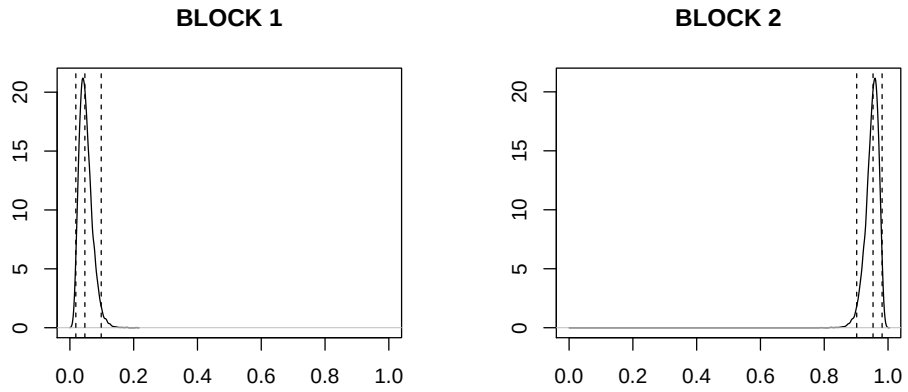


Figure 4: Marginal posterior densities of proportions of blocks 1 and 2; dashed lines indicate 2.5%, 50%, and 97.5% quantiles.



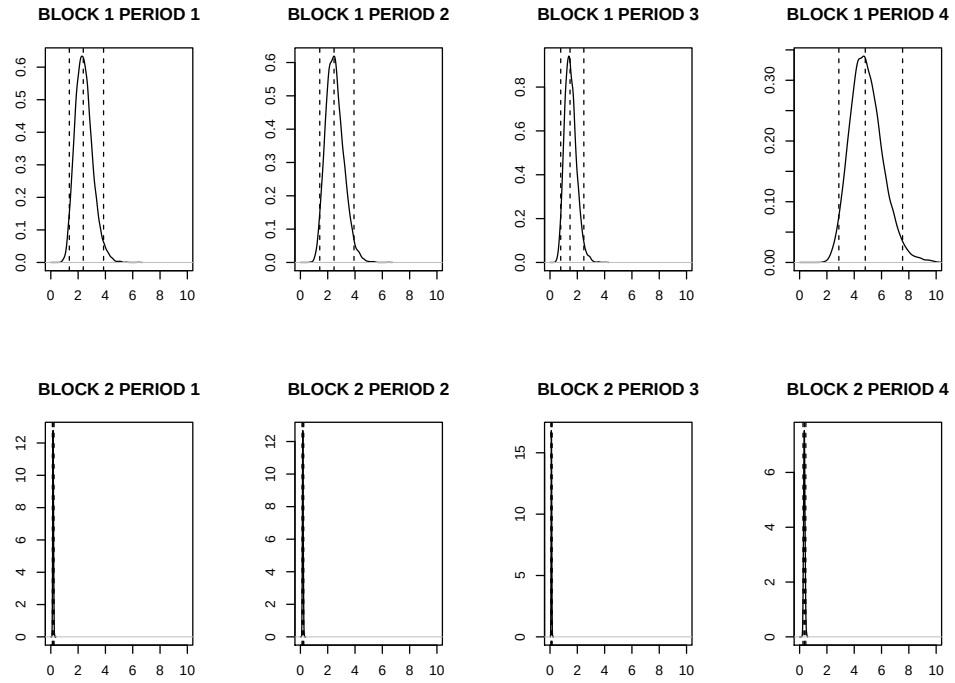
(2003). A related, but distinct data set is described in Pahor et al. (2004).

To detect shadow-financial companies, we consider $K = 2$ blocks, motivated by Pahor's expectation that the ownership network consists of non-financial and shadow-financial companies. We did explore models with 3 blocks, but found much more posterior uncertainty about the block memberships of companies, which may be an indication of model overfit. In other applications where the number of blocks K is unknown, K can be selected by model selection tools. While the development of model selection tools is doubtless an important problem, it is beyond the scope of our paper and is not needed in our application.

Let $Z_{i1} = 1$ and $Z_{i2} = 0$ if i belongs to block 1 and $Z_{i1} = 0$ and $Z_{i2} = 1$ otherwise. The rate of change of company i ($i = 1, \ldots, 165$) in period h ($h = 1, \ldots, 4$) is of the form

$$\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) = \exp[\theta_{1,h} + \theta_{1,5} Z_{i,2}], \tag{6}$$

Figure 5: Marginal posterior densities of the rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ of companies i in blocks 1 and 2, as defined in (6); dashed lines indicate 2.5%, 50%, and 97.5% quantiles.



where $\theta_{1,h}$ is the baseline rate parameter of period h ($h = 1, \ldots, 4$), which is shared by blocks 1 and 2, and $\theta_{1,5}$ represents the deviation of block 2 from the baseline rate parameter. The inclusion of the rate parameters $\theta_{1,h}$ ($h = 1, \ldots, 4$) and $\theta_{1,5}$ allows one subset of companies to buy and sell stock more frequently than the other. The conditional probability that company i changes its relationship to company j, given that it changes its relationship to some company, is assumed to be of the form

$$p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2) = \exp\left[\eta_{i,2,1}\, c_{i,2,1}(j, \mathbf{Y}) + \eta_{i,2,2}\, c_{i,2,2}(j, \mathbf{Y}) + \eta_{i,2,3}\, c_{i,2,3}(j, \mathbf{Y}) - \psi_i(\mathbf{Z}, \boldsymbol{\theta}_2)\right],$$

where the change statistics $c_{i,2,1}(j, \mathbf{Y})$, $c_{i,2,2}(j, \mathbf{Y})$, and $c_{i,2,3}(j, \mathbf{Y})$ correspond to the change in the number of relationships, reciprocated relationships, and transitive relationships due to the change in relationship $y_{i,j}$, and the parameters $\eta_{i,2,1}$, $\eta_{i,2,2}$, and $\eta_{i,2,3}$ are given by

- $\eta_{i,2,1} = \theta_{2,h} + \theta_{2,5} Z_{i,2}$, where $\theta_{2,h}$ is the baseline outdegree parameter of period h ($h = 1, \ldots, 4$) and $\theta_{2,5}$ represents the deviation of block 2 from the baseline outdegree parameter;
- $\eta_{i,2,2} = \theta_{2,6}$ is the reciprocity parameter;
- $\eta_{i,2,3} = \theta_{2,7}$ is the transitivity parameter.

Table 1: 95% posterior confidence intervals of parameters. The rates refer to the rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ of companies i in blocks 1 and 2, as defined in (6).

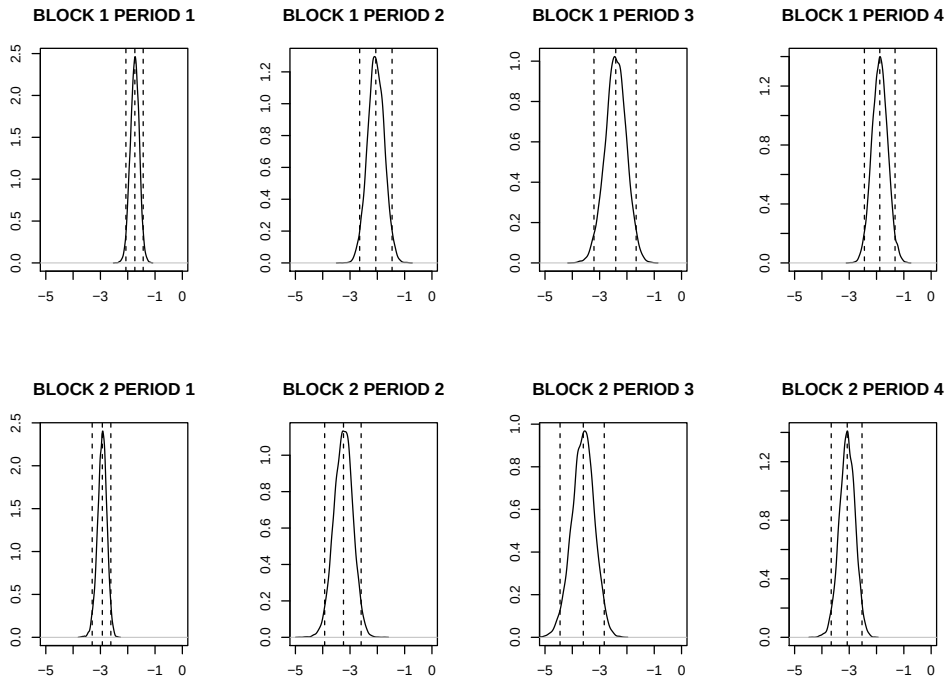
	period 1	period 2	period 3	period 4
rate block 1	$(1.36, 3.87)$	$(1.42, 3.93)$	$(.79, 2.48)$	$(2.87, 7.54)$
rate block 2	$(.11, .24)$	$(.12, .24)$	$(.06, .16)$	$(.25, .46)$
outdegree block 1	$(-2.07, -1.43)$	$(-2.65, -1.46)$	$(-3.21, -1.66)$	$(-2.44, -1.32)$
outdegree block 2	$(-3.30, -2.62)$	$(-3.93, -2.59)$	$(-4.45, -2.83)$	$(-3.66, -2.53)$
reciprocity	$(1.29, 3.41)$	$(1.29, 3.41)$	$(1.29, 3.41)$	$(1.29, 3.41)$
transitivity	$(-.45, .75)$	$(-.45, .75)$	$(-.45, .75)$	$(-.45, .75)$

The inclusion of the outdegree parameters $\theta_{2,h}$ $(h = 1, \ldots, 4)$ and $\theta_{2,5}$ allows one subset of companies to accumulate more stock through time than the other. We choose the Dirichlet$(2, 2)$ prior for the proportions α_1 and α_2 of blocks 1 and 2, Gamma$(1.0, 0.1)$ for the rate parameters $\exp(\theta_{1,h})$ $(h = 1, \ldots, 4)$, and $N(0, 4)$ for the remaining parameters. We generated a Markov chain Monte Carlo sample of size 120,000, discarding the first 20,000 iterations as burn-in iterations and recording every 10-th iteration of the last 100,000 iterations. To detect signs of non-convergence, we exploited the convergence checks of Warnes and Burrows (2010) and, upon discarding the first 20,000 Markov chain Monte Carlo sample points and relabeling the remaining Markov chain Monte Carlo sample points, we inspected trace plots of the rates of change, outdegree, reciprocity, and transitivity parameters, shown in Figures 1, 2, and 3. These convergence checks did not reveal signs of non-convergence. 95% posterior confidence intervals of the parameters are shown in Table 1.

The marginal posterior of the proportions of blocks 1 and 2 (see Figure 4) suggests that there is a small subset of companies, corresponding to block 1 with less than 5% of the companies (posterior median 4.70%), and a large subset of companies, corresponding to block 2 with about 95% of the companies (posterior median 95.31%).

These two subsets of companies deviate from each other in terms of rate of change and outdegree (see Figures 5 and 6). Both the rate of change and the outdegree parameter of block 1 exceed the rate of change and the outdegree parameter of block 2 and, since the rates of change of block 2 tend to be close to 0, it seems that it is the companies of block 1 which shape the evolution of the ownership network. In short, there seems to be a small subset of companies (block 1) which outpaces a large subset of companies (block 2) in terms of the rate of change as well as the desire to accumulate stock of other companies. network.figure.tex In view of Pahor's conjecture, it is tempting to interpret the small subset of companies (block 1) as shadow-financial companies and the large subset of companies (block 2) as non-financial companies. It is possible to make probabilistic statements about which companies belong to blocks 1 and 2, helping detect which companies are shadow-financial companies and which companies are non-financial companies. We do not present them here, because the number

Figure 6: Marginal posterior densities of outdegree parameters of blocks 1 and 2; dashed lines indicate 2.5%, 50%, and 97.5% quantiles.



of companies is large and the individual companies are not well-known.

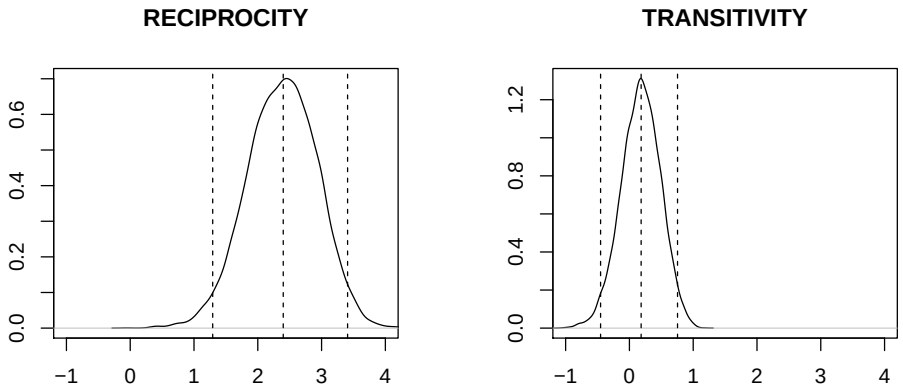
It is worth noting that the rates of change of both subsets of companies in period 4 seem to exceed the rates of change in periods 1—3, which may reflect changes in the economic environment (markets) or legal environment (rules and regulations). In addition, Figure 7 suggests that companies are interested in reciprocating relationships, which may be explained by the desire to align interests and form strategic alliances. Last, but not least, while Pahor (2003) reported a positive tendency towards transitivity among ownerships, Figure 7 suggests that there is no transitivity among ownerships when the partition of the set of companies into shadow-financial companies and non-financial companies is taken into account.

5 Discussion

We have assumed here that a population of nodes is partitioned into unobserved subpopulations, called blocks, and that the parameters of the unobserved continuous-time Markov process which generates the observed networks depend on the subpopulations.

An interesting extension of the proposed modeling framework would be to use subpopulations to restrict the range of dependence. Constraining the range of dependence to sub-

Figure 7: Marginal posterior densities of reciprocity and transitivity parameter; dashed lines indicate 2.5%, 50%, and 97.5% quantiles.



populations makes sense, because it is unreasonable to assume that each edge can depend on all other edges when the population of interest is large. Schweinberger and Handcock (2015) explored such ideas in the context of cross-sectional network data, assuming that the dependence induced by exponential-family random graph models is restricted to subpopulations. Schweinberger and Stewart (2019) used these local dependence models to establish the first statistical consistency results for exponential-family random graphs with non-trivial dependence, and Schweinberger (2019) showed that unobserved block structure can be recovered with high probability under weak dependence and smoothness conditions. Constraining the range of dependence induced by continuous-time Markov processes to subpopulations would likewise make sense, and constitutes an interesting direction for future research.

A second interesting extension would be to extend these models to discrete-time network and behavior data. That would enable researchers to, e.g., detect subsets of nodes that are more prone to social influence than others.

An implementation of the proposed modeling framework in `Delphi`, which builds on the third generation of the Siena software (Snijders et al. 2010), can be found at `www.stat.rice.edu/~ms88/siena/code.html`. An `R` script for solving the label-switching problem described in Section 3.4, based on `R` package `hergm` (Schweinberger and Luna 2018), can be found in Appendix B.

Acknowledgements

I acknowledge support from the National Science Foundation (NSF awards DMS-1513644 and DMS-1812119) and the Netherlands Organisation for Scientific Research (NWO award Rubicon-44606029), and would like to thank Marko Pahor for his willingness to share his data with me.

A Markov chain Monte Carlo algorithm

We combine the following Markov chain Monte Carlo steps by means of cycling or mixing (Tierney 1994). Where possible, we sample from full conditional distributions. Otherwise, we use Metropolis-Hastings steps.

Block structure $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. Sample

$$\mathbf{Z}_i \mid \alpha_{i,1}, \ldots, \alpha_{i,K} \overset{\text{ind}}{\sim} \text{Multinomial}(1; \alpha_{i,1}, \ldots, \alpha_{i,K}),\ i \in \mathcal{N}, \tag{7}$$

where

$$\alpha_{i,k} = \frac{L_i(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, Z_{i,k} = 1)}{\sum_{\mathbf{Z}_i} L_i(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, Z_{i,l} = 1)} \tag{8}$$

and

$$\begin{aligned} L_i(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, Z_{i,k} = 1) \ &= \ \alpha_k \\ \times \ &\left\{ \prod_{m: i_m = i}^{M} \exp\left[-\lambda(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\, h_m\right] \lambda_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\, p_{i_m}(j_m \mid \mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_2) \right\} \\ \times \ &\exp\left[-\lambda(\mathbf{Y}_M, \mathbf{Z}, \boldsymbol{\theta}_1)\, h_{M+1}\right], \end{aligned}$$

where the summation in the denominator of (7) is with respect to all K possible values of $\mathbf{Z}$, the product in (A) is with respect to all changes of directed edges $y_{i,k}$ from node i, and

$$\lambda(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) = \sum_{k=1}^{n} \lambda_k(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1).$$

If either $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ or $p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2)$ do not depend on $\mathbf{Z}$, then the corresponding terms of (A) cancel.

Sequence of changes A_M. Sampling A_M subject to the constraints $\mathbf{Y}(t_0) = \mathbf{Y}_0$ and $\mathbf{Y}(t_1) = \mathbf{Y}_1$ requires non-standard Markov chain Monte Carlo steps that are too space-consuming to describe here. We use Markov chain Monte Carlo steps along the lines of Snijders et al. (2010).

Parameter $\boldsymbol{\alpha}$. If the prior of $\boldsymbol{\alpha}$ is given by a truncated stick-breaking prior, the full conditional distribution of $\boldsymbol{\alpha}$ can be sampled by sampling

$$V_k^\star \overset{\text{ind}}{\sim} \text{Beta}\left(A_k + n_k, B_k + \textstyle\sum_{j=k+1}^{K} n_j\right),\ k = 1, \ldots, K-1,$$

and setting

$$\begin{aligned} \alpha_1 &= V_1^\star \\ \alpha_k &= V_k^\star \prod_{j=1}^{k-1} (1 - V_j^\star),\ k = 2, \ldots, K-1 \\ \alpha_K &= 1 - \sum_{k=1}^{K-1} \alpha_k, \end{aligned}$$

where n_k is the number of nodes in block k ($k = 1, \ldots, K$).

Parameters θ_1 and $\boldsymbol{\theta}_1$. If the rates of change $\lambda_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ are constant and given by θ_1 and the prior of θ_1 is given by Gamma(C, D), we sample θ_1 from its full conditional distribution Gamma$(C+M, D+n)$. Otherwise, we update $\boldsymbol{\theta}_1$ by random-walk Metropolis-Hastings steps, generating candidates from multivariate Gaussian distributions.

Parameter $\boldsymbol{\theta}_2$. We update $\boldsymbol{\theta}_2$ by random-walk Metropolis-Hastings steps, generating candidates from multivariate Gaussian distributions.

B R script for solving the label-switching problem

The label-switching problem described in Section 3.4 can be solved by using R package `hergm` (Schweinberger and Luna 2018).

The following R script undoes the label-switching in the sample of block memberships used in Section 3.4:

```
library(hergm)
set.seed(0)
z <- c(1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 1, 1, 2, 2, 1, 1)
z <- matrix(z, nrow=4, ncol=4, byrow=T)
s <- hergm.relabel_1(max_number=2, indicator=z, number_runs=5, verbose=1)
```

where the first argument specifies the number of blocks; the second argument specifies the sample of block memberships in matrix form, where rows correspond to samples and columns correspond to block memberships of nodes; the third argument specifies the number of runs of the relabeling algorithm, with starting values chosen at random; and the last argument specifies the amount of detail reported by the relabeling algorithm.

While the original sample shows evidence of label-switching,

```
> z
     [,1] [,2] [,3] [,4]
[1,]    1    1    2    2
[2,]    1    1    2    2
[3,]    2    2    1    1
[4,]    2    2    1    1
```

the R script undoes the label-switching,

```
> s$indicator
     [,1] [,2] [,3] [,4]
[1,]    1    1    2    2
[2,]    1    1    2    2
[3,]    1    1    2    2
[4,]    1    1    2    2
```

and reports estimates of posterior classification probabilities,

```
> s$p
     [,1] [,2]
[1,]    1    0
[2,]    1    0
[3,]    0    1
[4,]    0    1
```

where the rows correspond to nodes; the columns correspond to blocks; and element (i, k) of the matrix can be interpreted as an estimate of the posterior probability that node i belongs to block k. In addition, the `R` script reports the optimal permutations obtained at the last iteration of the relabeling algorithm,

```
> s$min_permutations
     [,1] [,2]
[1,]    1    2
[2,]    1    2
[3,]    2    1
[4,]    2    1
```

which can be used to undo the label-switching in samples of block-dependent parameters from the posterior. Here, rows $1, \ldots, 4$ correspond to the optimal permutations of samples $1, \ldots, 4$, respectively.

References

E. Airoldi, D. Blei, S. Fienberg, and E. Xing (2008), "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014.

P. Block, J. Koskinen, J. Hollway, C. E. G.Steglich, and C. Stadtfeld (2018), "Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power," *Social Networks*, 52, 180–191.

C. T. Butts (2008), "A relational event framework for social action," *Sociological Methodolgy*, 38, 155–200.

R. J. Connor, and J. E. Mosiman (1969), "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution," *Journal of the American Statistical Association*, 64, 194–206.

D. Durante, and D. B. Dunson (2014), "Nonparametric Bayes dynamic modelling of relational data," *Biometrika*, 101, 125–138.

T. Ferguson (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

S. E. Fienberg, and S. Wasserman (1981), "Categorical data analysis of single sociometric relations," in *Sociological Methodology*, ed. S. Leinhardt, San Francisco, CA: Jossey-Bass, pp. 156–192.

W. Fu, L. Song, and X. E. (2009), "Dynamic mixed membership blockmodel for evolving networks," in *Proceedings of the 26th Annual International Conference on Machine Learning*.

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

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

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

P. W. Holland, and S. Leinhardt (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.

H. Ishwaran, and L. F. James (2001), "Gibbs Sampling Methods for Stick-breaking Priors," *Journal of the American Statistical Association*, 96, 161–173.

S. Karlin, and H. M. Taylor (1975), *A first course in stochastic processes*, New York: Academic Press.

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

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

J. H. Koskinen, and T. A. B. Snijders (2007), "Bayesian Inference for Dynamic Social Network Data," *Journal of Statistical Planning and Inference*, 137, 3930–3938.

R. W. Krause, M. Huisman, and T. A. B. Snijders (2018), "Multiple imputation for longitudinal network data," *Italian Journal of Applied Statistics*, 30, 33–57.

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

F. Lorrain, and H. C. White (1971), "Structural equivalence of individuals in social networks," *Journal of Mathematical Sociology*, 1, 49–80.

N. M. D. Niezink, and T. A. B. Snijders (2017), "Co-evolution of social networks and continuous actor attributes," *The Annals of Applied Statistics*, 11, 1948–1973.

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

V. Ouzienko, Y. Guo, and Z. Obradovic (2011), "A decoupled exponential random graph model for prediction of structure and attributes in temporal social networks," *Statistical Analysis and Data Mining*, 4, 470–486.

M. Pahor (2003), "Causes and Consequences of Companies' Activity in Ownership Network," Ph.D. thesis, Faculty of Economics, University of Ljubljana, Slovenia.

M. Pahor, J. Prasnikar, and A. Ferligoj (2004), "Building a corporate network in a transition economy: the case of Slovenia," *Post-Communist Economics*, 16, 307–331.

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

M. Schweinberger (2019), "Consistent structure estimation of exponential-family random graph models with block structure," *Bernoulli*, to appear.

M. Schweinberger, and M. S. Handcock (2015), "Local dependence in random graph models: characterization, properties and statistical inference," *Journal of the Royal Statistical Society, Series B*, 77, 647–676.

M. Schweinberger, and P. Luna (2018), "HERGM: Hierarchical exponential-family random graph models," *Journal of Statistical Software*, 85, 1–39.

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

M. Schweinberger, and J. Stewart (2019), "Concentration and consistency results for canonical and curved exponential-family models of random graphs," *The Annals of Statistics*, to appear.

D. K. Sewell (2017), "Network autocorrelation models with egocentric data," *Social Networks*, 49, 113–123.

D. K. Sewell, and Y. Chen (2015), "Latent space models for dynamic networks," *Journal of the American Statistical Association*, 110, 1646–1657.

— (2016), "Latent Space Approaches to Community Detection in Dynamic Networks," *Bayesian Analysis*.

D. K. Sewell, Y. Chen, W. Bernhard, and T. Sulkin (2016), "Model-based longitudinal clustering with varying cluster assignments," *Statistica Sinica*, 26, 205–233.

T. A. B. Snijders (2001), "The Statistical Evaluation of Social Network Dynamics," in *Sociological Methodology*, eds. M. Sobel, and M. Becker, Boston and London: Basil Blackwell, pp. 361–395.

— (2017), "Stochastic actor-oriented models for network dynamics," *Annual Review of Statistics and its Application*, 4, 343–363.

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

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

T. A. B. Snijders, C. E. G. Steglich, M. Schweinberger, and M. Huisman (2010), *Manual for Siena 3.0*, Department of Statistics, University of Oxford, UK.

C. Stadtfeld, T. A. B. Snijders, C. E. G. Steglich, and M. van Duijn (2018), "Statistical Power in Longitudinal Network Studies," *Sociological Methods and Research*, 1–35.

C. E. G. Steglich, T. A. B. Snijders, and M. Pearson (2010), "Dynamic Networks and

Behavior: Separating Selection from Influence," *Sociological Methodology*, 40, 329–393.

M. Stephens (2000), "Dealing with label-switching in mixture models," *Journal of the Royal Statistical Society, Series B*, 62, 795–809.

Y. W. Teh (2010), "Dirichlet Processes," in *Encyclopedia of Machine Learning*, eds. C. Sammut, and G. I. Webb, Springer-Verlag.

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

G. R. Warnes, and R. Burrows (2010), *R package mcgibbsit: Warnes and Raftery's MCGibbsit MCMC diagnostic.*

S. Wasserman (1977), "Random directed graph distributions and the triad census in social networks," *Journal of Mathematical Sociology*, 5, 61–86.

— (1980), "Analyzing Social Networks as Stochastic Processes," *Journal of the American Statistical Association*, 75, 280–294.

S. Wasserman, and C. Anderson (1987), "Stochastic a posteriori blockmodels: Construction and assessment," *Social Networks*, 9, 1–36.

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

Y. Zhao, E. Levina, and J. Zhu (2012), "Consistency of community detection in networks under degree-corrected stochastic block models," *The Annals of Statistics*, 40, 2266–2292.