Estimating Continuous-Time Markov Processes with Unobserved Equivalence Classes from 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 popular approach to modeling discrete-time network data assumes that the discrete-time network data were generated by an unobserved continuous-time Markov process. While such models can capture a wide range of dependencies in discrete-time network data 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 equivalence classes with distinct probability laws. The resulting models can capture unobserved heterogeneity across nodes and admit model-based clustering of nodes based on network properties chosen by researchers. We use Bayesian methods to demonstrate the usefulness of the models, applying them to discrete-time observations of the ownership network of companies in Slovenia in its critical transition from a communist command and control economy to a market economy.

Keywords: finite mixture models, model-based clustering, social networks, stochastic actor-oriented models

1 Introduction

The statistical analysis of network data has seen a surge of interest in statistics and related fields (Kolaczyk 2009). Network data help understand a connected world, how connections depend on other connections, how connections change over time, and how connections affect other outcomes of interest (e.g., public health, national security).

We focus here on longitudinal network data, consisting of observations of a network with a given population of nodes at discrete time points. A widely used approach to modeling discrete-time network data assumes that the discrete-time network data were generated by an unobserved continuous-time Markov process, which are known as stochastic actor-oriented models in social network analysis (Snijders 2017); some recent developments are described in, e.g., Koskinen and Snijders (2007), Steglich et al. (2010), Snijders et al. (2010), Niezink and Snijders (2017), Block et al. (2018), and Krause et al. (2018). While such models can capture a wide range of dependencies in discrete-time network data 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 equivalence classes with distinct probability laws. 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 to the unobserved block structure and parameters of continuous-time Markov processes from discrete-time network data, we propose Bayesian inference implemented by Markov chain Monte Carlo methods, building on Koskinen and Snijders (2007) and Snijders et al. (2010). The issue of non-identifiable parameters, arising from the invariance of the likelihood function to permutations of block labels, is solved in a Bayesian decision theoretic framework, by defining a suitable loss function and minimizing the posterior expected loss along the lines of Stephens (2000). To reduce computing time, we propose a novel stochastic version of the relabelling algorithm. We demonstrate the usefulness of the models by applying them to discrete-time observations of the ownership network of companies in Slovenia in its critical transition from a communist command and control economy to a market economy (Pahor 2003; Pahor, Prasnikar, and Ferligoj 2004). We 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 shares of other companies. These results lend support to the conjecture of Pahor (2003) that the ownership network consists of a small subset of shadow financial companies that trades shares more often than others, and accumulates more shares than others.

The remainder of the paper is structured as follows. Section 2 proposes stochastic block models of discrete-time relational data. Section 3 proposes Markov chain Monte Carlo methods to generate samples from the posterior distribution and an algorithm to solve the label-switching problem. In Section 4, we demonstrate the methods by an application to discrete-time observations of the ownership network of companies in Slovenia in its critical transition from a communist command and control-style economy to a market economy.

Other, related models It is worth noting that the equivalence classes resemble blocks in stochastic block models (Nowicki and Snijders 2001; Bickel and Chen 2009). Stochastic block models of cross-sectional network data were introduced by Holland et al. (1983) and Nowicki and Snijders (2001) and extended by Airoldi et al. (2008), and combined with latent space models (Handcock, Raftery, and Tantrum 2007) and exponential-family random graph models (Schweinberger and Handcock 2015). Stochastic block models of continuous-time network data were introduced by Fu et al. (2009). But, in contrast to the models we propose, these stochastic block models assume that relationships are independent conditional on block memberships, and hence cannot capture a wide range dependencies among relationships.

Other, related models are discrete-time latent space models (Sewell and Chen 2015) and discrete-time exponential-family random graph models (Hanneke et al. 2010; Krivitsky and Handcock 2014), among others. However, the first class of models does not allow to model a wide range of dependencies (despite capturing a weak tendency towards transitivity), while the second class of models cannot capture unobserved heterogeneity (although it can capture observed heterogeneity through covariates).

2 Model

We consider relational data in the form of $n \times n$ matrices $\mathbf{Y}$, where the rows and columns of $\mathbf{Y}$ correspond to nodes $1, \dots, n$ and the elements $y_{i,j} \in \{0, 1\}$ of $\mathbf{Y}$ correspond to directed relationships between ordered pairs of nodes (i, j), corresponding to the directed edges of a directed graph. By convention, self-relationships are discarded by restricting the elements on the main diagonal of $\mathbf{Y}$ to be 0. The set of nodes is denoted by $\mathcal{N}$ and the set of 0–1 matrices with 0 main diagonal is denoted by $\mathcal{Y}$. Let $\{\mathbf{Y}(t),\, t \in \mathcal{T}\}$ be a continuous-time random process with finite state space $\mathcal{Y}$, and assume that observations of $\{\mathbf{Y}(t),\, t \in \mathcal{T}\}$ at discrete times in time interval $\mathcal{T}$ are available.

2.1 Model

Motivated by the incomplete nature of the data (discrete-time observations of continuous-time random processes), we consider simple complete-data models as follows.

Suppose that there are K blocks, and that $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are vectors, where element Z_{ik} of vector $\mathbf{Z}_i$ is 1 if node $i = 1, \dots, n$ is member of block $k = 1, \dots, K$ and 0 otherwise. A stochastic blockmodel assumes that the $\mathbf{Z}_i$ are i.i.d. random vectors:

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

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

We condition on the first observation of the random process $\mathbf{Y}(t)$, and denote the time by $t_0 \in \mathcal{T}$ and the state of the random process $\mathbf{Y}(t)$ at time $t = t_0$ by $\mathbf{Y}_0 \in \mathcal{Y}$. Conditional on $\mathbf{Z}$ and $\mathbf{Y}(t_0) = \mathbf{Y}_0$, $\mathbf{Y}(t)$ is governed by a right-continuous Markov as follows. Suppose that the Markov process $\mathbf{Y}(t)$ is at graph $\mathbf{Y}$ at time t, where $\mathbf{Y} \in \mathcal{Y}$ and $t \in \mathcal{T}$. The transition probability of going from graph $\mathbf{Y}$ to graph $\mathbf{Y}^\star \neq \mathbf{Y} \in \mathcal{Y}$ during $(t, t+h)$ is assumed to be of the form

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

where

$$P[Y_{i,j}(t+h) = 1 - Y_{i,j} \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}] = q_{i,j}(\mathbf{Y}, \mathbf{Z})\, h + o(h) \tag{2}$$

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

$$q_{i,j}(\mathbf{Y}, \mathbf{Z}) = \lim\nolimits_{h \downarrow 0} \frac{P[Y_{i,j}(t+h) = 1 - Y_{i,j} \mid \mathbf{Y}(t) = \mathbf{Y}, \mathbf{Z}]}{h}$$

denotes the rate of change of $Y_{i,j}$ given $\mathbf{Y}(t) = \mathbf{Y}$. Two remarks are in place. First, transition probabilities of the form (1) assume that changes of edges in short time intervals $(t, t+h)$ are independent and changes of graphs are local in the sense that the probability that more than one edge changes is $o(h)$ (Holland and Leinhardt 1977; Wasserman 1977, 1980; Snijders 2001, 2006, 2009). Second, while changes of edges in short time intervals $(t, t+h)$ are independent, the changes of edges can depend on the graph at time t, allowing to model transitivity and other forms of dependence.

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). Let

$$q_{i,j}(\mathbf{Y}, \mathbf{Z}) = q_i(\mathbf{Y}, \mathbf{Z})\; p_i(j \mid \mathbf{Y}, \mathbf{Z}), \tag{3}$$

where the function $q_i(\mathbf{Y}, \mathbf{Z})$ satisfies $q_i(\mathbf{Y}, \mathbf{Z}) > 0$ (all i) and can be interpreted as the rate of change of row i of matrix $\mathbf{Y}$, while the function $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ satisfies $0 < p_i(j \mid \mathbf{Y}, \mathbf{Z}) < 1$ (all i, j) and $\sum_{j \neq i}^{n} p_i(j \mid \mathbf{Y}, \mathbf{Z}) = 1$ (all i) and can be interpreted as the conditional probability that element $y_{i,j}$ of column j changes given that row i changes. The change of element $y_{i,j}$ can be interpreted along the lines of discrete choice models in economics (McFadden 1974): the change of element $y_{i,j}$ in row i of $\mathbf{Y}$ can be interpreted as actor i choosing an actor j from the discrete set of choices $\mathcal{N}_i$ with probability $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ and changing $y_{i,j}$ to $1 - y_{i,j}$.

The rate of change $q_i(\mathbf{Y}, \mathbf{Z})$ and the conditional probability mass function $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ can be parameterized as follows. A convenient exponential form of the rate of change $q_i(\mathbf{Y}, \mathbf{Z})$ is given by

$$q_i(\mathbf{Y}, \mathbf{Z}) \equiv q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) = \exp[\boldsymbol{\eta}_{i1}^T(\mathbf{Z}, \boldsymbol{\theta}_1)\, s_{i1}(\mathbf{Y})],$$

where $\boldsymbol{\eta}_{i1}(\mathbf{Z}, \boldsymbol{\theta}_1)$ is a vector of parameters and $s_{i1}(\mathbf{Y})$ is a vector of statistics. An exponential family form of the conditional probability mass function $p_i(j \mid \mathbf{Y}, \mathbf{Z})$ with support $\mathcal{N}_i = \mathcal{N} \setminus i$ is given by

$$p_i(j \mid \mathbf{Y}, \mathbf{Z}) \equiv p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2) = \exp[\boldsymbol{\eta}_{i2}^T(\mathbf{Z}, \boldsymbol{\theta}_2) s_{i2}(j, \mathbf{Y}) - \psi_i(\mathbf{Z}, \boldsymbol{\theta}_2)],\; j \in \mathcal{N}_i,$$

where $\boldsymbol{\eta}_{i2}(\mathbf{Z}, \boldsymbol{\theta})$ is a vector of parameters and $s_{i2}(j, \mathbf{Y})$ is a vector of statistics, and

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

ensures that $p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2)$ sums to 1.

The vectors of parameters $\boldsymbol{\eta}_{i1}(\mathbf{Z}, \boldsymbol{\theta}_1)$ and $\boldsymbol{\eta}_{i2}(\mathbf{Z}, \boldsymbol{\theta}_2)$ of node i depend on the block membership $\mathbf{Z}_i$ of node i:

$$\boldsymbol{\eta}_{i1}(\mathbf{Z}, \boldsymbol{\theta}_1) = \boldsymbol{\theta}_1^T \mathbf{Z}_i,\ i = 1, \dots, n,$$

and

$$\boldsymbol{\eta}_{i2}(\mathbf{Z}, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_2^T \mathbf{Z}_i,\ i = 1, \dots, n,$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are matrices of parameters, with columns corresponding to the block-dependent parameter vectors. Some parameters may be shared across blocks, in which case some rows of $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$ consist of constants. We give examples of model specifications in Section 4.

3 Bayesian estimation

We start by deriving the complete-data likelihood function, given complete observation of the block structure $\mathbf{Z}$ and the Markov process $\mathbf{Y}(t)$, in Section 3.1. Since, by the incomplete nature of the data (discrete observations of continuous-time Markov process), the incomplete-data likelihood is intractable, we exploit Markov chain Monte Carlo data augmentation methods to generate samples from the posterior. We describe the general Markov chain Monte Carlo data augmentation approach in Section 3.3, and describe the specific Markov chain Monte Carlo steps to generate samples from posteriors in Section A. Solutions of the label-switching problem, which is rooted in the invariance of the complete-data likelihood to the labeling of the blocks, are discussed in Section 3.4. Throughout, we assume that there are two observation points $t_0 < t_1$, since, by the Markov property of the random process, the extension to multiple, non-overlapping time intervals is straightforward.

3.1 Likelihood function

We start with the complete-data likelihood of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ given complete observations of block structure $\mathbf{Z}$ and Markov process $\mathbf{Y}(t)$ in some time interval $[t_0, t_1]$, where $t_0 < t_1$. A complete observation of the Markov process in time interval $[t_0, t_1]$ 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), where the ordered pair of nodes (i_m, j_m) corresponds to the directed edge $y_{i_m j_m}$ which turns into $1 - y_{i_m j_m}$ at time $t_0 + \sum_{k=1}^m h_k$.

The complete-data likelihood of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ given W_M and $\mathbf{Z}$ factors according to

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

The complete-data likelihood of $\boldsymbol{\alpha}$ given $\mathbf{Z}$ is given by

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

By the theory of continuous-time Markov processes (Karlin and Taylor 1975; Norris 1997; Resnick 2002), the complete-data likelihood of $\boldsymbol{\theta}_1$ given W_M and $\mathbf{Z}$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}_1; W_M, \mathbf{Z}) &= \left\{\prod_{m=1}^{M} \exp\left[-q(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\, h_m\right]\, q_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\right\} \\ &\times \exp\left[-q(\mathbf{Y}_M, \mathbf{Z}, \boldsymbol{\theta}_1)\, h_{M+1}\right], \end{aligned}$$

where

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

and

$$h_{M+1} = t_1 - \sum_{m=1}^{M} h_m.$$

The complete-data likelihood of $\boldsymbol{\theta}_2$ given W_M and $\mathbf{Z}$ is given by

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

The incomplete-data likelihood of $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ given W_M can be obtained by summing the complete-data likelihood $L(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, \mathbf{Z})$ with respect to all possible values of $\mathbf{Z}$, which is intractable since there are $K^n = \exp[n \log K]$ possible values of $\mathbf{Z}$. The incomplete-data likelihood of $\boldsymbol{\alpha}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ given discrete observations of the Markov process is likewise intractable. To generate samples from the posterior distribution, we therefore use Markov chain Monte Carlo data augmentation (see Section 3).

3.2 Priors

We consider non-parametric stick-breaking priors (Ferguson 1973; Ishwaran and James 2001; Teh 2010), which help to sidestep the problematic selection of the number of blocks K.

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

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

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 $q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ are constant across $\mathbf{Y}$, $\mathbf{Z}$, and $i \in \mathcal{N}$,

$$q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) = \exp[\boldsymbol{\eta}_{i1}^T(\mathbf{Z}, \boldsymbol{\theta}_1) s_{i1}(\mathbf{Y})] = \rho > 0, \; i \in \mathcal{N},$$

then it is convenient to use the ρ-parameterization and the conjugate ρ-prior given by

$$\rho \,|C, D \sim \mathrm{Gamma}(C, D).$$

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

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

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

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

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

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

3.3 Markov chain Monte Carlo data augmentation

By the incomplete nature of the data (discrete observations of continuous-time Markov process), the incomplete-data likelihood is intractable (see Section 3.1). We therefore exploit Markov chain Monte Carlo data augmentation methods to generate samples from the posterior.

Suppose that observations of the continuous-time Markov process at times t_0 and t_1 are available, where $t_0 < t_1$. A complete observation of the continuous-time Markov process in time interval $[t_0, t_1]$ would be constituted by the number of changes M in time interval $[t_0, t_1]$ and the sequence $W_M = (h_m, i_m, j_m)_{m=1}^M$. The augmented data considered here are constituted by the number of changes M in time interval $[t_0, t_1]$ and the sequence $A_M = (i_m, j_m)^M$ rather than the sequence $W_M = (h_m, i_m, j_m)_{m=1}^M$, i.e., the holding times are integrated out. The resulting Markov chain Monte Carlo algorithm is thought to be preferable to other Markov chain Monte Carlo algorithms, Markov chain Monte Carlo algorithms that can handle state spaces with changing dimension (Richardson and Green 1997; Koskinen and Snijders 2007), both in terms of simplicity and Markov chain Monte Carlo simulation variance (Snijders et al. 2010).

The distribution of the number of changes M in time interval $[t_0, t_1]$ can be obtained either exactly or approximately. If the rates of change $q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ are constant and given by ρ, the distribution of the number of changes M in time interval $[t_0, t_1]$ is Poisson$(\rho\,(t_1 - t_0)\,n)$ and the probability of M changes in time interval $[t_0, t_1]$ is given by

$$P[T_M \leq t_1 < T_{M+1}] = \exp[\rho\,(t_1 - t_0)\,n]\frac{[\rho\,(t_1 - t_0)\,n]^M}{M!}.$$

Otherwise the distribution of the number of changes M in time interval $[t_0, t_1]$ can be approximated as follows. The distribution of $\sum_{m=1}^{M} h_m$ is a convolution of negative exponential distributions with parameters $q_{i_1}(\mathbf{Y}_0, \mathbf{Z})$, $\ldots$, $q_{i_M}(\mathbf{Y}_{M-1}, \mathbf{Z})$ and thus, by the central limit theorem, the distribution of $T_M = \sum_{m=1}^{M} h_m$ is approximately Gaussian $N(\mu_T, \sigma_T^2)$ with mean $\mu_T = \sum_{h=1}^{M} 1/q_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z})$ and variance $\sigma_T^2 = \sum_{h=1}^{M} 1/q_{i_m}^2(\mathbf{Y}_{m-1}, \mathbf{Z})$. Along the lines of Snijders et al. (2010), it can be shown that the probability that there are M changes in time interval $[t_0, t_1]$ is given by

$$P[T_M \leq t_1 < T_{M+1}] \approx \frac{p_{\mu_T, \sigma_T^2}(t_1)}{q_{i_M}(\mathbf{Y}_{M-1}, \mathbf{Z})},$$

where $p_{\mu_T, \sigma_T^2}(t_1)$ is the Gaussian probability density of t_1.

The probability of M changes in time interval $[t_0, t_1]$ and the sequence of changes A_M is therefore given by

$$P[T_M \leq t_1 < T_{M+1}] \prod_{m=1}^{M} \exp\left[-q(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1)\,h_m\right]\,q_{i_m}(\mathbf{Y}_{m-1}, \mathbf{Z}, \boldsymbol{\theta}_1),$$

where $P[T_M \leq t_1 < T_{M+1}]$ is either given by (3.3) or (3.3).

Markov chain Monte Carlo steps to sample from the posterior are outlined in Appendix A.

3.4 Label-switching problem

We solve the label-switching problem of Markov chain Monte Carlo algorithms, which stems from the invariance of the likelihood function to the labeling of blocks, by following a Bayesian decision-theoretic approach along the lines of Stephens (2000) and Schweinberger and Handcock (2015). In other words, we choose a loss function and minimize the posterior expected loss.

A simple loss function used by Schweinberger and Handcock (2015) is given by

$$f(\mathbf{Q}; \mathbf{Z}) = \min_{\nu} f_0(\mathbf{Q}; \nu(\mathbf{Z})),$$

where

$$f_0(\mathbf{Q}; \mathbf{Z}) = -\log \prod_{i=1}^{n} s_{iC_i}, \tag{4}$$

Figure 1: Trace plots of rate parameters of blocks 1 and 2

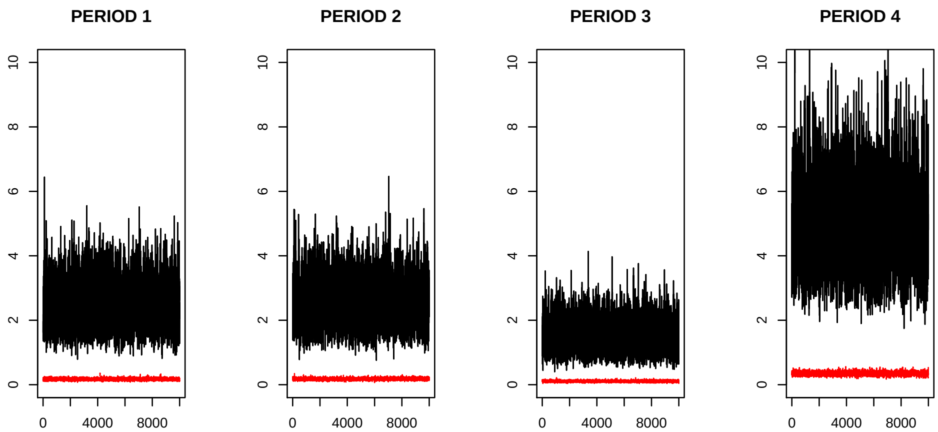
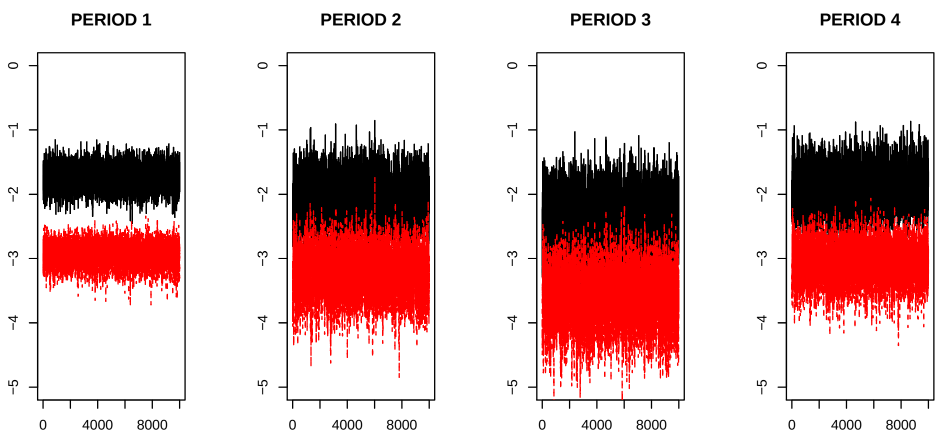


Figure 2: Trace plots of outdegree parameters of blocks 1 and 2



where $\mathbf{Q} = (q_{ik})$, q_{ik} is the probability that node i is reported to be member of block k, $C_i = \sum_{k=1}^{K} k Z_{ik}$ is the block of node i, and $\nu(\mathbf{Z})$ is a permutation of $\mathbf{Z}$. It is desired to report the $\mathbf{Q}$ such that the posterior expectation of loss (3.4) is minimal. The posterior expected loss can be approximated by a Markov chain Monte Carlo sample average. Algorithms are discussed by Schweinberger and Handcock (2015).

4 Applications

We consider an application to cross-ownership networks where the number of blocks is known but the block membership is unknown. Pahor (2003); Pahor, Prasnikar, and Ferligoj (2004) studied the directed cross-ownerships among 413 non-financial companies in Slovenia observed at 5 time points between 2000 and 2002, where $y_{i,j} = 1$ means that company i holds stock market shares of company j. The observations fall into a period in which Slovenia

Figure 3: Trace plots of reciprocity and transitivity parameter

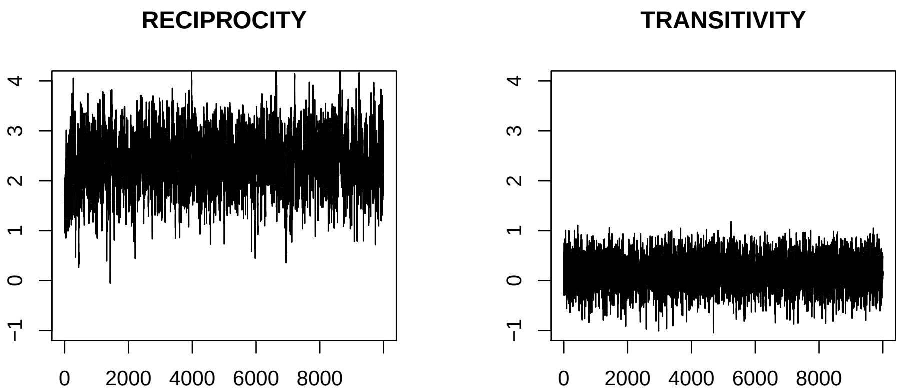
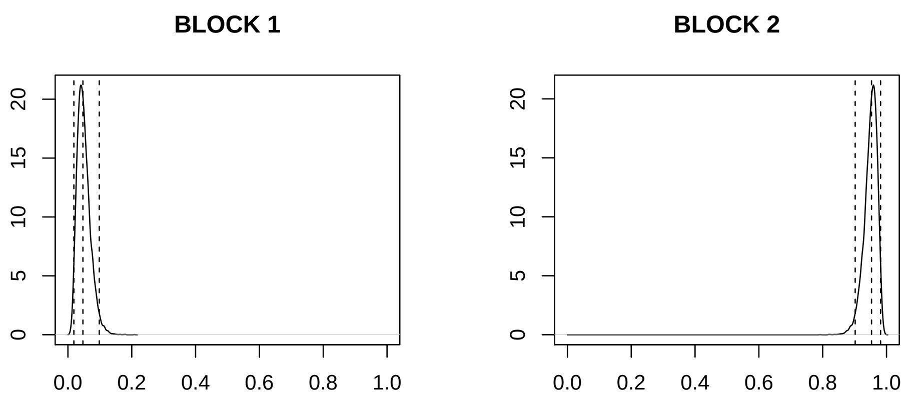
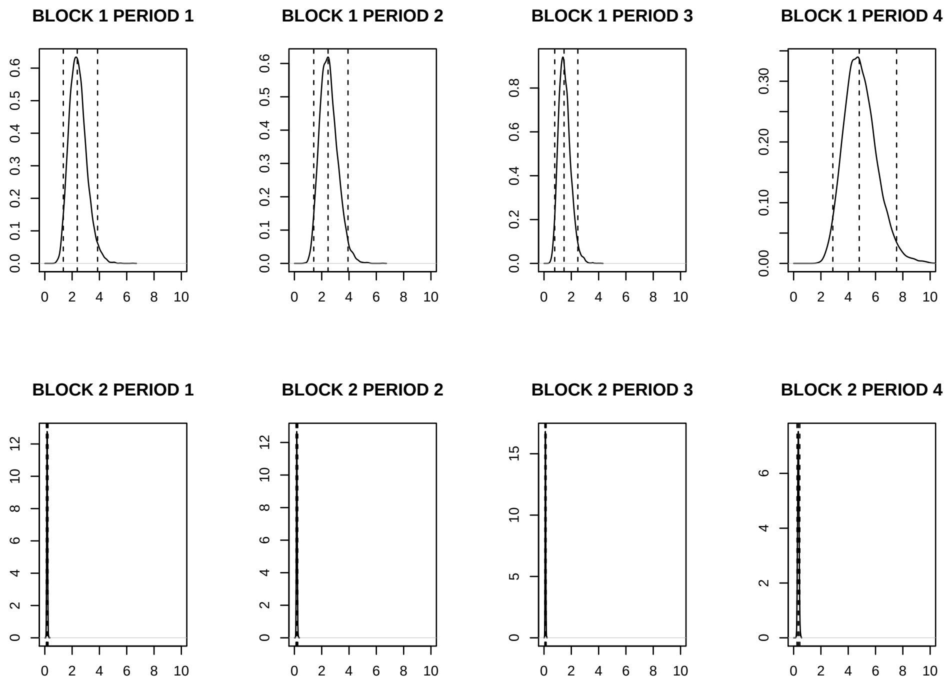


Figure 4: Marginal posterior densities of proportions of blocks 1 and 2; dashed lines indicate 2.5%, 50%, and 97.5% quantiles



managed to make the transition from a communist command and control-style economy to a market economy at the level of Western Europe. Pahor (Pahor 2003; Pahor, Prasnikar, and Ferligoj 2004, and personal communication) suspected that the data set includes 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 investing in and trading shares of other companies. Shadow financial companies are thought to buy and sell shares more frequently and accumulate more shares through time than non-financial companies. The problem is that it was not observed which companies were shadow financial companies. We use a subset of $n = 165$ companies located in the central region of Slovenia around its capital Ljubljana, the most prosperous region of Slovenia, and use the models described above to find out which companies were shadow financial companies and how shadow financial

Figure 5: Marginal posterior densities of rate parameters of blocks 1 and 2; dashed lines indicate 2.5%, 50%, and 97.5% quantiles



companies deviate from non-financial companies.

To capture shadow financial and non-financial companies, we consider $K = 2$ blocks and let Z_{i1} and Z_{i2} be indicators of block memberships, where $Z_{i1} = 1$ and $Z_{i2} = 0$ if i belongs to block 1 and $Z_{i1} = 0$ and $Z_{i2} = 1$ otherwise. The rate of change of period h ($h = 1, \ldots, 4$) and node i ($i = 1, \ldots, 165$) is of the form

$$q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1) = \exp[\theta_{1h} + \theta_{15} Z_{i1} + \theta_{16} Z_{i2}],$$

where θ_{1h} is the baseline rate parameter of period h ($h = 1, \ldots, 4$) and θ_{15} and θ_{16} represent the deviation of the rate parameters of blocks 1 and 2 from the baseline rate parameter, respectively, where θ_{15} is constrained to 0 to make the model identifiable. The inclusion of the rate parameters θ_{15} and θ_{16} allows one subset of companies to buy and sell shares 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

$$p_i(j \mid \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_2) = \exp\left[\eta_{i21}\, c_{i21}(\mathbf{Y}) + \eta_{i22}\, c_{i22}(\mathbf{Y}) + \eta_{i23}\, c_{i23}(\mathbf{Y}) - \psi(\mathbf{Z}, \boldsymbol{\theta}_2)\right],$$

where the change statistics $c_{i21}(j, \mathbf{Y})$, $c_{i22}(j, \mathbf{Y})$, and $c_{i23}(j, \mathbf{Y})$ correspond to the change in

Table 1: 95% posterior confidence intervals of parameters

	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 number of relationships, reciprocated relationships, and transitive relationships due to the change in relationship $y_{i,j}$, and the parameters η_{i21}, η_{i22}, and η_{i23} are given by

- $\eta_{i21} = \theta_{2h} + \theta_{25} Z_{i1} + \theta_{26} Z_{i2}$, where θ_{2h} is the baseline outdegree parameter of period h ($h = 1, \ldots, 4$) and θ_{25} and θ_{26} represent the deviation in outdegree parameters of blocks 1 and 2 from the baseline outdegree parameter, respectively, where θ_{25} is constrained to 0 to make the model identifiable;

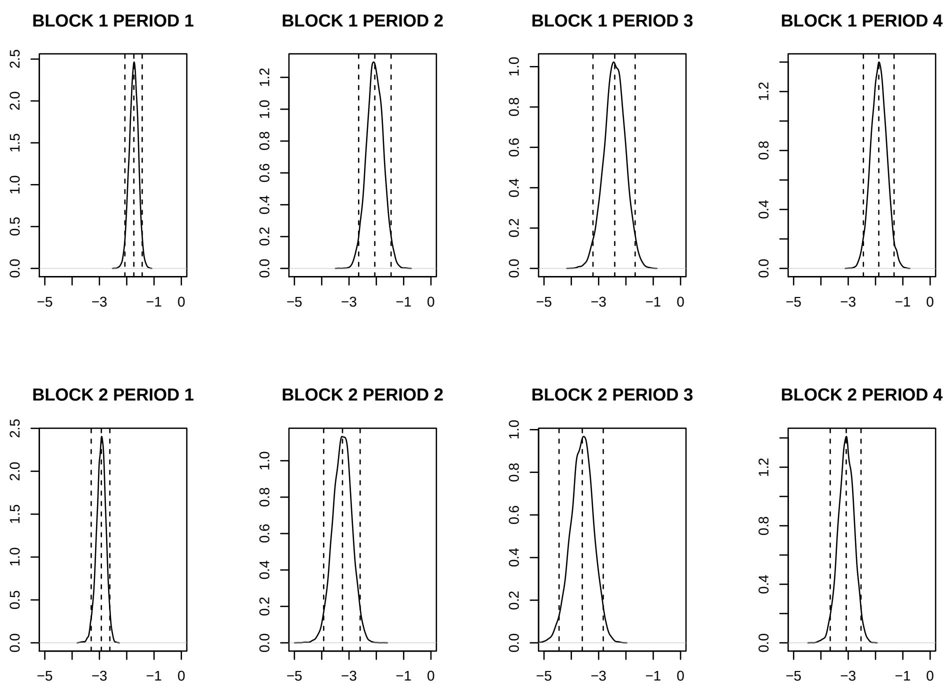
- $\eta_{i22} = \theta_{27}$ is the reciprocity parameter;

- and $\eta_{i23} = \theta_{28}$ is the transitivity parameter.

The inclusion of the outdegree parameters θ_{25} and θ_{26} allows one subset of companies to accumulate more shares through time than the other. The observed number of changes between the 5 observations of the cross-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, so that the number of changes is small and the observed networks sparse. We therefore consider the inclusion of transitivity parameter warranted and do not expect model degeneracy (Section 2). 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 baseline rate parameters $\rho_h = \exp[\theta_{1h}]$ ($h = 1, \ldots, 4$), 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

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

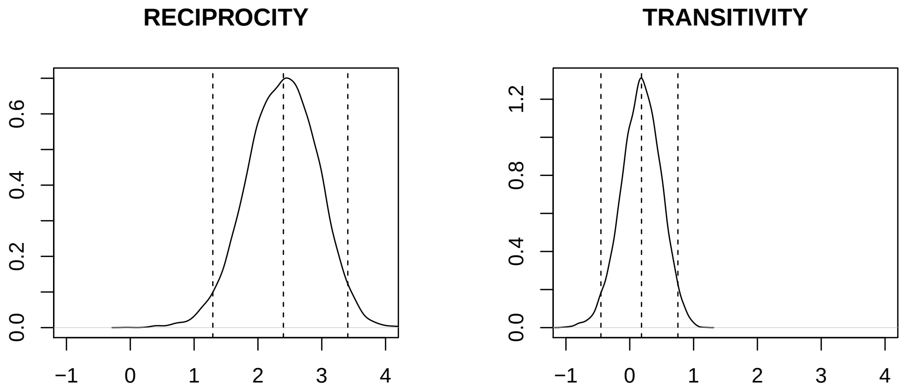


companies (posterior median 4.70%), and a large subset of companies, corresponding to block 2 with about 95% of the companies (posterior median 95.31%).

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 shares of other companies. In view of the expectations of Pahor, 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 of companies is large and the individual companies are not well-known.

It is worthwhile to mention 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

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



addition, companies seem to desire reciprocated relationships (see Figure 7), which may be explained by the desire to align interests. There is, however, not much evidence to suggest that companies invest in transitive relationships, though the large outdegree parameter of companies of block 1 implies that companies of block 1 tend to increase the number of relationships to other companies of block 1 and, as a by-product, tend to increase the number transitive relationships within block 1.

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 processes generating the observed networks depend on the subpopulations. One interesting extension of the proposed modeling framework would be to use subpopulations to limit the range of dependence. Limiting the range of dependence to subpopulations makes sense, because it is unreasonable to assume that each edge can depend on all other edges when the population of interest is large. In fact, models which do not contrain the range of dependence and induce strong dependence throughout the network can suffer from the so-called phenomenon of model near-degeneracy (Handcock 2003; Schweinberger 2011; Chatterjee and Diaconis 2013). Schweinberger and Handcock (2015) explored this idea in the context of cross-sectional network data, assuming that the dependence induced by exponential-family random graph models is limited to subpopulations, and Schweinberger and Stewart (2019) and Schweinberger (2018) used these local dependence models to establish the first statistical consistency results for exponential-family random graphs with non-trivial dependence. Limiting the range of dependence induced by continuous-time Markov processes to subpopulations would likewise make sense, and constitutes an interesting direction for future research.

Acknowledgements

We acknowledge support from the National Science Foundation (NSF awards DMS-1513644 and DMS-1812119) and would like to thank Marko Pahor for his willingness to share his data with us.

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

$$\mathbf{Z}_i \mid \alpha_{i,1}, \ldots, \alpha_{i,K} \overset{\text{ind}}{\sim} \text{Multinomial}(1; \alpha_{i,1}, \ldots, \alpha_{i,K}),\ i \in \mathcal{N}, \tag{5}$$

where

$$\alpha_{i,k} = \frac{L_i(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, Z_{ik} = 1)}{\displaystyle\sum_{\mathbf{Z}_i} L_i(\boldsymbol{\alpha}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; W_M, Z_{il} = 1)} \tag{6}$$

and

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

where the summation in the denominator of (5) 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_{ik} from node i, and

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

If either $q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ 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 Markov chain Monte Carlo steps along the lines of Snijders, Koskinen, and Schweinberger (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

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

and setting

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

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

Parameters ρ and $\boldsymbol{\theta}_1$. If the rates of change $q_i(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}_1)$ are constant and given by ρ and the prior of ρ is given by Gamma(C, D), we sample ρ 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.

References

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

P. J. Bickel, and A. Chen (2009), "A nonparametric view of network models and Newman-Girvan and other modularities," in *Proceedings of the National Academy of Sciences*, Vol. 106, pp. 21068–21073.

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.

S. Chatterjee, and P. Diaconis (2013), "Estimating and understanding exponential random graph models," *The Annals of Statistics*, 41, 2428–2461.

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.

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

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*.

M. S. Handcock (2003), "Assessing degeneracy in statistical models of social networks," Tech. rep., Center for Statistics and the Social Sciences, University of Washington, www.csss.washington.edu/Papers.

M. S. Handcock, A. E. Raftery, and J. M. Tantrum (2007), "Model-based clustering for social networks," *Journal of the Royal Statistical Society, Series A (with discussion)*, 170, 301–354.

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

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.

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.

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

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.

J. R. Norris (1997), *Markov Chains*, Cambridge: Cambridge University Press.

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

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.

S. I. Resnick (2002), *Adventures in Stochastic Processes*, Boston: Birkhäuser.

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

M. Schweinberger (2011), "Instability, sensitivity, and degeneracy of discrete exponential families," *Journal of the American Statistical Association*, 106, 1361–1370.

— (2018), "Consistent structure estimation of exponential-family random graph models with block structure," Available at https://arxiv.org/abs/1702.07801.

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

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.

— (2006), "Statistical methods for network dynamics," in *Proceedings of XLIII Scientific Meeting, Italian Statistical Society*, eds. S. R. Luchini et al., pp. 281–296.

— (2009), "Longitudinal Methods of Network Analysis," in *Encyclopedia of Complexity and System Science*, ed. B. Meyers, Springer-Verlag, pp. 5998–6013.

— (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.

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.