Modeling the dynamics of social networks using Bayesian hierarchical blockmodels

Abel Rodríguez
Department of Applied Mathematics and Statistics
University of California, Santa Cruz

Abstract

We introduce a new class of dynamic models for networks that extends stochastic blockmodels to settings where the interactions between a group of actors are observed at multiple points in time. Our goal is to identify structural changes in model features such as differential attachment, homophily by attributes, transitivity and clustering as the network evolves. Our focus is on Bayesian inference, so the models are constructed hierarchically by combining different classes of Bayesian nonparametric priors. The methods are illustrated through a simulation study and two real data sets.

1 Introduction

A social network is a model for the interactions between multiple individuals u organizations (often called actors), who are connected to each other by one or more specific types of relationships, such as collaboration, kinship, financial exchange, dislikes or beliefs. Typically, the data associated with the network includes a measurement of the presence/absence or the level of interdependence among some or all of the actors in the network, along with predictors describing the features of the actors. The objectives of the model might include the estimation of global network features (such as the degree of connectivity), or local features (such as the unobserved interdependency between two actors).

Social networks are widely used in areas as distinct as psychology, sociology, epidemiology and finance. For example, consider the application of social networks to two problems in defense and national security. In the last 20 years, the armed conflicts around the world have progressively tended to become fourth generation warfare (4GW), i.e., conflicts in which one of the major participants is not a state but rather a violent non-state actor. Hence, the enemy in 4GW typically comprises a large number of small and semi-independent units, which interact and coordinate to varying degrees. Identifying factions encompassing these actors and understanding how they interact is a critical intelligence task that can be addressed using social network models. Similarly, social network modeling can be used in epidemiological applications to understand how emergent diseases are transmitted, allowing for better design of interventions and public response plans.

Constructing probabilistic models that yield networks with realistic features is a major challenge in the development of formal inferential tools for social network data. In practice, social networks exhibit differential propensities (some actors tend to be more popular than others, see for example Davis & Leinhardt, 1972), transitivity (two actors that have a tie to a third actor are more likely to be tied together than actors that do not, see for example White et al., 1976), homophily by attributes (the tendency of actors with similar – or completely opposite – features to have a higher probability of presenting a link, McPherson et al., 2001), and clustering (the tendency of individuals to form communities beyond what can be explained by popularity, transitivity and homophily, Wasserman & Faust, 1994).

Exponentially parameterized random graph (EPRG) models, also known as p^* models in the psychology and sociology literature (Frank & Strauss, 1986; Wasserman & Pattison, 1996), extend the work of Besag (1974) from spatial statistics to general network models. EPRG models attempt to place a probability distribution directly on the full sociomatrix. This distribution usually depends on low-dimensional summaries of the network topology such as the number of edges, the number of two-stars and the number of triangles (Snijders, 2002). Although EPRG models can in principle model transitivity in network structure, work by Besag (2000), Handcock (2000) and Snijders (2002) suggests that they struggle to capture local structure associated with realistic social networks.

Latent social space (LSS) models (Hoff et al., 2002; Handcock et al., 2007) represent a second alternative to understand social networks. They foster transitivity by associating a coordinate in a latent Euclidean space with each actor in the network; if actor i is close to actor j in the social space, and actor j is close to actor k, then typically actors i is close to actor k. The location of each actor in the social space is estimated from the data rather than fixed in advance, which allows us to automatically learn about the underlying similarities between the subjects. The drawback of LSS models it that they cannot easily model clustering or distinguish between homophily and stochastic equivalence (Hoff, 2007).

In this paper we focus on stochastic blockmodels (Wang & Wong, 1987; Airoldi et al., 2008), which divide actors into tightly knit factions, and then model interactions between these factions. Stochastic blockmodels can be conceived as LSS models where the social space is discrete; they are particularly appealing because of their interpretability: factions are meaningful social constructs that are often driven by unobserved (or unobservable) variables.

The models we develop are constructed hierarchically by combining different classes of Bayesian nonparametric priors. Network models that employ nonparametric Bayesian priors have recently been proposed by Kemp et al. (2006) and Xu et al. (2006); they treat the number of factions (along with faction membership) as a random quantity that is to be estimated from the data. We extend these models in two directions. First, we develop a framework for network models based on generalized linear models that allows us to account not only for clustering, but also for differential attachment and homophily. Second, we embed this single-network models into a nonparametric version of a hidden Markov model to generate a hierarchical model that allows us to identify structural changes in the network. Since posterior distributions for these models are not analytically tractable, all computation is carried out using Markov chain Monte Carlo (MCMC) algorithms (Robert & Casella,

2005).

The remaining of the paper is organized as follows: Section 2 describes a general frameworks for modeling a single social network and reviews the use of nonparametric mixture priors in the context of social networks. This framework is illustrated in Section 3 with a classical example that demonstrates the usefulness of social network models in defense and national security applications. Section 4 extends the single-network model to a setting where interactions between a given group of actors are observed over time; the objective of the model is to identify changes in the latent parameters driving the structure of the network. This dynamic model is illustrated in Section 5 using first a simulation study and then a dataset of roll call voting patterns in the Senate. Finally, Section 6 presents a brief discussion of our models and future research directions.

2 Modeling a single social network

We introduce our notation by describing a model for a single social network. For the purpose of this paper, a network corresponds to a $n \times n$ matrix-variate random variable $\mathbf{Y} = [y_{ij}]$, where y_{ij} measures the strength of the relationship between actor i and actor j. The network is called undirected if $y_{ij} = y_{ji}$ and directed otherwise. Similarly, the network is called acyclical if subjects do not interact with themselves (in which case the diagonal of $\mathbf{Y}$ is typically taken to be structural zeros), and cyclical otherwise. In addition to the response matrix $\mathbf{Y}$, covariate vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ might be available for each of the actors in the network.

It is natural to assume that actors in a social network are (partially) exchangeable, which suggests a model where the entries of $\mathbf{Y}$ are conditionally independent given a $n \times n$ matrix of parameters $\mathbf{\Lambda} = [\lambda_{ij}]$. In the sequel we model $y_{ij}|\lambda_{ij}$ using a generalized linear model

$$y_{ij}|\lambda_{ij} \sim \Psi(y_{ij}|\lambda_{ij})$$

where $\Psi(\cdot|\lambda_{ij})$ is distribution in the exponential family with location parameter $\lambda_{ij} \in \mathcal{X}$. This allows for a rich class of models that can be used to explain different types of data. For example, if we only know whether the relationship is present of absent, then we could take $y_{ij} \in \{0, 1\}$ where $y_{ij} = 1$ if the relationship is present and $y_{ij} = 0$ otherwise. In that case, it is natural to assume that $y_{ij}|\lambda_{ij} \sim \mathsf{Ber}(\lambda_{ij})$ where $\lambda_{ij} = \mathsf{Pr}(y_{ij} = 1)$. In contrast, y_{ij} might corresponds to the number of times actor i and actor j interacted over a period of time, in which case $y_{ij} \in \mathbb{N}$ and it is natural to take $y_{ij}|\lambda_{ij} \sim \mathsf{Poi}(\lambda_{ij})$, where $\lambda_{ij} = \mathsf{E}(y_{ij})$.

In order to generate models that display realistic features, we link the parameter λ_{ij} with a linear predictor using an appropriate link function $g : \mathcal{X} \to \mathbb{R}$, so that

$$g(\lambda_{ij}) = \mu + \alpha_i + \alpha_j + \theta_{ij} + \mathbf{w}_{ij}\boldsymbol{\beta} \tag{1}$$

where μ is the global average response across all subjects, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is a vector of subject-specific random effects, $\mathbf{\Theta} = [\theta_{ij}]$ is an $n \times n$ matrix of random effects associated with the interactions between individuals, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_R)'$ is a vector of fixed effects associated with a covariate vector $\mathbf{w}_{ij} = (w_{ij1}, \ldots, w_{ijR})$ (which is constructed from $\mathbf{x}_i$ and $\mathbf{x}_j$). Note

that introducing a linear predictor that incorporates these four terms allow us to account for differential propensities, homophily by attributes, transitivity, and clustering. In particular, the additive terms α_i and α_j capture differential propensities, the interaction term θ_{ij} captures transitity in the network, and the regression term $\mathbf{w}_{ij}\boldsymbol{\beta}$ captures homophily by attributes. By suitably choosing the distribution for the random effects $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$ the model can also incorporate clustering (more on this below).

Framing network models in the context of generalized linear models is particularly appealing because it dramatically simplifies computation and allows us to adapt hierarchical specifications used in other contexts to answer questions related to the topology of the network.

2.1 Modeling the random effects

We start by discussing how to model differential propensities through the individual random effects. A natural approach to model subject-specific random effects in $\boldsymbol{\alpha}$ is to assume that they are independently sampled from a common distribution G; for example, we could take $\alpha_i \sim G = \mathsf{N}(0, \tau^2)$. This approach is appealing because it is easily interpretable; we can think of τ^2 as controlling the dispersion of individual propensities about the population average. In addition, conditional normality can usually can lead to simpler computational algorithms because it allows us to exploit conditional conjugacy to generate simple Gibbs sampling algorithms. Typically τ^2 will be unknown, so an additional prior needs to be incorporated into the model.

Alternatively, rather than assuming that G belongs to a given parametric family that depends only on a small number of parameters, we use the Dirichlet Process (DP) (Ferguson, 1973; Blackwell & MacQueen, 1973; Sethuraman, 1994) to provide a more flexible specification for the random-effect distribution. A random distribution G is said to follow a Dirichlet process with baseline probability measure G_0 and precision parameter $a > 0$, denoted $G \sim \mathsf{DP}(a, G_0)$, if it constructed as

$$G(\cdot) = \sum_{k=1}^{\infty} \omega_k \delta_{\alpha_k^*}(\cdot) \tag{2}$$

where $\delta_\alpha(\cdot)$ denotes the degenerate measure at α, $\alpha_1^*, \alpha_2^*, \ldots$ are independent and identically distributed samples from G_0 and ω_k are constructed from another independent and identically distributed sequence $u_1, u_2, \ldots$ where $u_k \sim \mathsf{Beta}(1, a)$ and $\omega_k = u_k \prod_{r<k}(1 - u_r)$. The joint distribution on the weights $\boldsymbol{\omega} = (\omega_1, \omega_2, ...)$ induced by this prior is called a stick-breaking prior with parameter a, denoted $\boldsymbol{\omega} \sim \mathsf{SB}(a)$.

Note that, for any measurable set A, we have $\mathsf{E}(G(A)) = G_0(A)$. Therefore, we can think of the baseline measure as providing a center for the nonparametric prior. This suggests setting $G_0 = \mathsf{N}(\mu, \tau^2)$, i.e., centering G around the simpler parametric model described at the beginning of this section. Similarly, $\mathsf{Var}(G(A)) = G_0(A)\{1-G_0(A)\}/(1+a)$, so a controls how close the samples from the DP are to G_0.

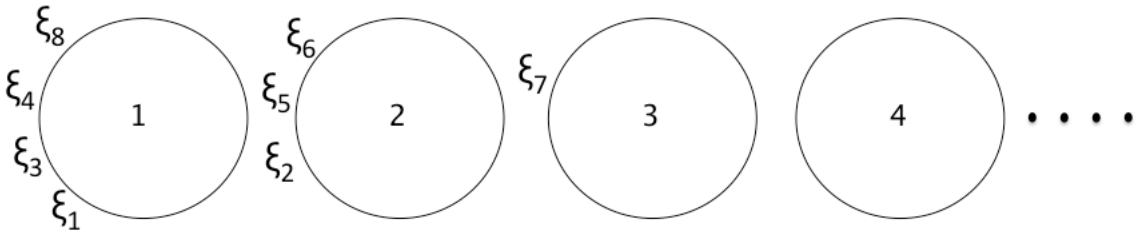


Figure 1: Graphical representation of the Chinese restaurant process, which describes the process by which observations are allocated to clusters under the Dirichlet process prior. In this analogy, customers sequentially sit at the tables of an infinite-capacity restaurant. The next customer sits at one of the currently occupied table with probability proportional to the number of customers in that table, and open a new table with probability proportional to a. In this case, customer 9 would sit table 1 with probability $4/(8+a)$, at table 2 with probability $3/(8+a)$, at table 4 with probability $1/(8+a)$ and at a new table with probability $a/(8+a)$.

The discrete nature of G implies ties among the random effects. Blackwell & MacQueen (1973) showed that the joint distribution of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ (after integrating out the unknown distribution G) can be written in terms of a collection of unique values $\alpha_1^*, \alpha_2^*, \ldots$ and indicators $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ by setting $\alpha_i = \alpha_{\xi_i}^*$. As in the constructive definition in (2), the unique values are independent and identically distributed with $\alpha_k^* \sim G_0$. On the other hand, the indicators in $\boldsymbol{\xi}$ follow a simple set of predictive rules satisfying $\xi_1 = 1$ and

$$\xi_i | \xi_{i-1}, \ldots, \xi_1 \sim \sum_{k=1}^{K^{i-1}} \frac{m_k^i}{a+i-1} \delta_k + \frac{a}{a+i-1} \delta_{K^{i-1}+1} \qquad i \geq 2, \tag{3}$$

where $K^{i-1} = \max\{\xi_1, \ldots, \xi_{i-1}\}$ is the number of distinct values among the first $i-1$ indicators, and $m_k^i = \sum_{j=1}^{i-1} \mathbf{1}(\xi_j = k)$ is the number of indicators taking the value k, for $k \leq K^{i-1}$.

This generative model for $\boldsymbol{\xi}$ is often called the Chinese Restaurant Process (Pitman, 1995, 1996), and is denoted $\boldsymbol{\xi} \sim \mathsf{CRP}(a)$. In this analogy, one imagines observations representing customers arriving to a Dim Sum restaurant with an infinite number of tables. The value of ξ_i indicates what table customer i is assigned to; the first customer sits in table 1, and subsequent customers sit at one of the occupied tables with probability proportional to the number of people already sitting in the table, or at a new table with probability proportional to c. Customers sitting at the k-th table "share" a common dish for dinner, which corresponds to α_k^*. A graphical representation of the CRP is presented in Figure 1.

A similar approach, first presented by Kemp et al. (2006) and Xu et al. (2006), can be used to generate a prior for the matrix of interactions $\boldsymbol{\Theta}$. As before we introduce a sequence of indicators $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n) \sim \mathsf{CRP}(b)$, where ζ_i is associated with actor i and $b > 0$ is a

precision parameter. To construct the interactions, we pair $\boldsymbol{\zeta}$ with a matrix of unique values $\boldsymbol{\Theta}^* = [\theta^*_{kl}]$, where $\theta^*_{kl} \sim H_0$ for some baseline measure H_0, and let $\theta_{ij} = \theta^*_{\zeta_i,\zeta_j}$. Note that, since the CRP produces an exchangeable sequence of indicators and the entries of $\boldsymbol{\Theta}^*$ are independently sampled from H_0, the resulting matrix $\boldsymbol{\Theta}$ is also invariant to the labeling of the actors.

Our specification of the random effects using almost-surely discrete distribution implies clustering among the actors in the network, both at the level of the α_is and the θ_{ij}s. However, the interpretation of both classes of cluster are subtly different. The indicators in $\boldsymbol{\xi}$ allow us to identify subjects with similar popularity levels, which is a measure that is specific to each single individual. On the other hand, the indicator $\boldsymbol{\zeta}$ cluster actors according to the way they interact with each other; hence, the clustering is based on the properties of pairs of subjects, making it natural to think of the clusters induced by $\boldsymbol{\zeta}$ as "factions" or "communities" within the network.

2.2 Modeling the fixed effects

Often, identifying features of the actors that affect the topology of the network is of key importance in social network analysis. Since the covariates enter the model through the regression term $\mathbf{w}_{ij}\boldsymbol{\beta}$, variable selection can be accomplished by carefully selecting a prior for $\boldsymbol{\beta}$.

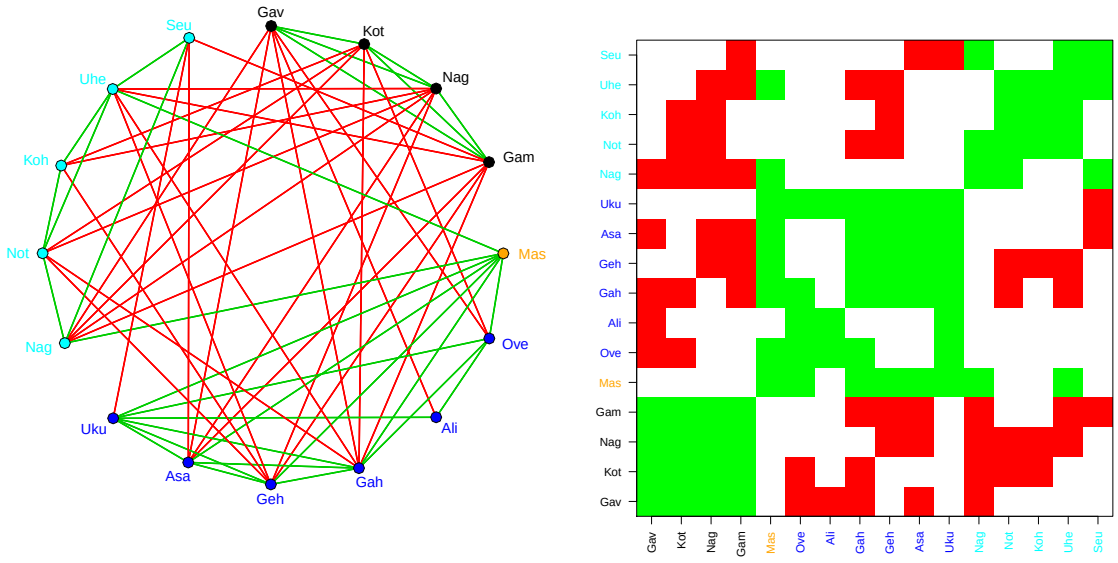
Mixture priors (George & McCulloch, 1997) are particularly appealing because they combine simplicity of implementation with interpretability. Covariates are assumed independently a priori from a zero-inflated normal distribution,

$$\beta_r|\varpi, \kappa^2 \sim \varpi \mathsf{N}(0, \kappa^2) + (1 - \varpi)\delta_0$$

The model can be more conveniently written in terms of indicators $\varsigma_1, \ldots, \varsigma_R$ such that, for $r = 1, \ldots, R$, $\varsigma_r = 0$ if $\beta_r = 0$ and $\varsigma_r = 1$ if $\beta_r \neq 0$. A priori, each variable is assumed to be present in the model independently of the others with probability ϖ, i.e., $\mathsf{Pr}(\varsigma_r) = \varpi$. In order to adjust for multiple comparisons, we model ϖ hierarchically by setting $\varpi \sim \mathsf{Beta}(c, d)$ (Scott & Berger, 2006). Alternatively to mixture priors, g-priors (Liang et al., 2008) or other shrinkage priors (Polson & Scott, 2010) could be used.

3 An illustration: Modeling a network of political alliances

To illustrate how our hierarchical Bayesian models can capture latent communities in social networks, we present an analysis of a network of political alliances and enmities among 16 Gahuku-Gama sub-tribes of Eastern Central Highlands of New Guinea, originally documented by Read (1954) and available on-line from `http://vlado.fmf.uni-lj.si/pub/networks/data/UciNet/UciData.htm`. The data, which corresponds to an undirected relationship, corresponds to a 16×16 symmetric matrix $\mathbf{Y} = [y_{ij}]$, where $y_{ij} = 1$ if the clans



(a) (b)

Figure 2: Raw network of political alliances among 16 Gahuku-Gama sub-tribes of Eastern Central Highlands of New Guinea. Panel (a) presents a circular plot where vertices correspond to tribes, green lines correspond to alliances among tribes, and red lines correspond to enmities. Panel (b) presents the same raw data in matrix form; again green cells correspond to alliances, red cells to enmities and white cells to neutrality. In both cases, the names of the tribes have been colored and reordered to reflect faction membership as elicited from our analysis.

are allied, $y_{ij} = -1$ if they are enemies, and $y_{ij} = 0$ if they are neutral. Analysis of similar datasets arises in intelligence tasks in countries like Iraq, Pakistan, and Afghanistan, where a large number of militias and armed groups are simultaneously active.

Figure 2(a) presents a circular plot for the network, where vertices correspond to tribes, green lines correspond to alliances among tribes, and red lines correspond to enmities. Figure 2(b) shows the same data in matrix form, where green cells correspond to alliances, red cells to enmities and white cells to neutrality. In both cases, the names of the tribes have been colored and reordered to reflect faction membership as elicited from our analysis. These figures strongly suggest the presence of three factions among the 16 tribes, with close alliances among member of the same faction and moderate to mild enmities among the factions.

Since the main goal of the analysis is identifying factions and covariates are not available for this dataset, we define the probabilities of the three possible outcomes using a simple continuations ratio probit model,

$$\mathsf{Pr}(y_{ij} = -1) = \Phi(\theta_{ij}), \qquad \mathsf{Pr}(y_{ij} = 0 | y_{ij} \neq -1) = \Phi(\phi_{ij}),$$

where Φ is the standard cumulative normal distribution, $\theta_{ij} = \theta^*_{\zeta_i,\zeta_j}$, $\phi_{ij} = \phi^*_{\zeta_i,\zeta_j}$, $\boldsymbol{\zeta} \sim \mathsf{CRP}(b)$,

$\theta_{kl} \sim \mathsf{N}(\varrho_1, \tau_1^2)$ and $\phi_{kl} \sim \mathsf{N}(\varrho_2, \tau_2^2)$. The model is completed with a hierarchical specification for the hyperparameters $b \sim \mathsf{Gam}(1,1)$, $\varrho_1 \sim \mathsf{N}(0,1)$, $\tau_1^2 \sim \mathsf{IGam}(2,1)$, $\varrho_2 \sim \mathsf{N}(0,1)$ and $\tau_2^2 \sim \mathsf{IGam}(2,1)$, where $\mathsf{IGam}(e,f)$ denotes the inverse-gamma distribution with e degrees of freedom and mean $f/(e-1)$.

We fitted the model using an iterative MCMC algorithm (see appendix A for details). All inferences reported below are based on 20,000 iterations obtained after a burn-in period of 5,000 iterations. Visual inspection of trace plots of enmity and alliance probabilities does not suggest any mixing problem.

Figure 3(a) presents the average posterior pairwise incidence matrix for this problem. The (i,j) entry of the matrix provides the posterior probability that tribes i and j belong to the same cluster. As expected, the plot suggests that there are three factions in this network; however, there is fair amount of uncertainty surrounding some aspects of the cluster structure. In particular, the first cluster (composed of four tribes, Gaveve, Kotuni, Nagamiza and Gama) is very well defined, but there is a small probability (around .25) that clusters two and three are really one cluster. Similarly, the Masilakidzuha tribe would seem to be somewhat of an outlier within cluster 2, and there is close to a 0.5 probability that this tribe belongs to a cluster of its own.

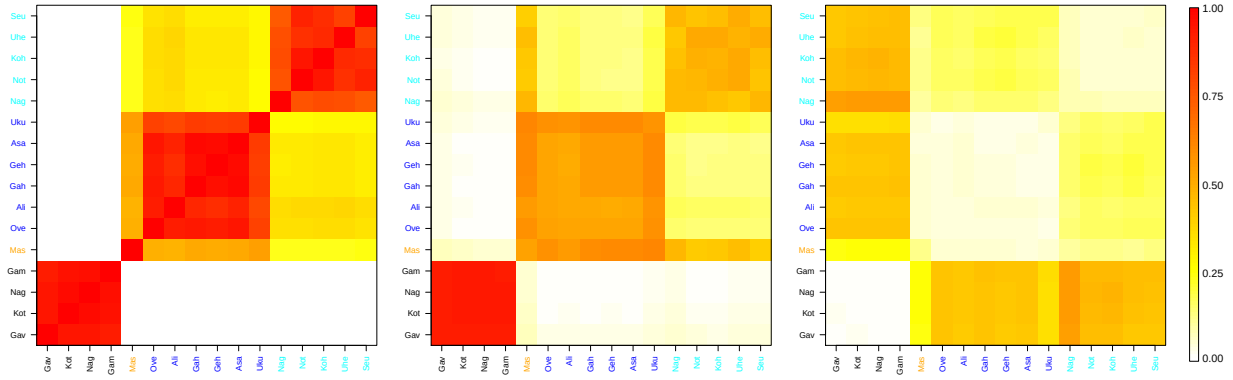
In addition to the pairwise clustering structure, we present in Figure 3(b) the posterior mean for the probability that any two tribes are allies, $\mathsf{E}\{[1-\Phi(\phi_{ij})][1-\Phi(\theta_{ij})]|\mathbf{Y}\}$, while in Figure 3(c) shows the posterior means for the probability that any two tribes are enemies, $\mathsf{E}\{\Phi(\theta_{ij})|\mathbf{Y}\}$. Note that his estimates average over the uncertainty in the number of clusters, providing a more accurate estimate of the interactions among actors. As expected, the probability of alliances within the same faction are very high (specially among members of the first faction), while the probability of alliances across factions tends to be low. On the other hand, there is a moderate probability of enmities between members of faction 1 and the members of the other two factions, but the members of the second factions tend to be mostly neutral towards the members of the third faction.

4 Identifying structural changes in social networks

Many interesting applications of network models involve systems where the structure of the network evolves in time. This section extends the hierarchical blockmodel of Section 2 to model a sequence of social networks observed over time. We account for time dependence by introducing a hidden Markov model for the parameters of the network, which allows us to identify points in time when structural changes seem to have occured.

In the sequel, we assume that the interactions among a group of n actors are observed over T consecutive time intervals. Hence, the data consists of a sequence $\mathbf{Y}_1, \ldots \mathbf{Y}_T$ of $n \times n$ matrices where $\mathbf{Y}_t = [y_{ijt}]$ and y_{ijt} corresponds to the strength of the relationship between actors i and j at time t. As in Section 2, we model y_{ijt} conditionally on a linear predictor λ_{ijt} using an exponential family likelihood,

$$y_{ijt}|\lambda_{ijt} \sim \Psi(y_{ijt}|\lambda_{ijt})$$



(a) Average pairwise incidence matrix. (b) Probability of an alliance. (c) Probability of an enmity.

Figure 3: Summary of the posterior distribution of model parameters for the tribes dataset. Panel (a) corresponds to the average pairwise incidence matrix, where entry (i, j) of the matrix provides the posterior probability that tribes i and j belong to the same cluster. Red squares correspond to clearly defined factions in the network. Panels (b) and (c) provide posterior means for the posterior probability that any two tribes are allis or enemies.

where the linear predictor λ_{ijt} is connected through a link function g to state-specific parameters,

$$g(\lambda_{ijt}) = \mu_{\gamma_t} + \alpha_{i,\gamma_t} + \alpha_{j,\gamma_t} + \theta_{i,j,\gamma_t} + \mathbf{w}_{ijt}\boldsymbol{\beta}_{\gamma_t}$$

where $\gamma_t \in \{1, 2, ...\}$ is a state indictor associated with time t, μ_l is the mean response when the system is in state l, α_{il} corresponds to the propensity of individual i in state l, θ_{ijl} captures transitivity between subjects i and j in state l, and $\boldsymbol{\beta}_l$ captures the effects of the actor's attributes on the network in state l.

The state of the system is assumed to evolve according to a discrete-time Markovian processes where

$$\gamma_t|\gamma_{t-1} \sim \mathsf{Multinomial}(\boldsymbol{\pi}_{\gamma_{t-1}}) \qquad \boldsymbol{\pi}_l|\boldsymbol{\eta} \sim \mathsf{DP}(d, \boldsymbol{\eta}) \qquad \boldsymbol{\eta} \sim \mathsf{SB}(e)$$

and $\boldsymbol{\pi}_l = (\pi_{l1}, \pi_{l2}, \ldots)$ represents the vector of transition out of state l. This specification implies that $\mathsf{E}(\boldsymbol{\pi}_l|\boldsymbol{\eta}) = \boldsymbol{\eta}$, hence $\boldsymbol{\eta}$ corresponds to the average transition probabilities out of a state. The precision parameter d controls the amount of variability of each $\boldsymbol{\pi}_l$ around $\boldsymbol{\eta}$, while e controls the way mass is spread over the entries of $\boldsymbol{\eta}$. Using a hierarchy of stick-breaking priors to generate infinite-dimensional hidden Markov models was first proposed by (Beal et al., 2001) and (Teh et al., 2006) and allows us to treat the number of states in the system as a random variable that needs to be estimated from the data.

Given the states, networks are assumed to be independent from each other and the parameters associated with each state constructed by following essentially the same recipe

we provided in Section 2. In particular, for each state $l = 1, 2, \ldots$ we set $\alpha_{i,l} = \alpha^*_{\xi_{i,l},l}$ where $\alpha^*_{kl} \sim G_{l0}$ independently and $\boldsymbol{\xi}_l = (\xi_{1,l}, \ldots, \xi_{n,l}) \sim \mathsf{CRP}(a_l)$ is a set of unique indicators for the individuals at state l. Similarly, we construct the interactions by letting $\theta_{i,j,l} = \theta^*_{\zeta_{il},\zeta_{jl},l}$, where $\boldsymbol{\zeta}_l = (\zeta_{1,l}, \ldots, \zeta_{n,l}) \sim \mathsf{CRP}(b_l)$ is another set of unique indicators for the individuals at state l. Finally, for the regression coefficients we let $\beta_{rl} \sim \varpi_l \mathsf{N}(0, \kappa^2) + (1 - \varpi_l)\delta_0$ with $\varpi_l \sim \mathsf{Beta}(c, d)$ independently for each $l = 1, 2, \ldots$.

5 Illustrations

5.1 A simulation study

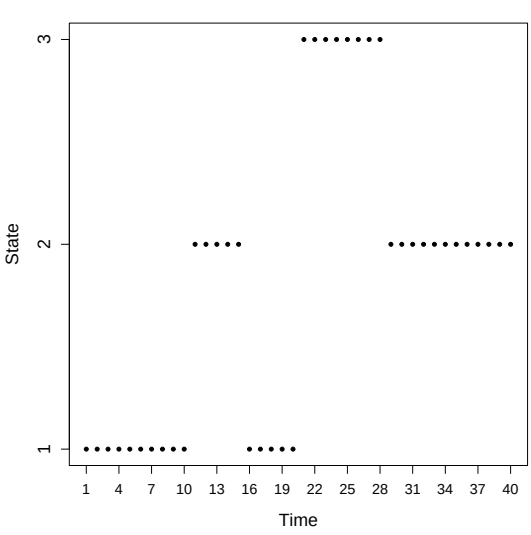
We first test our dynamic model for networks using a simulated dataset consisting of dyadic relationships among $n = 20$ actors observed over $T = 40$ time points. There are three possible states for the system. From $t = 1$ to $t = 10$ the network is in state 1, from $t = 11$ to $t = 15$ the system is in state 2, from $t = 16$ to $t = 20$ the system returns to state 1, from $t = 21$ to $t = 28$ the system moves to state 3, and from $t = 29$ to $t = 40$ the system moves back to state 2 (see Figure 4(a)). In state 1 individuals group into two factions consisting of actors 1 to 10 and 11 to 20 respectively; in state 2 we observe three factions consisting of subjects 1 to 6, 7 to 12 and 13 to 20; in state 3 we have four factions comprising of subjects 1 to 8, 9 to 13, 14 to 17 and 18 to 20. Figures 4(b), 4(c) and 4(d) show the true faction membership and the true interaction probability among subjects on each of these three states.

We model the data using a simplified version of the model described in Section 4. In particular, we let $y_{i,j,t} \sim \mathsf{Ber}(\Phi(\theta_{i,j,\gamma_t}))$ where $\theta_{i,j,l} = \theta^*_{\zeta_{i,l},\zeta_{j,l},l}$. In addition, we let $\theta^*_{k,k',l} \sim \mathsf{N}(\varrho_l, \tau_l^2)$ and $\boldsymbol{\zeta}_l \sim \mathsf{CRP}(b_l)$ independently for each state. The hyperparameters are also given independent priors $\varrho_l \sim \mathsf{N}(0, 1)$, $\tau_l^2 \sim \mathsf{IGam}(2, 1)$, $b_l \sim \mathsf{Gam}(1, 1)$. Finally, as discussed in Section 4, we have

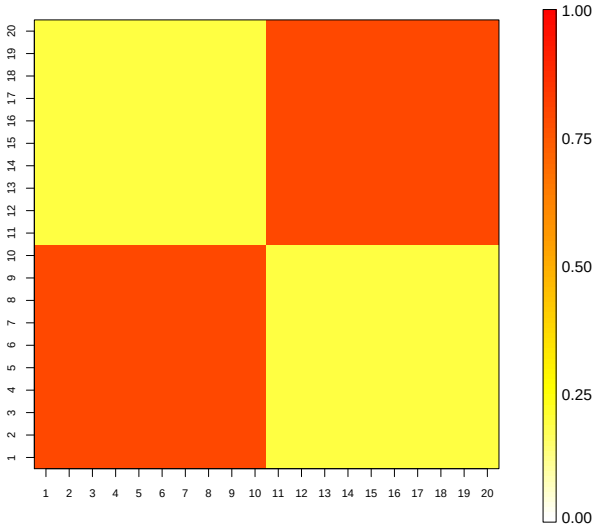
$$\gamma_t | \gamma_{t-1} \sim \mathsf{Multinomial}(\boldsymbol{\pi}_{\gamma_{t-1}}) \qquad \boldsymbol{\pi}_l | \boldsymbol{\eta} \sim \mathsf{DP}(d, \boldsymbol{\eta}) \qquad \boldsymbol{\eta} \sim \mathsf{SB}(e)$$

and we place priors on the parameters controlling the structure of the inifinite hidden Markov model, with $d \sim \mathsf{Gam}(1, 1)$ and $e \sim \mathsf{Gam}(1, 1)$. A graphical representation of the model, highlighting how the hidden Markov model relates to the CRP priors on faction membership can be seen in Figure 5.1.

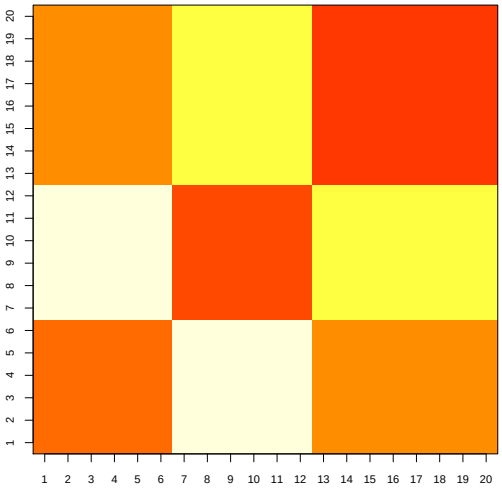
We fitted the model using an iterative MCMC algorithm; a full description can be seen in appendix B. All inferences reported below are based on 60,000 iterations obtained after a burn-in period of 10,000 iterations. The algorithms were implemented using `R`; one full run takes about 25 hours on a MacBook Pro laptop. To initialize the chain, we assign each network to a separate state and run the rest of the MCMC algorithm using this fixed state for a preliminary burn-in period of 3,000 iterations. This provides reasonable estimates of the faction structure for each network that can be used as an initial state for a final burn-in period consisting of 7,000 iterations in which the full algorithm is run. Visual inspection of trace plots for the interaction probabilities do not suggest any major problem with mixing;



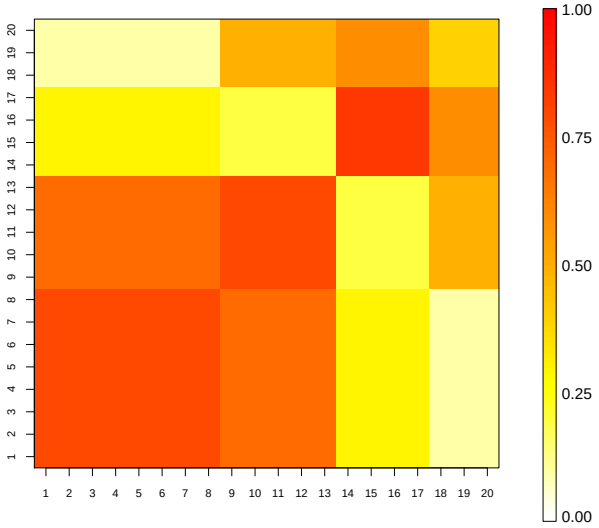
(a) True states



(b) Interaction probability for state 1



(c) Interaction probability for state 2



(d) Interaction probability for state 3

Figure 4: True states and interaction probabilities for the simulation study. The system is assumed to consist of $n = 40$ subjects observed over $T = 40$ periods. There are three states consisting of two, three and four factions respectively.

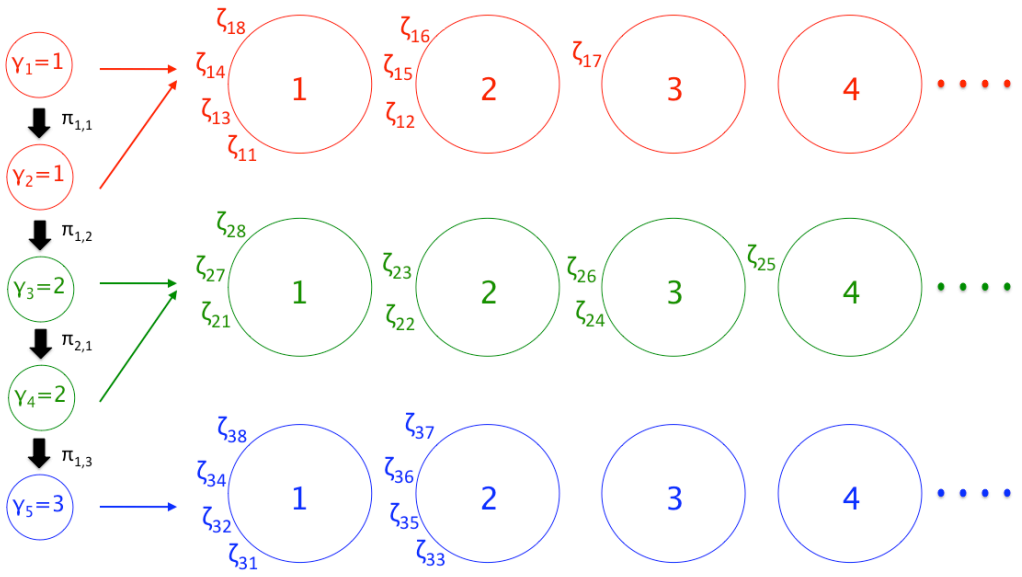
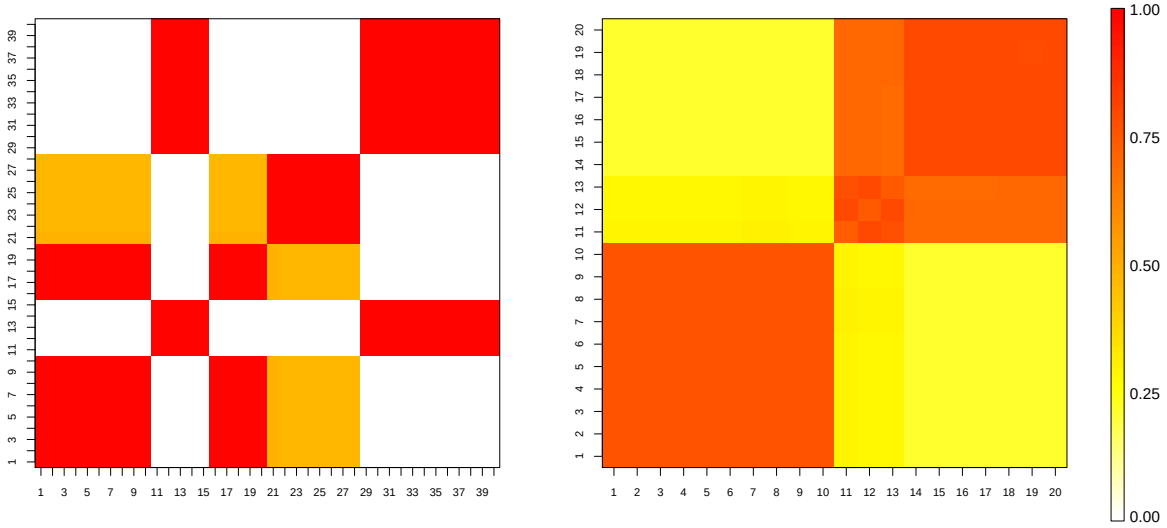


Figure 5: Graphical representation of the clustering mechanism associated with our dynamic relational model. Each state has associated with it a different CRP process, potentially controlled by a different parameter b_l. Each of these CRPs provides a prior on faction membership that is specific to each state.

however, trace plots for the number of states in the system suggests that mixing on $\boldsymbol{\zeta}$ is relatively slow.

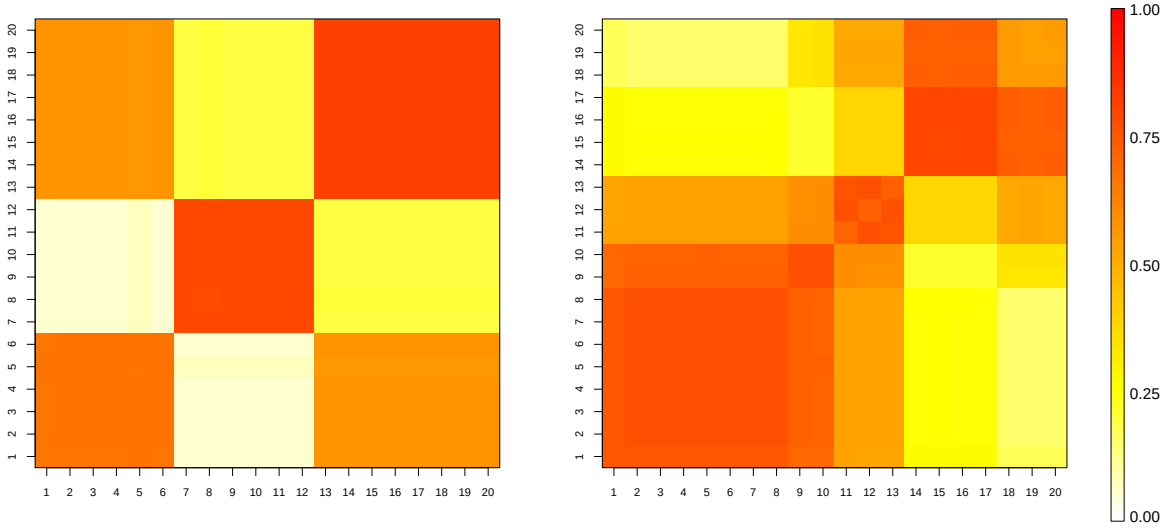
Posterior estimates for a representative simulated dataset are presented in Figures 6(a) to 6(d). In particular, Figure 6(a) presents posterior means for the probabilities that any two networks (corresponding to two different time points) belongs to the same state. Note that networks generated in state 1 are clearly separated from networks generated in state 2. Similarly networks generated in state 2 are never assigned to the same group with networks generated under state 3. However, the model finds it difficult to separate states 1 and 3; the networks observed at times 16 to 20 have between 0.45 and 0.48 probability of being clustered with at least one network generated at state 1. More generally, our algorithm visits models containing between 2 and 5 states (posterior probabilities are approximately 0.45, 0.53, 0.01, and < 0.01). The difficulties in identifying the third state are probably due to the fact that the number of actors associated with each of the four factions is very small for these networks, which translates into a high degree of uncertainty in the probability of interaction among factions.

Figures 6(b), 6(c) and 6(d) present posterior means for the probability that any two subjects interact for three of the networks (one associated with each of the original states). Note that, in spite of the uncertainty associated with the existence of state 3, the estimates for the probability of association between subjects for specific networks are relatively close to the true values. In particular, the estimates of interaction between subjects at time 1



(a) Pairwise probabilities for states

(b) Interaction probability at time 1



(c) Interaction probability at time 12

(d) Interaction probability at time 24

Figure 6: Posteriors estimates for a representative Monte Carlo simulation. Panel (a) shows the mean posterior pairwise clustering probabilities for every pair of states. Panels (b) to (d) show the pairwise interaction probability associated with three of the networks, one associated with each of the original states.

suffer very little due to the misclassification of networks associated with state 3. The main discrepancies arise in the estimation of the interaction probabilities in state 3, particularly in the estimation of the probability of interaction among subjects 18 to 20.

5.2 Roll call votes in the 109^{th} Senate

As a second illustration we consider a dataset consisting of roll call votes in the first Session of the 109^{th} U.S. Senate, which served during the year 2005; the data is available at `http://thomas.loc.gov/home/rollcallvotes.html`. We focus on a small subset of the votes consisting of 15 Senate confirmations, which include the confirmation of a number of high level officials (including Condoleeza Rice as Secretary of State on January 6 and of Alberto R. Gonzales as Attorney General on February 3), as well as the confirmation of a number of U.S. District Court Judge Judges. Since confirmations are influenced by the dynamically changing general political climate in the nation, we could expect some sort of temporal dependence in the votes.

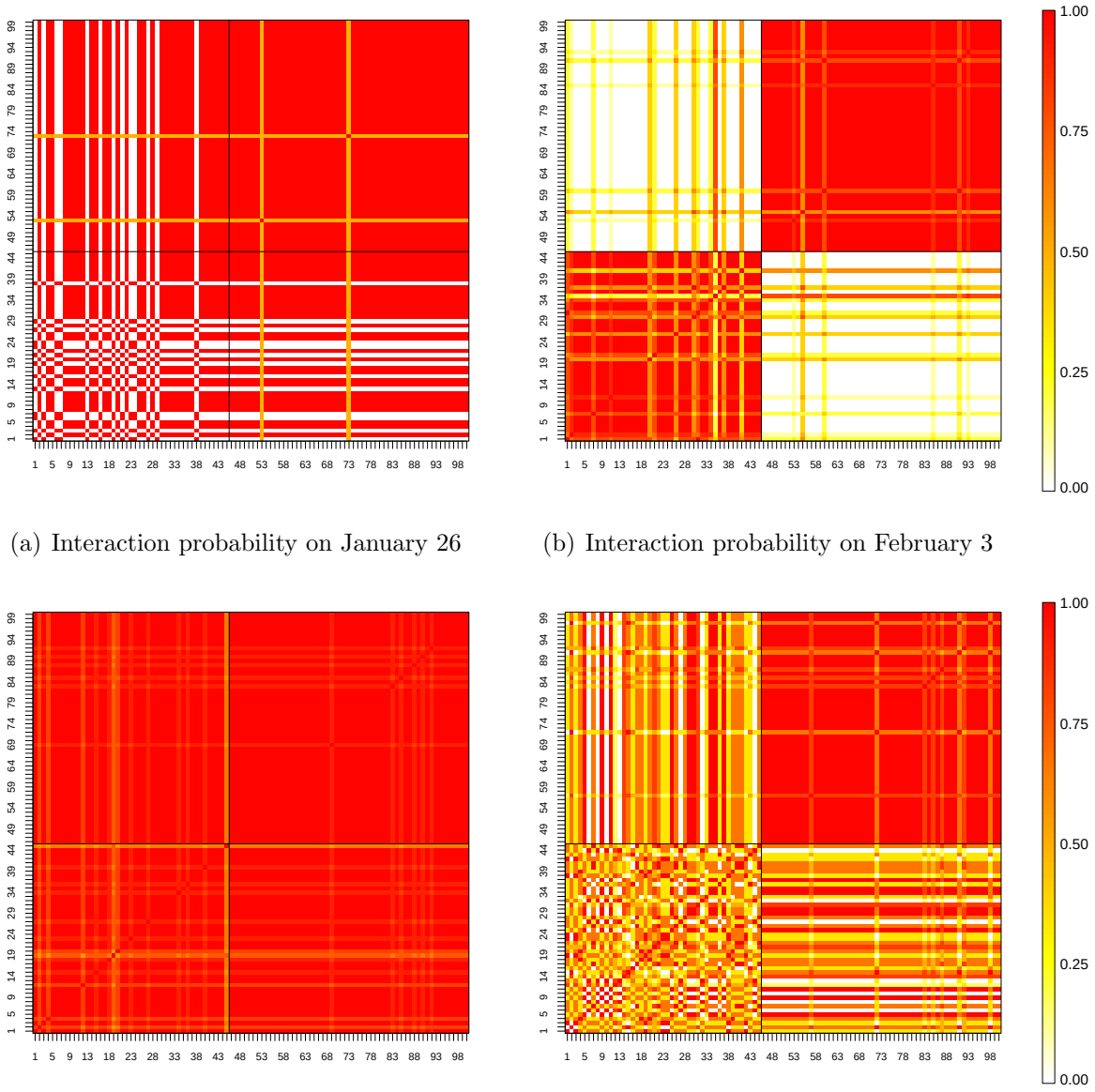
To construct the networks, two senators are said to interact if their vote agreed for a given confirmation. In this illustration we ignore party affiliation and fit the same dynamic model described in Section 5.1, which allows us to identify factions in the networks. We expect that one of the states of the network will correspond to votes along partisan lines (hence, factions will mirror party affiliation), and we are interested in identifying alternative configurations. Note that, since Senators might be absent from the vote on any given date, the dataset contains a number of missing values.

Again, we fit the model using the MCMC algorithm described in Appendix B. We assume that missing data is missing completely at random and augment the sampler by imputing the missing values from the likelihood function at each iteration. As in our simulation study we report posterior estimates based on 60,000 iterations obtained after a burn-in period of 10,000 iterations.

Our algorithm identifies four states for the network; Figures 7(a) to 7(d) presents the probability of pairwise interaction associated with them. States 2 and 3 are easy to interpret; state 2 (for which February 3 is a representative date) corresponds to a vote along partisan lines, while state 3 (for which April 11 is a good example) corresponds to a unanimous vote. On the other hand, states 1 (corresponding to January 26) and state 4 (July 18) correspond to different alignments of dissenting Democrat senators with the Republican majority. However, the features of both networks are very different; state 1 is a quasi-unanimous vote with a small number of dissenting Democrats, while in state 4 a number of democratic senators are unreliable swing voters.

6 Conclusions and future work

We have discussed probabilistic models for dynamic social networks that allow us to identify structural changes in the structure of the network. The models are extremely general and



(a) Interaction probability on January 26

(b) Interaction probability on February 3

(c) Interaction probability on April 11

(d) Interaction probability on July 18

Figure 7: Posteriors means for the interaction probabilities for four time points, each one associated with a different state in the system. Vertical lines mark partisan divisions, with senators Democrat senators being placed on the left/bottom area and Republican senators being on the right/top section of the plot. Senator Jeffords (I-VT) was considered Democrat for the purpose of this classification.

can account for differential attachment, homiphily by attributes, transitivity and clustering in the network as well as for binary, count and continuous data.

In order to explore the posterior distribution, we have introduced a Markov chain Monte Carlo algorithm based on the CRP representation of the process. Although the algorithms seem to perform reasonable well in our examples, there is plenty of room for improvement. In general, two important issues arise from our experiments. First, it is not clear that the algorithms will be able to mix well in more complex datasets with a substantially larger number of actors. To address this issue we plan to explore in the future alternatives such as slice samplers Walker (2007) and split-mergue algorithms Dahl (2003). The second issue is speed, specially in on-line problems where the data is collected sequentially and inferences must be performed after each new data point is observed. In this setting, sequential Monte Carlo algorithms Cappe et al. (2007); Carvalho et al. (2009); Rodriguez (2010) are a natural alternative that can provide fast computation.

Our models utilize nonparametric mixture priors to identify factions and model differential attachment. In particular, we use constructions based on the Dirichlet process priors and the Chinese restaurant process. However, the CRP is a restrictive allocation mechanism; in particular, it implies the prior assumption that the number of factions increases logarithmically with the number of actors in the system. In the future we plan to explore more general allocation mechanisms such as the Pitman-Yor process (Pitman, 1995), which allows for power-law rates of growth in the number of factions while preserving all the computational advantages associated with the CRP.

The dynamic models we describe in this paper are appealing because of their interpretability. States can be thought of as stable periods when the system is in equilibrium; variability within the state is assumed to be due to the uncertainty associated with the stochastic nature of the relationship and not due to any structural change in the system. However, the model implies that states are (conditionally) independent among themselves, an assumption that might not be supported by the data. Also, it is not clear that social systems are ever "in equilibrium", making hidden Markov models and other change-point models unrealistic. In the future we plan to explore more general mechanisms to introduce dynamic structure into network models that do not rely on discrete state space models.

A Computation in the single-network model using Markov chain Monte Carlo algorithms

To develop our MCMC sampler, introduce two sets of latent variables, $\{z_{ij1}\}$ and $\{z_{ij2}\}$, such that $z_{ij1} \sim \mathsf{N}(\theta_{ij}, 1)$ and $z_{ij2} \sim \mathsf{N}(\phi_{ij}, 1)$. In addition, let

$$y_{ij}|z_{ij1}, z_{ij2} = \begin{cases} -1 & z_{ij1} \geq 0 \\ 0 & z_{ij1} < 0, z_{ij2} \geq 0 \\ 1 & z_{ij1} < 0, z_{ij2} < 0 \end{cases}$$

Integrating z_{ij1} and z_{ij2} leads to $\mathsf{Pr}(y_{ij} = -1) = \Phi(\theta_{ij})$, $\mathsf{Pr}(y_{ij} = 0) = \Phi(\phi_{ij})\{1 - \Phi(\theta_{ij})\}$ and $\mathsf{Pr}(y_{ij} = 1) = \{1 - \Phi(\phi_{ij})\}\{1 - \Phi(\theta_{ij})\}$ as desired. Then, the joint posterior distribution is given by

$$p(\mathbf{Z}_1, \mathbf{Z}_2, \boldsymbol{\zeta}, \boldsymbol{\Theta}^*, \boldsymbol{\Phi}^*, c, \varrho_1, \varrho_2, \tau_1, \tau_2) \propto \left\{ \prod_{i=1}^{n} \prod_{j=i+1}^{n} p(y_{ij}|z_{ij1}, z_{ij2}) p(z_{ij1}|\theta^*_{\zeta_i,\zeta_j}) p(z_{ij2}|\phi^*_{\zeta_i,\zeta_j}) \right\}$$
$$\times \left\{ \prod_{k=1}^{\max_{i\leq n}\{\zeta_i\}} \prod_{l=k}^{\max_{i\leq n}\{\zeta_i\}} p(\theta^*_{lk}) p(\phi^*_{lk}) \right\} p(\boldsymbol{\zeta}) p(c) p(\varrho_1) p(\varrho_2) p(\tau_1) p(\tau_2)$$

where $\mathbf{Z}_k = [z_{ijk}]$ and $p(\boldsymbol{\zeta})$ is given by (3). Since the entries of $\boldsymbol{\zeta}$ are exchangeable a priori, the prior full conditional distribution for the faction indicator ζ_i is given by

$$\zeta_i|\boldsymbol{\zeta}^{-i}, \cdots \sim \sum_{k=1}^{K^{-i}} \frac{m_k^{-i}}{b + n - 1} \delta_k + \frac{b}{b + n - 1} \delta_{K^{-i}+1}$$

where $L^{-i} = \max_{j\neq i}\{\zeta_j\}$ is the current number of factions when observation i has been removed from the sample and $m_k^{-i} = \sum_{j\neq i} \mathbf{1}_{\{\zeta_j = k\}}$ is the number of components in faction k, possibly excluding observation i. The posterior full conditional is therefore given by:

$$\mathsf{Pr}(\zeta_i = k| \cdots) \propto$$
$$\begin{cases} m_k^{-i} \prod_{u=1}^{2} \prod_{k'=1}^{K^{-i}} \frac{\Upsilon\left(m_{k'}^{-i} + r_{k,k'}^{-i}, \sum_{(j,j')\in\Omega^i_{k'}\cup\Omega^{-i}_{k,k'}} z_{j,j',u}, \sum_{(j,j')\in\Omega^i_{k'}\cup\Omega^{-i}_{k,k'}} z^2_{j,j',u}, \varrho_u, \tau_u^2\right)}{\Upsilon\left(r_{k,k'}^{-i}, \sum_{(j,j')\in\Omega^{-i}_{k,k'}} z_{j,j',u}, \sum_{(j,j')\in\Omega^{-i}_{k,k'}} z^2_{j,j',u}, \varrho_u, \tau_u^2\right)} & k \leq K^{-i} \\ b \prod_{u=1}^{2} \prod_{k'=1}^{K^{-i}} \Upsilon\left(m_{k'}^{-i}, \sum_{(j,j')\in\Omega^i_{k'}} z_{j,j',u}, \sum_{(j,j')\in\Omega^i_{k'}} z^2_{j,j',u}, \varrho_u, \tau_u^2\right) & k = K^{-i} + 1 \end{cases}$$

where $\Omega^i_{k'} = \{(j, j') : \zeta_{j'} = k', j = i, j' \neq i\}$, $\Omega^{-i}_{k,k'} = \{(j, j') : \zeta_j = k, \zeta_{j'} = k', j \neq i, j' \neq i\}$, $r^{-i}_{k,k'} = \sum_{(j,j')\in\Omega^{-i}_{k,k'}} 1$ and

$$\Upsilon(n, a, b, \varrho, \tau^2) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{1 + n\tau^2}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(b + \frac{\varrho}{\tau^2} - \frac{(a + \varrho/\tau^2)^2}{n + 1/\tau^2}\right)\right\}$$

The full conditional for the interaction terms are given by

$$\theta^*_{k,k'}| \cdots \sim \mathsf{N}\left(\left\{r_{k,k'} + \frac{1}{\tau_1^2}\right\}^{-1} \left\{\sum_{(i,j)\in\Omega_{k,k'}} z_{i,j,1} + \frac{\varrho_1}{\tau_1^2}\right\}, \left\{r_{k,k'} + \frac{1}{\tau_1^2}\right\}^{-1}\right)$$

and

$$\phi^*_{k,k'}| \cdots \sim \mathsf{N}\left(\left\{r_{k,k'} + \frac{1}{\tau_2^2}\right\}^{-1} \left\{\sum_{(i,j)\in\Omega_{k,k'}} z_{i,j,2} + \frac{\varrho_2}{\tau_2^2}\right\}, \left\{r_{k,k'} + \frac{1}{\tau_2^2}\right\}^{-1}\right).$$

where $\Omega_{k,k'} = \{(j,j') : \zeta_j = k, \zeta_{j'} = k', j' > j\}$, $r_{k,k'} = \sum_{(j,j')\in\Omega_{k,k'}} 1$.

The baseline parameters are sampled as

$$\varrho_1|\cdots \sim \mathsf{N}\left(\left\{\frac{K(K+1)}{2}+\tau_1^2\right\}^{-1}\left\{\sum_{k=1}^{K}\sum_{k'=k}^{K}\theta^*_{k,k'}\right\}, \left\{\frac{K(K+1)}{2\tau_1^2}+1\right\}^{-1}\right)$$
$$\tau_1|\cdots \sim \mathsf{IGam}\left(2+\frac{K(K+1)}{4}, 1+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=k}^{K}(\theta^*_{k,k'}-\varrho_1)^2\right)$$

and similarly

$$\varrho_2|\cdots \sim \mathsf{N}\left(\left\{\frac{K(K+1)}{2}+\tau_2^2\right\}^{-1}\left\{\sum_{k=1}^{K}\sum_{k'=k}^{K}\phi^*_{k,k'}\right\}, \left\{\frac{K(K+1)}{2\tau_2^2}+1\right\}^{-1}\right)$$
$$\tau_2|\cdots \sim \mathsf{IGam}\left(2+\frac{K(K+1)}{4}, 1+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=k}^{K}(\phi^*_{k,k'}-\varrho_2)^2\right)$$

Following Escobar & West (1995), the concentration parameter in the CRP is sampled by introducing another indicator variable χ such that

$$\chi|b,\cdots \sim \mathsf{Beta}(b+1,n)$$
$$b|\chi,\cdots \sim \frac{K}{n(1-\log\{\chi\})+K}\mathsf{Gam}(K+1,1-\log\{\chi\}) + \frac{n(1-\log\{\chi\})}{n(1-\log\{\chi\})+K}\mathsf{Gam}(K,1-\log\{\chi\})$$

Finally, the full conditional distributions for the latent variables in $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent from each other and given by

$$z_{i,j,1}|\cdots \sim \begin{cases}\mathsf{N}(\theta^*_{\zeta_i,\zeta_j},1)\mathbf{1}_{[0,\infty)} & y_{ij}=-1\\ \mathsf{N}(\theta^*_{\zeta_i,\zeta_j},1)\mathbf{1}_{(-\infty,0)} & y_{ij}\neq -1\end{cases}$$

and

$$z_{i,j,2}|\cdots \sim \begin{cases}\mathsf{N}(\phi^*_{\zeta_i,\zeta_j},1) & y_{ij}=-1\\ \mathsf{N}(\phi^*_{\zeta_i,\zeta_j},1)\mathbf{1}_{[0,\infty)} & y_{ij}=0\\ \mathsf{N}(\phi^*_{\zeta_i,\zeta_j},1)\mathbf{1}_{(-\infty,0)} & y_{ij}\neq 1\end{cases}$$

where $\mathsf{N}(\varrho,\tau^2)\mathbf{1}_A$ denotes the normal distribution truncated to the set A.

B Computation in the dynamic-network model using Markov chain Monte Carlo algorithms

As before, we introduce auxiliary variables $z_{ijt} \sim \mathsf{N}(\theta_{ijt}, 1)$ such that

$$y_{ijt}|z_{ijt} \sim \begin{cases} 1 & z_{ijt} \geq 0 \\ 0 & z_{ijt} < 0 \end{cases}$$

so that the posterior distribution is given by:

$$p(\mathbf{Z}_1, \ldots, \mathbf{Z}_T, \{\mathbf{\Theta}_l^*\}, \{\boldsymbol{\zeta}_l\}, \{b_l\}, \boldsymbol{\gamma}, \{\varrho_l\}, \{\tau_l^2\}, \boldsymbol{\eta}, d, e, |\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$$

As argued in Teh et al. (2006) and van Gael et al. (2008), the full conditional distribution for γ_t is given by

$$\mathsf{Pr}(\gamma_t = l|\cdots) \propto \begin{cases} \left(s^{-t}_{\gamma_{t-1},l} + d\eta_l\right) \frac{s^{-t}_{l,\gamma_{t+1}} + d\eta_{\gamma_{t+1}}}{s^{-t}_{l,\cdot} + d} \Delta_l^{-t} & l \leq L^{-t}, l \neq \gamma_{t-1}, \\ \left(s^{-t}_{\gamma_{t-1},l} + d\eta_l\right) \frac{s^{-t}_{l,\gamma_{t+1}} + d\eta_{\gamma_{t+1}} + 1}{s^{-t}_{l,\cdot} + d + 1} \Delta_l^{-t} & l = \gamma_{t-1} = \gamma_{t+1}, \\ \left(s^{-t}_{\gamma_{t-1},l} + d\eta_l\right) \frac{s^{-t}_{l,\gamma_{t+1}} + d\eta_{\gamma_{t+1}}}{s^{-t}_{l,\cdot} + d + 1} \Delta_l^{-t} & l = \gamma_{t-1} \neq \gamma_{t+1}, \\ d\eta_l \eta_{\gamma_{t+1}} \Delta_l^{-t} & l = L^{-t} + 1. \end{cases}$$

where

$$\Delta_l^{-t} = \prod_{k=1}^{K_l} \prod_{k'=k}^{K_l} \frac{\Upsilon\left(w^t_{k,k',l} + w^{-t}_{k,k',l}, \sum_{(j,j',t') \in \Lambda^t_{k,k',l} \cup \Lambda^{-t}_{k,k',l}} z_{j,j',t'}, \sum_{(j,j',t') \in \Lambda^t_{k,k',l} \cup \Lambda^{-t}_{k,k',l}} z^2_{j,j',t'}, \varrho_l, \tau_l^2\right)}{\Upsilon\left(w^{-t}_{k,k',l}, \sum_{(j,j',t') \in \Lambda^{-t}_{k,k',l}} z_{j,j',t'}, \sum_{(j,j',t') \in \Lambda^{-t}_{k,k',l}} z^2_{j,j',t'}, \varrho_l, \tau_l^2\right)}$$

for $l \leq L^{-t}$ and

$$\Delta_l^{-t} = \prod_{k=1}^{K_l} \prod_{k'=k}^{K_l} \Upsilon\left(w^t_{k,k',l}, \sum_{(j,j',t') \in \Lambda^t_{k,k',l}} z_{j,j',t'}, \sum_{(j,j',t') \in \Lambda^t_{k,k',l}} z^2_{j,j',t'}, \varrho_l, \tau_l^2\right)$$

for $l = L^{-t} + 1$. The vector of indicators $\boldsymbol{\zeta}_{L^{-t}+1}$ is generated by first sampling $\varrho_{L^{-t}+1} \sim \mathsf{N}(0,1)$, $\tau_l^2 \sim \mathsf{IGam}(2,1)$ and $b_l \sim \mathsf{Gam}(1,1)$, setting $\zeta_{1,l} = 1$ and sequentially sampling $\zeta_{i,l}|b_l, \zeta_{i-1,l} \ldots, \zeta_{1,l}$ following the CRP prior in (3).

In the previous expression, $\Lambda^t_{k,k',l} = \{(j,j',t') : t' = t, \zeta_{j,l} = k, \zeta_{j',l} = k'\}$, $\Lambda^{-t}_{k,k',l} = \{(j,j',t') : \gamma_t = l, t' \neq t, \zeta_{j,l} = k, \zeta_{j',l} = k'\}$, $w^t_{k,k',l} = \sum_{(j,j',t') \in \Lambda^t_{k,k',l}} 1$, $w^{-t}_{k,k',l} = \sum_{(j,j',t') \in \Lambda^{-t}_{k,k',l}} 1$, $s^{-t}_{l,l'}$ denotes the number of transitions from state l to state l' excluding those involving the γ_t, and L^{-t} is the number of states excluding the t-th time point. If a new empty state needs to be created, we increase the number of states by one and the vector $\boldsymbol{\eta}$ is updated

by exploiting its stick-breaking structure and setting $\eta_{L+1}^{new} = v\gamma_{L+1}$, $\eta_{L+2}^{new} = (1-v)\gamma_{L+1}$ were $v \sim \mathsf{Beta}(\alpha_0, 1)$. Also,

$$\Upsilon(n, a, b, \varrho, \tau^2) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{1+n\tau^2}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(b + \frac{\varrho}{\tau^2} - \frac{(a+\varrho/\tau^2)^2}{n+1/\tau^2}\right)\right\}$$

The vector $\boldsymbol{\eta}$ can be resampled by introducing independent auxiliary variables $\{q_{l,l'}\}$ for $l, l' \in \{1, \ldots, L\}$ such that

$$\mathsf{Pr}(q_{l,l'} = q) \propto S(s_{l,l'}, q)(d\gamma_{l'})^q, \qquad\qquad q \in \{1, \ldots, s_{l,l'}\},$$

where $S(\cdot, \cdot)$ denotes the Stirling number of the first kind. Conditional on these auxiliary variables we can update $\boldsymbol{\eta}$ by sampling

$$(\eta_1, \ldots, \eta_{L+1}) \sim \mathsf{Dir}(q_{\cdot,1}, \ldots, q_{\cdot,L}, e),$$

where $q_{\cdot l'} = \sum_{l=1}^{L} q_{l,l'}$.

To sample d, we additionally introduce auxiliary variables $\iota_1, \ldots, \iota_L$ and $u_1, \ldots, u_L$ such that $\iota_l | d \sim \mathsf{Beta}(d+1, s_{l,\cdot})$ and $u_l | d \sim \mathsf{Ber}(s_{l,\cdot}/(d + s_{l,\cdot}))$, where $s_{l,\cdot} = \sum_{l'=1}^{L} s_{l,l'}$. Then, d is sampled from its full conditional distribution,

$$d | \{u_l\}, \{\iota_l\} \sim \mathsf{Gam}\left(1 + q_{\cdot\cdot} - \sum_{l=1}^{L} u_l, 1 - \sum_{l=1}^{L} \log \iota_l\right)$$

where $q_{\cdot\cdot} = \sum_{l=1}^{L}\sum_{l'=1}^{L} q_{l,l'}$. To sample e, we introduce an auxiliary variable υ such that $\upsilon | e \sim \mathsf{Beta}(e+1, q_{\cdot,\cdot})$. Conditional on υ, e follows a mixture distribution

$$\begin{aligned} e | \upsilon, \cdots \sim \frac{L}{q_{\cdot\cdot}(1 - \log\{\upsilon\}) + L} \mathsf{Gam}(L+1, 1 - \log\{\upsilon\}) \\ + \frac{q_{\cdot\cdot}(1 - \log\{\upsilon\})}{q_{\cdot\cdot}(1 - \log\{\upsilon\}) + L} \mathsf{Gam}(L, 1 - \log\{\upsilon\}). \end{aligned}$$

Conditional on the states $\gamma_1, \ldots, \gamma_T$ we can sample the state-specific parameters. In particular, the faction membership indicators $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_L$ are sampled from:

$$\mathsf{Pr}(\zeta_{il} = k | \cdots) \propto \begin{cases} m_{k,l}^{-i} \prod_{k'=1}^{K_l^{-i}} \frac{\Upsilon\left(m_{k',l}^{-i} + r_{k,k',l}^{-i}, \sum_{(j,j',t)\in\Omega_{k',l}^{i}\cup\Omega_{k,k',l}^{-i}} z_{j,j',t}, \sum_{(j,j',t)\in\Omega_{k',l}^{i}\cup\Omega_{k,k',l}^{-i}} z_{j,j',t}^2, \varrho_l, \tau_l^2\right)}{\Upsilon\left(r_{k,k',l}^{-i}, \sum_{(j,j',t)\in\Omega_{k,k',l}^{-i}} z_{j,j',t}, \sum_{(j,j',t)\in\Omega_{k,k',l}^{-i}} z_{j,j',t}^2, \varrho_l, \tau_l^2\right)} & k \leq K_l^{-i} \\ b_l \prod_{k'=1}^{K_l^{-i}} \Upsilon\left(m_{k',l}^{-i}, \sum_{(j,j',t)\in\Omega_{k',l}^{i}} z_{j,j',t}, \sum_{(j,j',t)\in\Omega_{k',l}^{i}} z_{j,j',t}^2, \varrho_l, \tau_l^2\right) & k = K_l^{-i} + 1 \end{cases}$$

where $m_{k,l}^{-i} = \sum_{j\neq i} \mathbf{1}_{\{\zeta_{j,l}=k\}}$, $K_l^{-i} = \max_{j\neq i}\{\zeta_{j,l}\}$, $\Omega_{k',l}^{i} = \{(j,j') : \gamma_t = l, \zeta_{j',l} = k', j = i, j' \neq i\}$, $\Omega_{k,k',l}^{-i} = \{(j,j',t) : \gamma_t = l, \zeta_{j,l} = k, \zeta_{j',l} = k', j \neq i, j' \neq i\}$, $r_{k,k',l}^{-i} = \sum_{(j,j',t)\in\Omega_{k,k',l}^{-i}} \mathbf{1}_{\{\zeta_{j,l}=k, \zeta_{j',l}=k'\}}$.

The precision parameters for each state are sampled from

$$\chi_l|b_l,\cdots \sim \mathsf{Beta}(b_l+1,n)$$
$$b_l|\chi_l,\cdots \sim \frac{K_l}{n(1-\log\{\chi_l\})+K_l}\mathsf{Gam}(K_l+1,1-\log\{\chi_l\}) + \frac{n(1-\log\{\chi_l\})}{n(1-\log\{\chi_l\})+K_l}\mathsf{Gam}(K_l,1-\log\{\chi_l\})$$

The state-specific interaction terms are sampled from

$$\theta^*_{k,k',l}|\cdots \sim \mathsf{N}\left(\left\{r_{k,k',l}+\frac{1}{\tau_1^2}\right\}^{-1}\left\{\sum_{(i,j,t)\in\Omega_{k,k',l}} z_{i,j,t}+\frac{\varrho_1}{\tau_1^2}\right\},\left\{r_{k,k',l}+\frac{1}{\tau_1^2}\right\}^{-1}\right)$$

where $\Omega_{k,k',l}=\{(j,j',t):\zeta_{j,l}=k,\zeta_{j',l}=k',j'>j\}$, $r_{k,k',l}=\sum_{(j,j',t)\in\Omega_{k,k',l}}\mathbf{1}_{\{\zeta_j=k,\zeta_{j'}=k'\}}$.

The baseline parameters for the l-th class are given by

$$\varrho_l|\cdots \sim \mathsf{N}\left(\left\{\frac{K_l(K_l+1)}{2}+\tau_1^2\right\}^{-1}\left\{\sum_{k=1}^{K_l}\sum_{k'=k}^{K_l}\theta^*_{k,k',l}\right\},\left\{\frac{K_l(K_l+1)}{2\tau_1^2}+1\right\}^{-1}\right)$$
$$\tau_l|\cdots \sim \mathsf{IGam}\left(2+\frac{K_l(K_l+1)}{4},1+\frac{1}{2}\sum_{k=1}^{K_l}\sum_{k'=k}^{K_l}(\theta^*_{k,k',l}-\varrho_1)^2\right)$$

The auxiliary variables are sampled as

$$z_{i,j,t}|\cdots \sim \begin{cases}\mathsf{N}(\theta^*_{\zeta_i,\zeta_j,\gamma_t},1)\mathbf{1}_{[0,\infty)} & y_{ij}=1\\ \mathsf{N}(\theta^*_{\zeta_i,\zeta_j,\gamma_t},1)\mathbf{1}_{(-\infty,0)} & y_{ij}=0\end{cases}$$

References

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

Beal, M. J., Ghahramani, Z. & Rasmussen, C. E. (2001). The infinite hidden Markov model. In *Proceedings of Fourteenth Annual Conference on Neural Information Processing Systems.*

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

Besag, J. (2000). Markov chain Monte Carlo for statistical inference. Technical report, University of Washington, Center for Statistics and the Social Sciences.

Blackwell, D. & MacQueen, J. B. (1973). Ferguson distribution via Pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

Cappe, O., Godsill, S. J. & Moulines, E. (2007). An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE* **95**, 899–924.

Carvalho, C. M., Lopes, H. F., Polson, N. G. & Taddy, M. (2009). Particle learning for general mixtures. Technical report, Department of Statistical Sciences - Duke University.

Dahl, D. (2003). An improved merge-split sampler for conjugate Dirichlet process mixture models. Technical report, Department of Statistics, University of Winsconsin.

Davis, J. A. & Leinhardt, S. (1972). The structure of positive interpersonal relations in small groups. In *Sociological Theories in Progress*, volume 2, pp. 218–251. Boston: Houghton Mifflin.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

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

van Gael, J., Saatci, Y., Teh, Y.-W. & Ghahramani, Z. (2008). Beam sampling for the infinite hidden markov model. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*.

George, E. I. & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

Handcock, M. S. (2000). Progress in statistical modeling of drug user and sexual networks. Technical report, University of Washington, Center for Statistics and the Social Sciences.

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

Hoff, P. D. (2007). Discussion of "Model-based clustering for social networks" by Handcock, Raftery and Tantrum. *Journal of the Royal Statistical Society, Series A* **170**, 339.

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

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T. & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the 22nd Annual Conference on Artificial Intelligence*.

Liang, F., Paulo, R., Molina, G., Clyde, M. A. & Berger, J. O. (2008). Mixtures of g-priors for Bayesian variable selection. *Journal of American Statistical Association* **103**, 410–423.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27**, 415–444.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102**, 145–158.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*, Eds. T. S. Ferguson, L. S. Shapeley & J. B. MacQueen, pp. 245–268. Hayward, CA:IMS.

Polson, N. G. & Scott, J. G. (2010). Shrink global ly, act locally: Sparse bayesian regularization and prediction. In *Valencia Statistics 9*.

Read, K. (1954). Cultures of the central highlands, New Guinea. *Southwestern Journal of Anthropology* **10**, 1–43.

Robert, C. P. & Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer, second edition edition.

Rodriguez, A. (2010). On line learning for the infinite hidden Markov model. Technical report, University of California - Santa Cruz.

Scott, J. G. & Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plan. Infer.* **136**, 2144–2162.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph modles. *Journal of Social Structure* **3**.

Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. (2006). Sharing clusters among related groups: Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**, 1566–1581.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Commnunications in Statistics, Part B - Simulation and Computation* **36**, 45–54.

Wang, Y. J. & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19.

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

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

White, H. C., Boorman, S. A. & Breiger, R. L. (1976). Social structure from multiple networks: I, Blockmodles of roles and positions. *American Journal of Sociology* **81**, 730–780.

Xu, Z., Tresp, V., Yu, K. & Kriegel, H.-P. (2006). Infinite hidden relational models. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*.