# Dynamic degree-corrected blockmodels for social networks: a nonparametric approach


## Linda S. L. Tan [1] and Maria De Iorio [2]

[1] Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore

[2] Department of Statistical Science, University College London, UK

---

**Address for correspondence:** Linda S. L. Tan, Department of Statistics and Applied Probability, Block S16, Level 7, 6 Science Drive 2, Faculty of Science, National University of Singapore, Singapore 117546.

**E-mail:** statsll@nus.edu.sg.

**Phone:** (+65) 6516 8473.

**Fax:** (+65) 6872 3919.


---


**Abstract:** A nonparametric approach to the modeling of social networks using degree-corrected stochastic blockmodels is proposed. The model for static network consists of a stochastic blockmodel using a probit regression formulation and popularity parameters are incorporated to account for degree heterogeneity. We specify a Dirichlet process prior to detect community structure as well as to induce clustering in the popularity parameters. This approach is flexible yet parsimonious as it allows the appropriate number of communities and popularity clusters to be determined automatically by the data. We further discuss and implement extensions of the static model to dynamic networks. In a Bayesian framework we perform posterior inference through MCMC algorithms. The models are illustrated using several real-world benchmark social networks.

**Key words:** community detection; stochastic blockmodels; degree correction; Dirichlet process

# 1 Introduction

Social networks play a central role in the dissemination of information (Westerman et al., 2014), formation of alliances (Gulati, 1998), transmission of disease (Cauchemez et al., 2011) and many other areas. It is thus important to understand the underlying structure of a social network and the behavioral patterns in the interactions. A common characteristic of social networks is that they often exhibit *community structure*, where certain groups of nodes (representing the social actors) are more densely connected within each group than across groups. The community structure may be present due to various factors such as similar interests, social stature or physical locations. Studying the nodal attributes associated with the communities can provide a greater understanding of the network topology, behavioral patterns and network dependent processes such as epidemic spreading. However, identifying the community structure in a network can be challenging as the number of communities is typically unknown and the communities can vary in size and rate of interaction. Moreover, results can be distorted if the broad degree distributions often observed in real networks are not taken into account (Karrer and Newman, 2011). In this article, we propose a nonparametric approach to community detection in social networks by using independent Dirichlet processes (Ferguson, 1973) to capture the blockstructure in the social network and induce clustering in the activity level of nodes.

The partitioning of nodes into structurally equivalent groups, such that nodes in the same group relate with other nodes in the same way, was first discussed by Lorrain and White (1971), followed by White et al. (1976). Building upon the work of Holland and Leinhardt (1981) and Fienberg and Wasserman (1981), Holland et al. (1983) generalized this deterministic concept and formulated stochastic blockmodels to allow for data variability. In a stochastic blockmodel, the nodes of a network are partitioned into groups and the distribution of ties between the nodes depends only on the group membership of the nodes and the probabilities of interactions between different groups. The stochastic blockmodel is generative and a wide variety of network structures, such as community, hierarchical or core-periphery, can be produced through different choices of the probability matrix. In *a priori* blockmodeling, exogenous actor attribute data are used to partition the nodes, while the discovery of blockstructures from relational data is referred to as *a posteriori* blockmodeling (Wasserman and Anderson, 1987). Snijders and Nowicki (1997) studied a posteriori blockmodeling for undirected networks when there are only two groups and derived procedures for finding the blockstructure using both maximum likelihood and Bayesian estimation. Nowicki and Snijders (2001) extend their approach to directed valued networks where the number of classes is fixed and address the nonidentifiability problem of the class labels. Handcock et al. (2007) consider a different clustering approach based on latent space models (Hoff et al., 2002), which assumes that the probability of a tie is dependent on the positions of the actors in some unobserved space and decreases with distance.

The stochastic blockmodel has been extended in many ways. To overcome the restriction that each actor can only belong to one group, Airoldi et al. (2008) develop mixed membership stochastic blockmodels (MMSB), where each node is associated with a membership vector describing the probability of the node belonging to each of the groups. Each node can also assume different group membership when interacting with different nodes. Latouche et al. (2011) considers overlapping stochastic blockmodels, where each node can belong simultaneously to multiple groups with independent probabilities. The infinite relational model introduced by Kemp et al. (2006) allows the number of groups to be determined automatically by the data by drawing the membership vector from a Chinese restaurant process (CRP, Pitman, 2006).

Karrer and Newman (2011) point out that the stochastic blockmodel often yield poor fits to real-world networks whose degree distributions are much broader than those generated by the stochastic blockmodel. To account for heterogeneity in the degrees of nodes, they propose degree-corrected stochastic blockmodels, which modify the probability of a tie between node $i$ in group $g_i$ and node $j$ in group $g_j$ from $\omega_{g_i g_j}$ to $\theta_i \theta_j \omega_{g_i g_j}$, where $\omega_{rs}$ denotes the probability of a tie between group $r$ and $s$ while $\theta_i$ measures the activity level or "popularity" of node $i$. Estimates of the parameters are derived using maximum-likelihood and they demonstrate that degree-corrected blockmodels lead to improved community detection. Gopalan et al. (2013) consider a related "assortative MMSB with node popularities" model, that uses a logit link and extends the MMSB to incorporate node popularities. A stochastic variational inference algorithm (Hoffman et al., 2013) is developed for posterior inference. Peng and Carvalho (2016) consider degree-corrected stochastic blockmodels using a Bayesian approach and a logistic regression formulation. Posterior inference is obtained via data augmentation with latent Pólya-Gamma variables and a canonically mapped centroid estimator that addresses label non-identifiability.

In this article, we focus on degree-corrected stochastic blockmodels for community detection in undirected social networks using a nonparametric Bayesian approach. The static model is formulated using probit regression and a Dirichlet process (DP) prior (Ferguson, 1973) is employed to capture the communities in the network and induce clustering among the popularity parameters. The DP is widely used in Bayesian nonparametric models, particularly in DP mixture models, as a prior over distributions. It is well known that the DP is almost surely discrete. If the random probability measure $G$ is a DP$(\alpha, G_0)$ with total mass parameter $\alpha$ and baseline distribution $G_0$, then $G$ can be represented as

$$G(\cdot) = \sum_{h \geq 1} w_h \delta_{\theta_h}(\cdot),$$

where $\delta_\theta$ is a point-mass at $\theta$ and the atoms $\{\theta_h\}_{h\geq 1}$ are such that $\theta_h \overset{\text{iid}}{\sim} G_0$. The weights follow a stick-breaking process (Sethuraman, 1994), $w_h = V_h \prod_{j<h}(1 - V_j)$, where $V_h \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha)$. Due to the discreteness of the DP, the prior induces clustering of the subjects in the sample based on the unique values of the parameters $\theta_h$, where the number of clusters $K$ is unknown and learned from the data. The nonparametric approach is highly flexible yet parsimonious as it does not require the number of communities to be fixed in advance and instead allows the appropriate number of communities and popularity clusters to be determined automatically by the data.

Our model integrates the approach of Kemp et al. (2006) who use the CRP to detect community structure and Ghosh et al. (2010), who use the DP to induce clustering among the "productivity" and "attractiveness" parameters of a variation of the $p_1$

model (Holland and Leinhardt, 1981) and a social relations model (Gill and Swartz, 2007). The infinite-degree-corrected stochastic block model (Herlau et al., 2014) also uses the CRP for community detection. However, they consider weighted links modeled using a Poisson distribution and they do not consider clustering of the degree parameters. The nonparametric network model introduced by Williamson (2016) also uses the DP, albeit for clustering of links instead of the nodes and represents a network as a possibly infinite sequence of node pairs.

We derive a Gibbs sampler for posterior inference and discuss some ways in which the static model can be extended to dynamic networks. There is an extensive literature on community detection in dynamic networks. For instance, Xing et al. (2010) extends MMSB to dynamic networks using state space models by allowing the membership of the nodes to vary with time while Sewell and Chen (2017) extend latent space models to dynamic networks. Here we focus on changes in the activity level of individual nodes and the dependency of a network on its previous state. The applicability of the proposed models are illustrated using benchmark social networks.

This article is organized as follows. In Section 2 we present a model for static networks and discuss extensions of this model to dynamic networks. In Section 3, we describe how posterior inference for the proposed models can be obtained using Gibbs samplers. In Section 4, we use the proposed models to analyze three real-world social networks. We conclude with a discussion of future research directions in Section 5. Online supplemental material for this article are available.

# 2 Nonparametric models for social networks

Let $N = \{1, \ldots, n\}$ be the set of actors of interest and $y = [y_{ij}]$ be a $n \times n$ adjacency matrix where $y_{ij}$ is an indicator of a link from actor $i$ to actor $j$. In this article, we focus on undirected networks without self-links and multiple links. Hence $y$ is symmetric and the diagonal elements of $y$ are zeros. When the network of interest is observed at multiple (discrete) time points, $T$, we let $y_t = [y_{t,ij}]$ be the $n \times n$ adjacency matrix representing the state of the network at time $t$ for $t = 1, \ldots, T$.

## 2.1 Static model

First we introduce a model for a static network $y$ that aims to detect community structure while incorporating actor heterogeneity through node-specific popularity parameters. Let $\mathcal{S} = \{(i,j)|1 \leq i < j \leq n\}$. For $(i,j) \in \mathcal{S}$, we assume

$$y_{ij}|p_{ij} \overset{\text{indep}}{\sim} \text{Bernoulli}(p_{ij}),$$

and introduce latent variable $\zeta_{ij}|\mu_{ij} \overset{\text{indep}}{\sim} N(\mu_{ij}, 1)$, where

$$\mu_{ij} = \theta_i + \theta_j + \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{z_i = z_j = k\}, \tag{2.1}$$

such that $y_{ij}|\zeta_{ij} = 1$ if $\zeta_{ij} > 0$ and $0$ if $\zeta_{ij} \leq 0$ and $\mathbb{1}\{\cdot\}$ denotes the indicator function. Here we exploit the latent variable representation of binary variables (Albert and Chib, 1993) and assume a probit link where $\Phi^{-1}(p_{ij}) = \mu_{ij}$ and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal. The parameter $\theta_i$ represents the *popularity* or *activity level* of actor $i$, $K \leq n$ denotes the total number of groups or communities in the network and $z_i \in \{1, \ldots, K\}$ represents the group membership of actor $i$. The coefficient $\beta_k^*$ measures the rate of interaction in the $k$th community.

Members within a community are assumed to interact with each other at a common rate. A high $\beta_k^*$ indicates a tight or close-knit community where members interact at a high rate while a low $\beta_k^*$ indicates a group with little interaction. The third term on the right-hand side of (2.1) resembles a stochastic blockmodel where non-diagonal entries of the probability matrix are set to a common value (not necessarily zero). In (2.1), the probability of interaction, $p_{ij}$, between actors $i$ and $j$ depends on their individual popularities as well as the interaction rate of their community if they belong to the same community. An interaction between actors from different communities is driven only by their popularities. Thus the presence of a link can be explained by homophily in terms of community membership or popularities and the popularity parameters $\{\theta_i\}$ and community assignments $\{z_i\}$ are *competing* to explain the observed network. Note that $K$ is unknown and the object of inference.

For model parsimony, a DP is used to induce clustering among the popularity parameters $\{\theta_i\}$. We assume

$$\begin{aligned} \theta_i|G &\overset{\text{iid}}{\sim} G \quad \text{for} \quad i = 1, \ldots, n, \\ G &\sim \text{DP}(\alpha, G_0), \end{aligned} \tag{2.2}$$

where the base distribution $G_0$ is $N(0, \sigma_\theta^2)$ and $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$. Let $\theta^* = [\theta_1^*, \ldots, \theta_L^*]^T$ denote the set of unique values among $\{\theta_1, \ldots, \theta_n\}$ and $c_i$ indicate the latent class associated with $\theta_i$ so that $\theta_i = \theta_{c_i}^*$. Clustering the popularity parameters results in a more parsimonious model representation assuming $L \ll n$. Moreover, such a model is easier to interpret and is useful in understanding the network structure. It also helps to identify influential nodes in the network who have high popularity and possibly connections across multiple communities (see also Vu et al., 2013).

To detect the communities in the network, we employ again a DP, $H$, which is independent of $G$. We introduce a $\beta_i$ for each actor $i$ where $\beta_i = \beta^*_{z_i}$ and assume

$$\begin{aligned} \beta_i|H &\overset{\text{iid}}{\sim} H \quad \text{for} \quad i = 1, \dots, n, \\ H &\sim \text{DP}(\nu, H_0), \end{aligned} \tag{2.3}$$

where $H_0$ is $N(0, \sigma^2_\beta)$ and $\nu \sim \text{Gamma}(a_\nu, b_\nu)$. Let $\beta^* = [\beta^*_1, \dots, \beta^*_K]^T$ be the set of unique values among $\{\beta_1, \dots, \beta_n\}$.

In the proposed approach, the number of clusters $L$ among $\{\theta_i\}$ and the number of communities $K$ are not fixed in advance. Instead, they are random and to be inferred from the data. The prior distribution on $L$ depends on the concentration parameter $\alpha$, with a larger $\alpha$ implying a larger $L$ a priori. To avoid overfitting, we opt for a prior for $\alpha$ that favors small values relative to $n$. We specify a gamma prior on $\alpha$ which facilitates computations. The prior specification on $\alpha$ also affects the prior distribution of the number $L$ of components. The conditional mean and variance of $L$ given the concentration parameter of the DP and the sample size $n$ are $\text{E}(L \mid \alpha) = \sum_{i=1}^n \frac{\alpha}{\alpha+i-1}$ and $\text{Var}(L \mid \alpha) = \sum_{i=1}^n \frac{\alpha(i-1)}{(\alpha+i-1)^2}$ (Liu, 1996; Jara et al., 2007). Jara et al. (2007) also derive the following marginal expressions for $\text{E}(L)$ and $\text{Var}(L)$ when $\alpha$ is given a $\text{Gamma}(a_0, b_0)$ prior:

$$\text{E}(L) \approx \frac{a_0}{b_0} A, \quad \text{Var}(L) \approx \text{E}(L^*) + \frac{a_0^2}{b_0^2} B + \left\{ \frac{a_0}{b_0} B + A \right\}^2 \frac{a_0}{b_0^2},$$

where $A = \psi_0(\frac{a_0+nb_0}{b_0}) - \psi_0(\frac{a_0}{b_0})$, $B = \psi_1(\frac{a_0+nb_0}{b_0}) - \psi_1(\frac{a_0}{b_0})$, and $\psi_0(\cdot)$ and $\psi_1(\cdot)$ represent the digamma and trigamma functions respectively. These may be used as reference for setting hyperparameter values in the gamma prior. For example, in the karate network in Section 4.1, where $n = 34$, a Gamma(5, 5) prior implies a prior mean of 4.1 and a prior variance of 3.8. The relation between $K$ and $\nu$ is similar.

Next, we propose extensions of the static model to dynamic networks that evolve over time. Suppose we observe networks $y_t = [y_{t,ij}]$ for $t = 1, \dots, T$. For the dynamic models below, we assume that for $t = 1, \dots, T$, $(i, j) \in \mathcal{S}$,

$$y_{t,ij} | p_{t,ij} \overset{\text{indep}}{\sim} \text{Bernoulli}(p_{t,ij}).$$

As before we consider the probit link function and introduce latent variables $\zeta_{t,ij} | \mu_{t,ij} \sim N(\mu_{t,ij}, 1)$ for $(i, j) \in \mathcal{S}$, $t = 1, \dots, T$ such that $y_{t,ij} | \zeta_{t,ij} = 1$ if $\zeta_{t,ij} > 0$ and 0 if $\zeta_{t,ij} \leq 0$. Thus, $p_{t,ij} = \Phi(\mu_{t,ij})$.

## 2.2 Dynamic model 1

Dynamic model I assumes that the community memberships remain unchanged over time but the popularities of the actors can vary with time. This assumption is appropriate for data where the communities arise due to factors that do not or are unlikely to vary drastically over time, for instance, gender, race, physical locations and job positions. In such cases, the changes in ties may be attributed to variations in the activity levels of individual nodes. For $(i, j) \in \mathcal{S}$ and $t = 1, \dots, T$, let

$$\mu_{t,ij} = \theta_{it} + \theta_{jt} + \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{z_i = z_j\}.$$

In resemblance of the static model, we assume that the $\{\theta_{it}\}$ are independent and induce clustering among them using a DP,

$$\theta_{it} | G \overset{\text{iid}}{\sim} G \;\; \text{for} \;\; i = 1, \dots, n, \;\; t = 1, \dots, T,$$
$$G \sim \text{DP}(G_0, \alpha),$$

where $G_0$ is $N(0, \sigma_\theta^2)$ and $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$. For this model, let $\theta^* = [\theta_1^*, \dots, \theta_L^*]^T$ denote the set of unique values among $\{\theta_{11}, \dots, \theta_{nT}\}$ and $c_{it}$ indicate the latent class associated with $\theta_{it}$ so that $\theta_{it} = \theta_{c_{it}}^*$ for $i = 1, \dots, n$, $t = 1, \dots, T$. The $\{\beta_k^*\}$ and $\{z_i\}$ are modeled using a DP as described in (2.3).

### 2.3 Dynamic model II

Dynamic model II extends the static model by allowing the tie between nodes $i$ and $j$ at time $t$ to depend on the existence of the tie at the previous time point. It assumes that the popularities and community memberships of the actors remain unchanged over time. For $(i, j) \in \mathcal{S}$ and $t = 1, \ldots, T$, let

$$\mu_{t,ij} = \eta y_{t-1,ij} \mathbb{1}\{t > 1\} + \theta_i + \theta_j + \sum_{k=1}^{K} \beta_k^* \mathbb{1}\{z_i = z_j\},$$

where $\eta \sim N(0, \sigma_\eta^2)$. The coefficient $\eta$ can be interpreted as a measure of the *persistence* of ties in the network once they are formed. A positive $\eta$ implies that a tie is more likely to be present at time $t$ if it was present at time $t-1$ than if it were not, conditional on their popularities and community memberships. On the other hand, a negative $\eta$ implies that a tie is more likely to be present at time $t$ if the tie was absent at the previous time point than if it were present. The parameters $\{\theta_i\}$ are modeled as in (2.2) and $\{z_i\}$ and $\{\beta_k^*\}$ are modeled as in (2.3). The popularities and communities inferred from this model smooths out the noise in the data and provide an overview of the behavior of actors over time.

## 3 Posterior inference

We use Gibbs samplers to perform posterior inference for the proposed models. To obtain the updates in the Gibbs sampler, we derive the posterior distribution of each variable conditional on the rest. Detailed derivations are given in the supplemental material. Sampling from the DP is performed using the methods described in Neal (2000) while the concentration parameters $\alpha$ and $\nu$ are sampled using the method described in Escobar and West (1995).

We introduce the following notations. Let $Z_{ij}$ be a binary vector of length $K$ where the $k$th element is 1 if $z_i = z_j = k$, 0 otherwise, and $Z = [Z_{12}, Z_{13}, \dots, Z_{(n-1),n}]^T$ be a $n(n-1)/2 \times K$ matrix. We define $\zeta_{ij} = \zeta_{ji}$ and $\zeta_{t,ij} = \zeta_{t,ji}$ for $1 \le j < i \le n$ and $t = 1, \dots, T$. Let $\mathcal{S}_m = \{(i,j) \in \mathcal{S} | c_i = c_j = m\}$, $\mathcal{S}_{t,m} = \{(i,j) \in \mathcal{S} | c_{it} = c_{jt} = m\}$, $\mathcal{P}_m = \{(i,j) | j \neq i, c_i = m, c_j \neq m\}$, and $\mathcal{P}_{t,m} = \{(i,j) | j \neq i, c_{it} = m, c_{jt} \neq m\}$. We use $\text{TN}(x|\mu, \sigma, a, b)$ to denote the truncated normal distribution with density $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})/(\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}))$, where $\phi(\cdot)$ denotes the density of the standard normal distribution. In the algorithms presented below, we use $K$ and $L$ to represent the current number of communities and popularity clusters respectively at each iteration and $\beta^* = [\beta_1^*, \dots, \beta_K^*]$ and $\theta^* = [\theta_1^*, \dots, \theta_L^*]$ to represent the states currently associated with the clusters.

For the static model, the joint distribution $p(y, \zeta, z, \beta^*, \nu, c, \theta^*, \alpha)$ is given by

$$p(c|\alpha)p(\alpha)p(z|\nu)p(\nu)p(\theta^*)p(\beta^*) \prod_{i<j} p(y_{ij}|\zeta_{ij})p(\zeta_{ij}|c_i, c_j, \theta^*, z_i, z_j, \beta^*),$$

where $\zeta = \{\zeta_{11}, \dots, \zeta_{n-1,n}\}$, $c = \{c_1, \dots, c_n\}$ and $z = \{z_1, \dots, z_n\}$. Note that $p(c|\alpha)$ and $p(z|\nu)$ are defined in Neal (2000) as

$$\begin{aligned} \text{P}(z_i = k | z_{-i}, \nu) &= \frac{m_{-i,k}}{n-1+\nu} \text{ for } k \in z_{-i}, \ \text{P}(z_i \neq z_j \text{ for all } j \neq i | z_{-i}, \nu) = \frac{\nu}{n-1+\nu}, \\ \text{P}(c_i = \ell | c_{-i}, \alpha) &= \frac{n_{-i,\ell}}{n-1+\alpha} \text{ for } \ell \in c_{-i}, \ \text{P}(c_i \neq c_j \text{ for all } j \neq i | c_{-i}, \alpha) = \frac{\alpha}{n-1+\alpha}, \end{aligned} \tag{3.1}$$

where $z_{-i} = z \backslash z_i$, $c_{-i} = c \backslash c_i$, $m_{-i,k} = \sum_{z_j \in z_{-i}} \mathbb{1}\{z_j = k\}$ and $n_{-i,\ell} = \sum_{c_j \in c_{-i}} \mathbb{1}\{c_j = \ell\}$. The Gibbs sampler for the static model is outlined in Algorithm 1. Step 2 supposes that the number of distinct values in $z_{-i}$ is $K'$. In the update, $z_i$ can either assume one of these $K'$ distinct values or a new value not assumed by any $z_j \in z_{-i}$. Algorithm (3.2) describes the probabilities of cluster assignment and $a$ is a constant that ensures these $K'+1$ probabilities sum to one. Similarly, (3.3)–(3.7) describe the update steps of cluster assignment.

For dynamic model I, the joint distribution $p(y, \zeta, z, \beta^*, \nu, c, \theta^*, \alpha)$ is given by

$$p(c|\alpha)p(\alpha)p(z|\nu)p(\nu)p(\theta^*)p(\beta^*)\prod_{t=1}^{T}\prod_{i<j}p(y_{t,ij}|\zeta_{t,ij})p(\zeta_{t,ij}|c_{it}, c_{jt}, \theta^*, z_i, z_j, \beta^*),$$

where $\zeta = \{\zeta_{1,11}, \ldots, \zeta_{T,n-1,n}\}$, $c = \{c_{11}, \ldots, c_{nT}\}$ and $z = \{z_1, \ldots, z_n\}$. Note that $p(c|\alpha)$ is defined as

$$\mathrm{P}(c_{it} = \ell | c_{-it}, \alpha) = \frac{n_{-it,\ell}}{nT - 1 + \alpha} \text{for } \ell \in c_{-it},$$
$$\mathrm{P}(c_{it} \text{ is not equal to any value in } c_{-it} | c_{-it}, \alpha) = \frac{\alpha}{nT - 1 + \alpha},$$

where $c_{-it} = c \backslash c_{it}$ and $n_{-it,\ell}$ is the number of indicators in $c_{-it}$ that are equal to $\ell$. The definition of $p(z|\nu)$ remains as in (3.1). The Gibbs sampler for dynamic model I is outlined in Algorithm 2.

For dynamic model II, the joint distribution is given by

$$p(y, \zeta, z, \beta^*, \nu, c, \theta^*, \alpha) = p(c|\alpha)p(\alpha)p(z|\nu)p(\nu)p(\theta^*)p(\beta^*)p(\eta)$$
$$\times\prod_{i<j}\Big\{\Big[\prod_{t\geq 1}p(y_{t,ij}|\zeta_{t,ij})\Big]p(\zeta_{1,ij}|c_i, c_j, \theta^*, z_i, z_j, \beta^*)\Big[\prod_{t\geq 2}p(\zeta_{t,ij}|c_i, c_j, \theta^*, z_i, z_j, \beta^*, \eta, y_{t-1,ij})\Big]\Big\},$$

where $\zeta = \{\zeta_{1,11}, \ldots, \zeta_{T,n-1,n}\}$, $c = \{c_1, \ldots, c_n\}$ and $z = \{z_1, \ldots, z_n\}$. Note that $p(c|\alpha)$ and $p(z|\nu)$ are as defined in (3.1). The Gibbs sampler for dynamic model II is outlined in Algorithm 3.

We code Algorithms 1–3 in Julia and all experiments are run on an Intel Core i5 CPU @ 3.30GHz, 8.0GB RAM. It is also possible to use softwares such as OpenBUGS to obtain posterior inference for the proposed models by considering a truncated DP approach (Ishwaran and Zarepour, 2000). However, we observe that the runtime in

OpenBUGS is significantly longer than Julia especially as the number of nodes $n$ and the number of time points $T$ increase. The computational complexity of Algorithms 1–3 scales as $\mathcal{O}(Tn^2)$ since the first step cycles through each time point and node pair. Moreover, the number of clusters in the DP ($L$ and $K$) is expected to grow logarithmically with the number of the nodes (Antoniak, 1974). We include Julia codes for Algorithms 1–3 and an OpenBUGS code for implementing the static model in the supplementary material. Extending the OpenBUGS code to the dynamic models is straightforward.

We have experimented with using logistic instead of probit regression via OpenBUGS and results are similar. The normal priors and gamma priors are adopted due to conjugacy so that the updates in the Gibbs sampler are of closed form. For the normal prior, we use a variance of one in the applications and doubling the variance results in minimal changes, although using a very large variance will result in instability in OpenBUGS. We think that a variance of 1 is sufficient since $p_{ij} = \Phi^{-1}(-3) = 0.001$ and $p_{ij} = \Phi^{-1}(3) = 0.999$. The model fit is more sensitive to the gamma prior parameters (see example in Section 4.1). We recommend using the results of Jara et al. (2007) for the expectation and variance of the number of clusters as a guideline for setting the gamma prior parameters and to vary the parameters to test the sensitivity of the model.

## 3.1 Cluster Analysis

We can perform posterior inference on the clustering structure using the MCMC output. We derive a posterior estimate of the clustering by computing the posterior similarity matrix $S$, which is a $n \times n$ symmetric matrix whose $(i, j)$ entry contains the posterior probability that actors $i$ and $j$ belong to the same cluster. This probability is estimated by the proportion of times actors $i$ and $j$ cluster together and it is not affected by the problem of "label-switching" (labels associated with clusters may change during MCMC runs, see e.g. Stephens, 2000) or the number of clusters varying across iterations.

We can also compute a single (hard) clustering estimate by using the maximum a posteriori (MAP) approach or methods based on the posterior similarity matrix or Rand index (see discussion in Fritsch and Ickstadt, 2009). Here we consider the Binder's loss function (Binder, 1978), which is defined as the total number of disagreements between the estimated and true clustering among all pairs of actors. The `R` package `mcclust` provides a function, `minbinder`, that can be used to find the clustering $c^* = [c_1^*, \ldots, c_n^*]$ which minimizes the posterior expectation of this loss. The posterior expected loss can be written as

$$\sum_{i<j} |\mathbb{1}_{\{c_i^* = c_j^*\}} - S_{ij}|, \tag{3.8}$$

where the sum is taken over all possible pairs of actors and $S_{ij}$ is the $(i, j)$ entry of the posterior similarity matrix.

# 4 Applications

We investigate the performance of the static and dynamic models on three well-known social network datasets and compare the fitted models with results obtained previously in published literature. The first is a karate club network studied by Zachary (1977) from 1970 to 1972, the second is a dolphins social network (Lusseau et al., 2003) and the third is a dataset collected by Kapferer (1972) at an African clothing factory in Zambia. These datasets are available at the UCI Network Data Repository (https://networkdata.ics.uci.edu/). The analysis on dolphins social networks is available in the supplemental material together with a simulated data example to test the sensitivity of the static model.

In the MCMC implementation, we run multiple chains from dispersed starting positions and use trace plots and kernel density plots to assess the convergence of the MCMC chains as well as the length of burn-in to be discarded. The parameters $L$, $K$, $\nu$, $\alpha$, $\eta$, $\beta$ and $\theta$ are monitored.

## 4.1 Karate club network

This dataset contains 78 undirected friendship links among 24 members, which are constructed based on interactions outside of club activities. Due to disputes over the price of karate lessons, the club was divided informally into two factions, led by the karate instructor "Mr Hi" (actor 1) and the president "John A." (actor 34) respectively (these names are pseudonyms). During the study, the club eventually split into two separate clubs when Mr Hi was fired for trying to raise lesson fees unilaterally and his supporters left to join the new club formed by Mr Hi. All members joined clubs following their own factions except actor 9, who crossed factions to join Mr Hi's club because he was only three weeks away from a test for black belt at the time of the split and he could not bear to give up his rank.

We fit the static model to this dataset using Algorithm 1, setting $a_\nu = b_\nu = a_\alpha = b_\alpha = 5$ and $\sigma_\theta^2 = \sigma_\beta^2 = 1$. Three chains were run in parallel, each consisting of 40,000 iterations with the first 30,000 discarded as burn-in. The total runtime is 172 seconds. A thinning factor of 5 was applied and the remaining 6000 samples were used

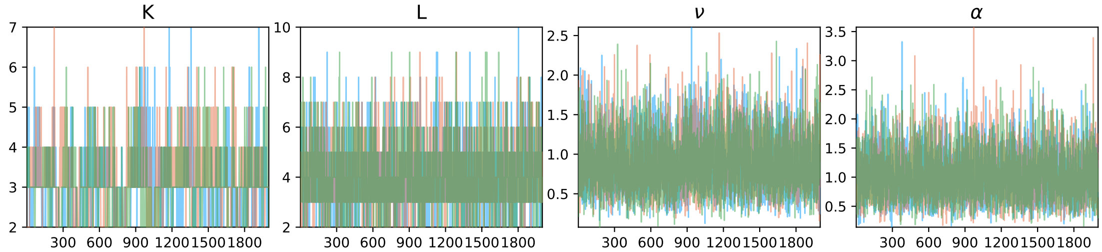


Figure 1: MCMC trace plots of $K$, $L$, $\nu$ and $\alpha$ (3 chains) after discarding burn-in and thinning.

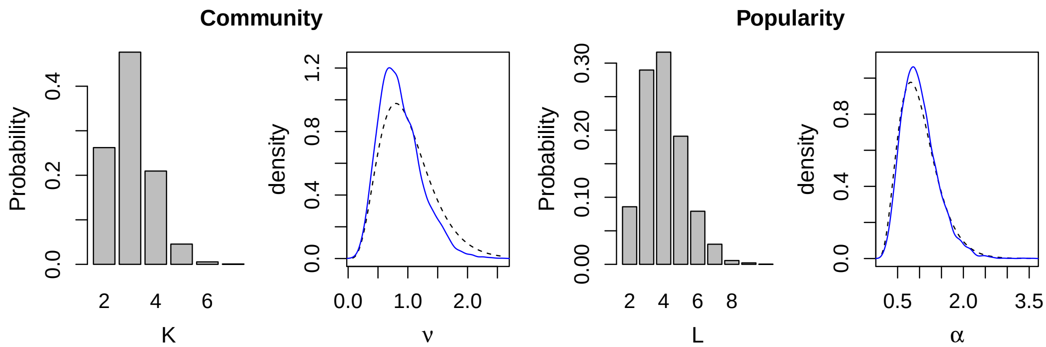


Figure 2: Posterior distributions of $K$, $\nu$ $L$ and $\alpha$. For $\nu$ and $\alpha$, the prior distributions are shown in dotted lines and the posterior distributions in solid (blue) lines.

for posterior inference. Figure 1 shows the trace plots for some of the parameters.

Figure 2 shows the posterior distributions of the number of communities ($K$), the number of popularity clusters ($L$), and the DP concentration parameters $\alpha$ and $\nu$. The mode of $K$ is 3 and that of $L$ is 4. The fitted model is quite parsimonious with a relatively small number of clusters for both popularity and community. Figure 3 shows the posterior similarity matrices for the clusterings according to community (left) and popularity (right). Figure 4 plots the posterior mean of $\theta_i$ against the degree for each actor. While the factional leaders, Mr Hi (actor 1) and John A. (actor 34), and a few other actors $\{2, 3, 33\}$ have high popularity, the rest of the members have much lower activity levels generally.

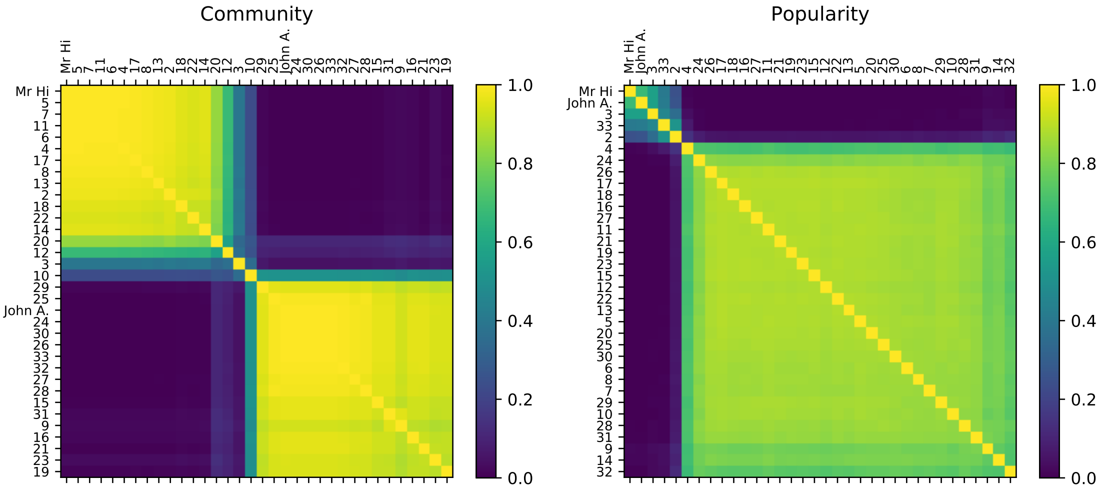


Figure 3: Posterior similarity matrices for community (left) and popularity (right).

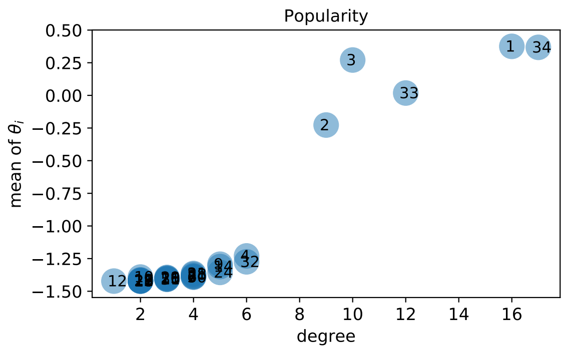


Figure 4: Plot of posterior mean of $\theta_i$ against actor $i$'s degree.

Using the similarity matrices, we compute hard clustering estimates using Binder's loss function. There are three communities, one containing a single node {3} and three popularity clusters. Figure 5 shows the fit of the static model to the karate club network. We run Algorithm 1 again, fixing $z$ and $c$ to obtain estimates of $\beta^*$ and $\theta^*$ conditional on clustering structure. The conditional posterior mean and standard deviation (in brackets) of these parameters are shown in the legend of Figure 5. There are three clusters for the popularity parameters $\{\theta_i\}$, the first contains {Mr Hi, John A., 3}, the second contains {2, 33} and the third contains all remaining members.

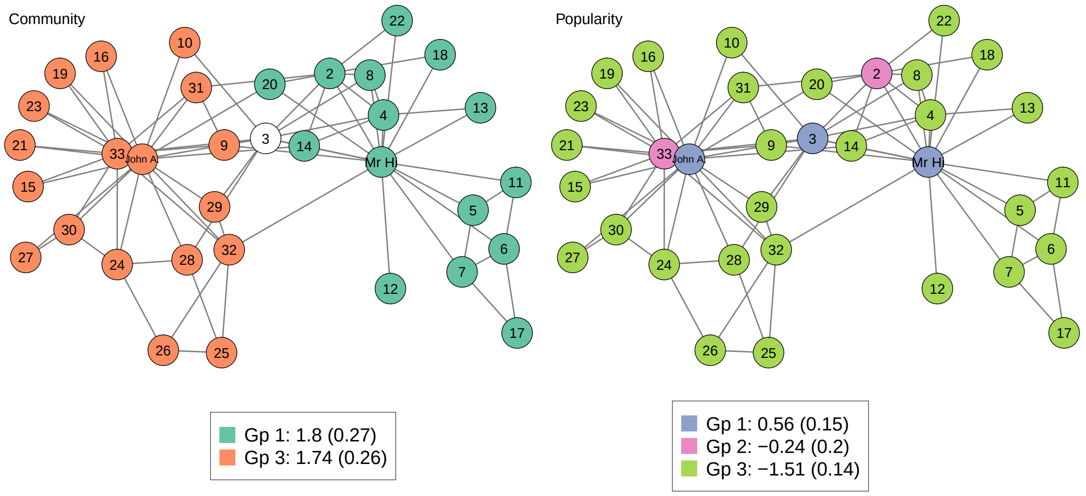


Figure 5: Fit of static model to karate club network. Nodes of the same color belong to the same cluster and singletons are not colored.

For the communities, we note that the $\beta_k^*$ for groups 1 and 3 are strongly positive, indicating a high interaction rate within each group. The posterior mean and standard deviation of $\beta_k^*$ for the singletons necessarily equal that of the prior distribution. Group 3 corresponds *exactly* to the faction led by John A. as concluded in Zachary (1977) while group 1 together with the singleton $\{3\}$ correspond to the faction led by Mr Hi. From the posterior probability matrix, actor 3 has a 0.4 probability of being clustered with members in group 1 and a 0.05 probability of being clustered with members in group 3. It is thus reasonable to combine actor 3 with group 1. Hence, our proposed static model is able to identify members in the factions accurately.

Incidentally, if we drop $\{\theta_i\}$ from the static model and consider just the blockmodel, we obtain five clusters, four of which are singletons: $\{1\}$, $\{3\}$, $\{33\}$, $\{34\}$ and the fifth cluster contains all other members. This result is similar to the phenomenon discussed in Karrer and Newman (2011), who note that the non-degree-corrected blockmodel with $K = 2$ splits the network into high-degree and low-degree nodes instead of by factions, while the degree-corrected version splits it according to factions albeit with one misclassification. In addition, Bickel and Chen (2009) observe that the non-degree-corrected blockmodel with $K = 4$ splits the network according to factions correctly after the merging of sub-communities. These observations highlight the importance of accounting for degree variation in blockmodels as well as the difficulties in determining an appropriate number of clusters. Our static model tries to address these issues using a nonparametric approach via the automatic clustering structures induced by the DP. We observed that the clusters identified by the static model can be sensitive to the DP concentration parameters in some cases. For example, if we adopt a more conservative prior, say by setting $a_\nu = b_\nu = a_\alpha = b_\alpha = 10$, then we obtain three communities, the first corresponds to the faction led by John A., the second contains $\{5, 6, 7, 11, 17\}$ and the third contains all remaining members. Here, the second cluster emerges as one with a higher interaction rate than the third. However, merging the second and third clusters still yields Mr Hi's faction.

While the clustering estimates return hard partitions of the network which are easy to interpret, the posterior similarity matrices reveal finer details regarding the degree of affiliation of actors towards the clusters that they are assigned to in the hard split. The posterior similarity matrix of popularities shows evidence of two main blocks but the partitioning among actors $\{1, 34, 3, 33, 2\}$ is not straightforward. For the posterior similarity matrix for communities, actor 10 is assigned to the cluster led by John A., but he has a slightly lower posterior probability ($\approx 0.5$) of being together with the other members in this cluster than the rest, and also has a positive posterior probability ($\approx 0.2$) of being in the same cluster as members in Mr Hi's faction.

## 4.2 Kapferer's tailor shop network

Kapferer (1972) collected data on the interactions among 39 workers in a tailor shop in Zambia, Southern Africa, from June 1965 to February 1966, and he examined how these social networks relate to major events taking place in the factory. The workers' duties can be classified into eight categories: head tailor (worker number 19), cutter (16), line 1 tailor (1–3, 5–7, 9, 11–14, 21, 24), button machiner (25–26), line 3 tailor (8, 15, 20, 22–23, 27–28), ironer (29, 33, 39), cotton boy (30–32, 34–38) and line 2 tailor (4, 10, 17–18). These positions require different levels of skills and some like the head tailor, cutter, line 1 tailors and button machiners were perceived as having more prestige. Here we focus on the symmetric "sociational" networks (based on convivial interactions) recorded at two time points, the first was before an aborted strike and the second was after a successful strike for higher wages. The network at the second time point (223 edges) is much denser than the first (158 edges) as the workers strive to be more united (thereby expanding their social relations) in their efforts to change the wage system. This dataset has been widely studied, for instance, by Mitchell (1989) and Nowicki and Snijders (2001) using block structures and Thiemichen et al. (2016) using Bayesian exponential random graph models. For each of the models fitted in this section, we run three MCMC chains in parallel. Each chain uses 15,000 iterations, with the first 5000 discarded as burn-in. A thinning factor of 5 was applied and posterior inferences are based on the remaining 6000 iterations. The hyperparameters are set as $a_\nu = b_\nu = a_\alpha = b_\alpha = 10$ and $\sigma_\theta^2 = \sigma_\beta^2 = \sigma_\eta^2 = 1$. In the analysis below, the nodes are colored according to the inferred clusters and singletons are not colored.

First, we fit the static model to the two time points separately to investigate whether

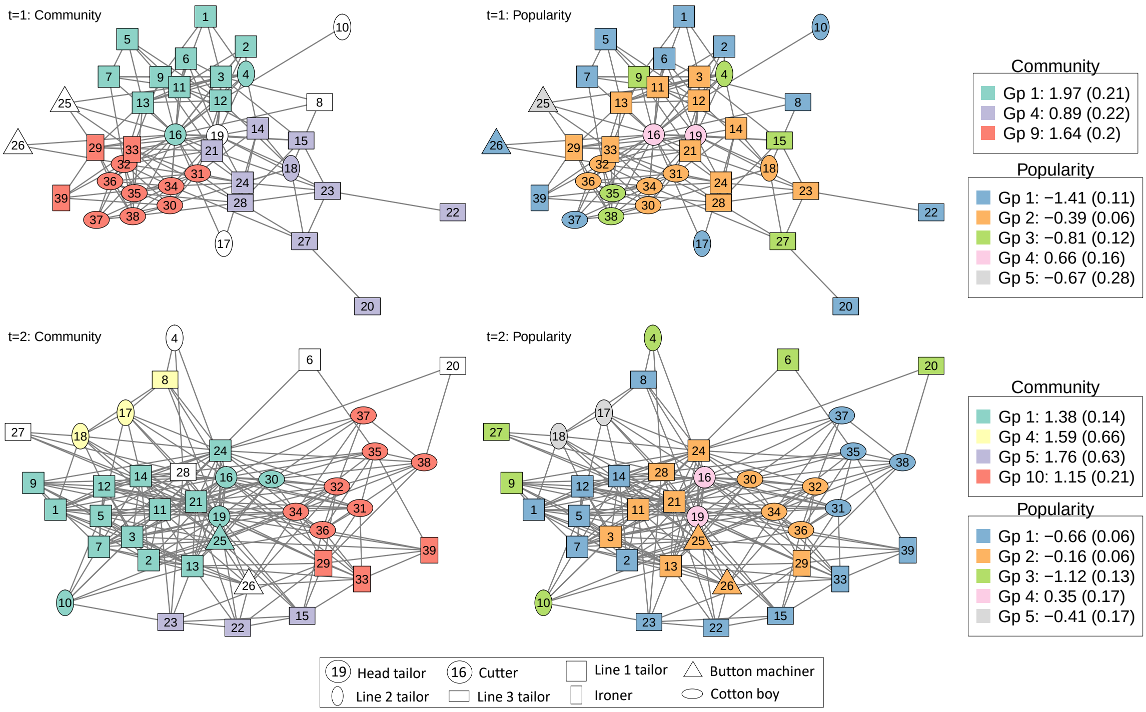


Figure 6: Fit of separate static models to Kapferer social networks at $t = 1$ and $t = 2$.

the assumption that community memberships remain unchanged over time is suitable. The results are shown in Figure 6. Three communities are detected at $t = 1$ and four at $t = 2$. It seems that there are some changes to the community memberships. In particular, some members in group 4 at $t = 1$ have joined group 1 at $t = 2$ while some others have become singletons. Most members in group 1 and especially group 9 remain intact. If we exclude actors who are singletons at either $t = 1$ or $t = 2$, then there are 28 actors left, of which 23 of the actors' community membership remain unchanged. An important observation is that the communities have a strong association with the workers' duties, which also provides some basis for the assumption that community memberships remain unchanged.

### *4.2.1 Dynamic model I*

Fitting dynamic model I using Algorithm 2 took 139 seconds. Dynamic model I assumes that the communities remain constant over time and the emergence or dissolution of ties are due to changes in the activity level of individual actors. The posterior distributions of $K$, $L$, $\nu$ and $\alpha$ are shown in Figure 7. The mode of $K$ is 6 and the mode of $L$ is 4.

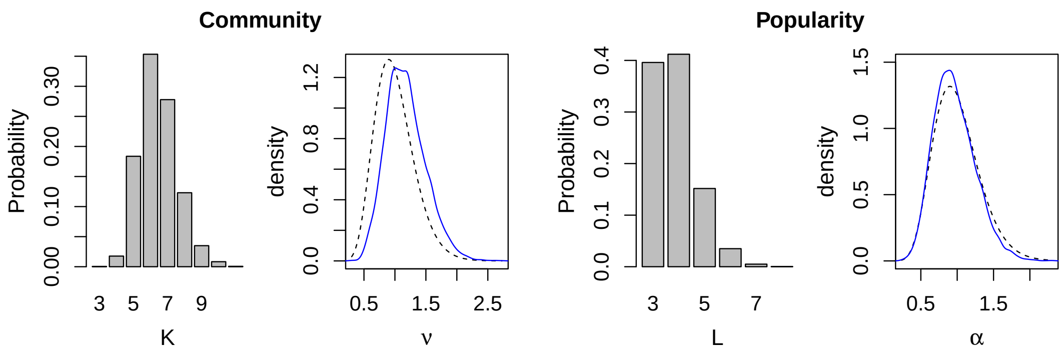


Figure 7: Posterior distributions of $K$, $\nu$, $L$ and $\alpha$. For $\nu$ and $\alpha$, the prior distributions are shown in dotted lines and the posterior distributions in solid (blue) lines.

Next we compute the posterior similarity matrices and use Binder's loss function to obtain hard clustering estimates. This yields nine communities and three popularity clusters. Of the nine communities, five are singletons so there are essentially only four communities. We run Algorithm 2 again, fixing $z$ and $c$ to obtain estimates of $\beta^*$ and $\theta^*$ conditional on the clustering structure. The results are shown in Figure 8, and the mean and standard deviation (in brackets) of $\beta^*$ and $\theta^*$ are reported for each group. The first row shows the four communities which are constant across the two time points. The shapes of the nodes represent the duties of the workers as explained in the legend. The plots indicate a high degree of job homophily in the communities even though these social networks are constructed based on casual interactions ("general

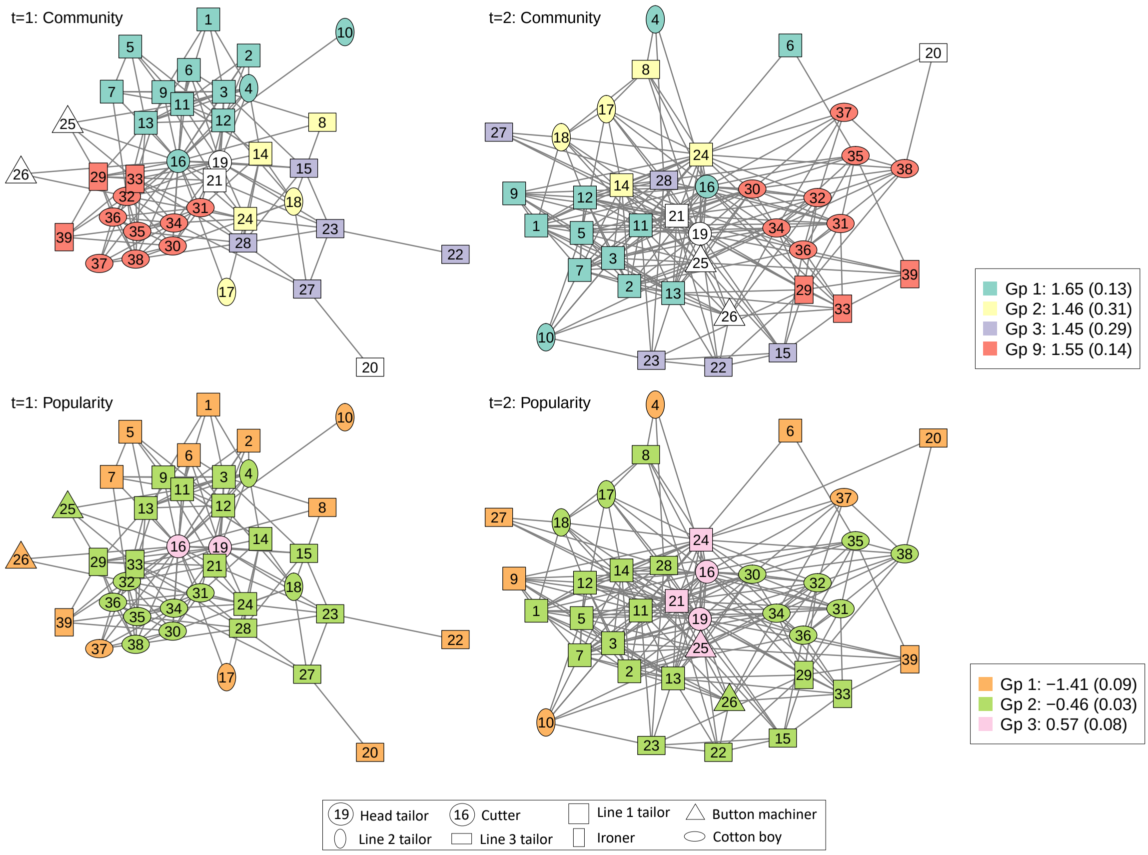


Figure 8: Fit of dynamic model I to Kapferer social networks.

conversation, the sharing of gossip and the enjoyment of a drink together", Kapferer, 1972). In particular, groups 1 and 2 consists of workers with jobs perceived to be of higher prestige: cutter, line 1 and line 2 tailors, group 3 consists of line 3 tailors and group 9 consists of all the ironers and cotton boys. The estimates of $\beta_k^*$ are strongly positive, indicating a high interaction rate within each group.

There are three popularity clusters with increasing means, $-1.41$ (group 1), $-0.46$ (group 2) and 0.57 (group 3). Thus, we can consider the three clusters as representing "low", "average" and "high" popularity. Actors 19 (head tailor) and 16 (cutter) are the only two actors with high popularity at $t = 1$, and they maintained high popularity at $t = 2$. This is not surprising since they are regarded by Kapferer

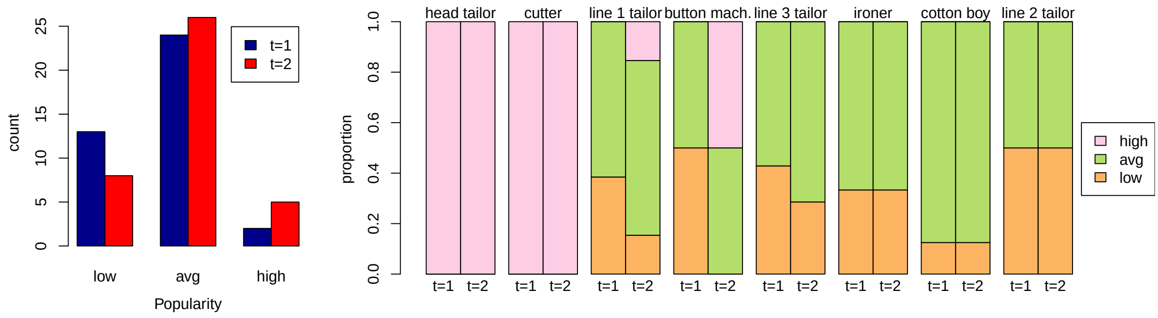


Figure 9: Barplot (left) shows the number of workers in each popularity cluster at the two time points. Barplot (right) shows the proportion of workers in each popularity cluster at each time for each job category.

(1972) to be in "supervisory" positions and play critical roles in the operation of the factory. From the barplot (left) in Figure 9, the number of workers with low popularity decreased from $t = 1$ to $t = 2$ while the number with average or high popularity increased. This reflects the efforts of the workers in expanding social ties after the first unsuccessful strike. Examining the results more closely using the barplot (right) in Figure 9, the proportion of workers with low and average popularity actually remained unchanged over the two time points for the ironers, cotton boys and line 2 tailors (positions with lower prestige). Changes in popularity arise mainly from line 1 tailors, button machiners and line 3 tailors. In particular, two line 1 tailors, $\{21, 24\}$ and a button machiner $\{25\}$ moved from average to high popularity. These observations are consistent with the analysis of Kapferer (1972), who noted that line 1 tailors made a strong attempt to expand their links after the first unsuccessful strike as they stand to benefit the most from the change in wage system. Mitchell (1989) also noted that the button machiner, Meshak (actor 25) played a crucial role in the unfolding events at the factory and was regarded as a supervisor by the factory owner

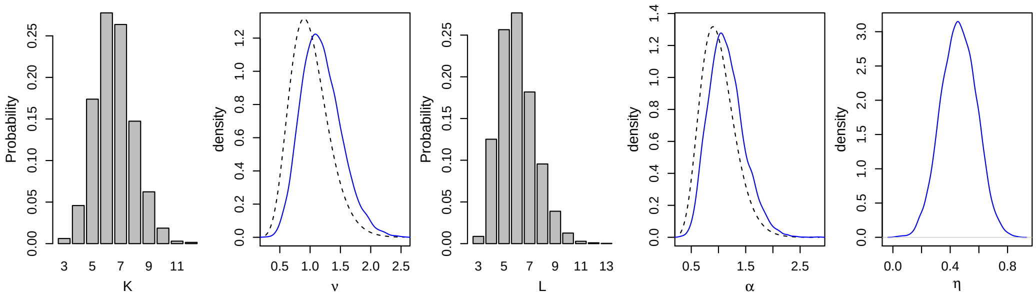


Figure 10: Posterior distributions of $K$, $\nu$, $L$, $\alpha$ and $\eta$. For $\nu$ and $\alpha$, prior distributions are shown in dotted lines and the posterior distributions in solid (blue) lines.

during the final period.

### *4.2.2 Dynamic model II*

Fitting dynamic model II using Algorithm 3 took 106 seconds. In this model, the parameter $\eta$ provides an indication of the persistence of ties. The probability that a tie is formed at any time point depends on whether a tie exists at the previous time point as well as the community membership of the nodes and their popularities. Figure 10 shows the posterior distributions of $K$, $\nu$, $L$, $\alpha$ and $\eta$ based on 6000 MCMC samples. The modes of $K$ and $L$ are both 6. The posterior mean of $\eta$ is 0.58 and its posterior mass is concentrated on positive values. This indicates that a tie is likely to persist at the second time point given that it existed at the first time point. Figure 11 shows the posterior similarity matrices. We note that the block structures are not clear-cut.

Figure 12 shows the posterior estimate of the cluster assignment obtained by minimizing Binder's loss function and the estimates of $\beta^*$ and $\theta^*$ for these clusterings. The communities detected are largely similar to that of dynamic model I except for changes to the assignment of individuals $\{14, 16, 19, 21\}$. A new "community" con-

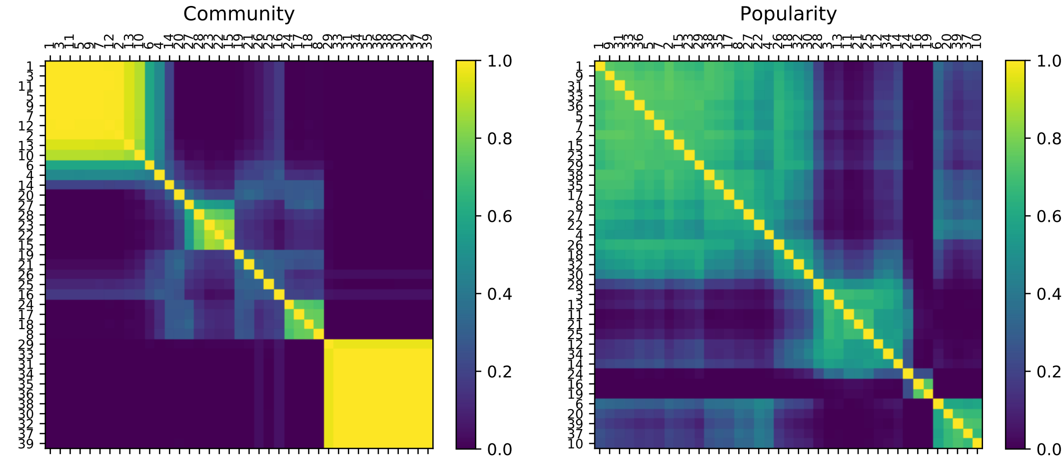


Figure 11: Posterior similarity matrices for community (left) and popularity (right).

sisting of $\{19, 21\}$ is present. However, the $\beta_k^*$ estimate for this group is 0.54 with a large standard deviation of 0.8. Thus, this is not truly a "community" in the sense that there is a high interaction rate between the actors.

The number of popularity clusters increased from three in dynamic model I to six in model II. In model II, the popularity of an actor summarizes his activity level across all time points. Figure 13 shows how the mean of $\theta_i$ varies with the degree of an actor at each time point. The head tailor and the cutter have significantly higher popularity than the other workers, followed by actor 24 (Ibrahim) and actors in popularity group 2. We note that group 2 includes several individuals who play significant roles in the factory's social relationships (Kapferer, 1972). These include Lyashi (11), who tried to win followers in support of his view of the factory structure, Hastings (13), who took on many supervisory duties for the cutter category at time 2, Meshak (25), who was regarded as a leader by the factory owner, and Mubanga (34), an influential figure among unskilled workers.

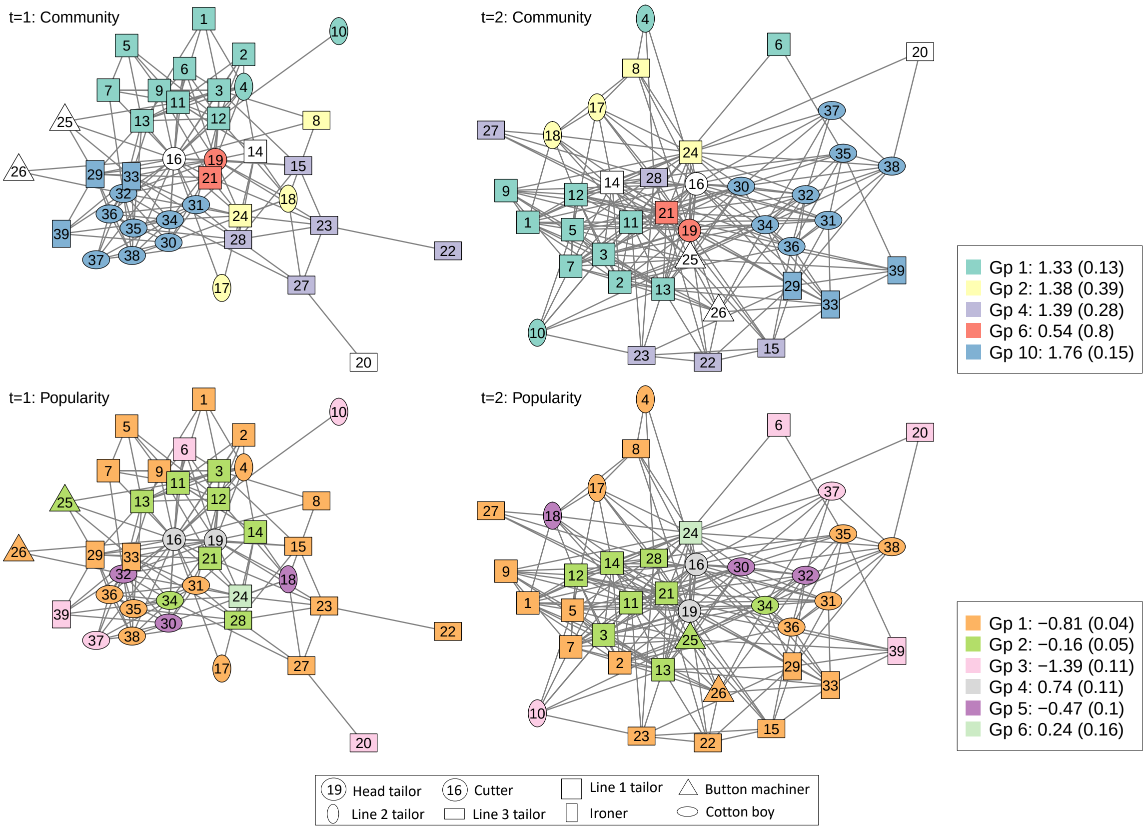


Figure 12: Fit of dynamic model II to Kapferer social networks.

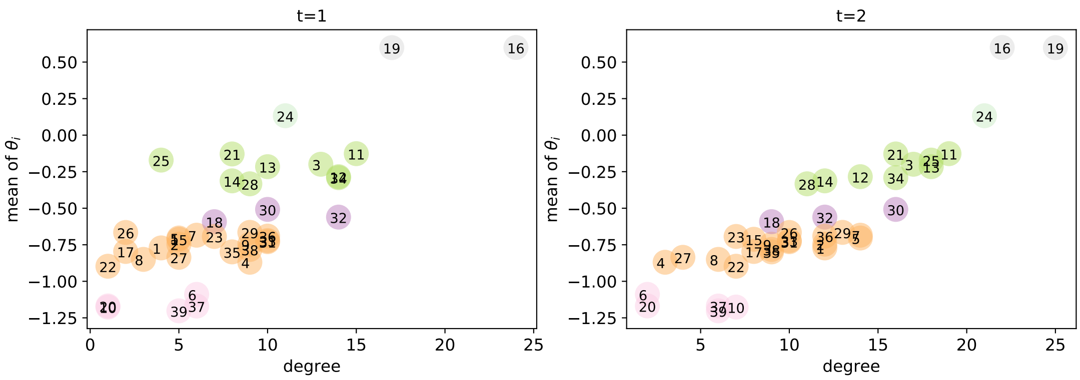


Figure 13: Plot of mean of $\theta_i$ against the actor $i$'s degree at $t = 1$ (left) and $t = 2$ (right).

# 5 Conclusion and future work

We propose a nonparametric Bayesian approach for detecting communities in social networks, using degree-corrected stochastic blockmodels. In the proposed static model, the number of communities and popularity clusters do not have to be fixed in advance and are inferred from the data automatically through the use of the DP. For the karate club network and the dolphins social network, we find that the static model returns sensible results although there is some sensitivity to the DP concentration parameters. The inferred popularity clusters also summarizes the popularities of the actors and helps in the identification of key players in the network. We discuss two extensions of the static model to dynamic networks. Dynamic model I enables the study of the change in activity level of actors over the entire duration while dynamic model II provides a measure of the persistence of links formed in the network. While the Gibbs samplers are feasible for small networks, they do not scale well to large networks and more efficient methods of estimation, such as variational approximation methods, need to be developed and will be object of future investigation.

# Acknowledgements

Linda Tan's research is supported by the NUS overseas postdoctoral fellowship and the start-up grant (R-155-000-190-133).

# References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, **9**, 1981–2014.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *Ann. Statist.*, **2**, 1152–1174.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, **106**, 21068–21073.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, **65**, 31–38.

Cauchemez, S., Bhattarai, A., Marchbanks, T. L., Fagan, R. P., Ostroff, S., Ferguson, N. M., Swerdlow, D., and the Pennsylvania H1N1 working group (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences*, **108**, 2825–2830.

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

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

Fienberg, S. E. and Wasserman, S. S. (1981). Categorical data analysis of single sociometric relations. *Sociological Methodology*, **12**, 156–192.

Fritsch, A. and Ickstadt, K. (2009). Improved criteria for clustering based on the posterior similarity matrix. *Bayesian Anal.*, **4**, 367–391.

Ghosh, P., Gill, P., Muthukumarana, S., and Swartz, T. (2010). A semiparametric bayesian approach to network modelling using dirichlet process prior distributions. *Australian & New Zealand Journal of Statistics*, **52**, 289–302.

Gill, P. S. and Swartz, T. B. (2007). Bayesian analysis of dyadic data. *American Journal of Mathematical and Management Sciences*, **27**, 73–92.

Gopalan, P. K., Wang, C., and Blei, D. (2013). Modeling overlapping communities with node popularities. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 2850–2858. Curran Associates, Inc.

Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, **19**, 293–317.

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

Herlau, T., Schmidt, M. N., and Mørup, M. (2014). Infinite-degree-corrected stochastic block model. *Phys. Rev. E*, **90**, 032819.

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

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, **14**, 1303–1347.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, **76**, 33–50.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, **5**, 109–137.

Ishwaran, H. and Zarepour, M. (2000). Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika*, **87**, 371.

Jara, A., García-Zattera, M. J., and Lesaffre, E. (2007). A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics & Data Analysis*, **51**, 5402–5415.

Kapferer, B. (1972). *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*. Manchester University Press.

Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, **83**, 016107.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*, pages 381–388. AAAI Press.

Latouche, P., Birmelé, E., and Ambroise, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *Ann. Appl. Stat.*, **5**, 309–336.

Liu, J. S. (1996). Nonparametric hierarchical bayes via sequential imputations. *The Annals of Statistics*, pages 911–930.

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

Lusseau, D., Schneider, K., Boisseau, O., Haase, P., Slooten, E., and Dawson, S. (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, **54**, 396–405.

Mitchell, J. C. (1989). Algorithms and network analysis: A test of some analytical procedures on kapferer's tailor shop material. In Freeman, L. C., White, D. R., and Romney, A. K., editors, *Research Methods in Social Network Analysis*, pages 365–391. George Mason University Press.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

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

Peng, L. and Carvalho, L. (2016). Bayesian degree-corrected stochastic blockmodels for community detection. *Electron. J. Statist.*, **10**, 2746–2779.

Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

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

Sewell, D. K. and Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Anal.*, **12**, 351–377.

Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, **14**, 75–100.

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

Thiemichen, S., Friel, N., Caimo, A., and Kauermann, G. (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks*, **46**, 11 – 28.

Vu, D. Q., Hunter, D. R., and Schweinberger, M. (2013). Model-based clustering of large networks. *Ann. Appl. Stat.*, **7**, 1010–1039.

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

Westerman, D., Spence, P. R., and Van Der Heide, B. (2014). Social media as information source: Recency of updates and credibility of information. *Journal of Computer-Mediated Communication*, **19**, 171–183.

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

Williamson, S. A. (2016). Nonparametric network models for link prediction. *Journal of Machine Learning Research*, **17**, 1–21.

Xing, E. P., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.*, **4**, 535–566.

Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, **33**, 452–473.

Initialize $z$, $c$, $\theta^*$ and $\beta^*$ and cycle through the following updates:

1. For $(i,j) \in \mathcal{S}$, draw $\zeta_{ij}$ from $\mathrm{TN}(\zeta_{ij}|\mu_{ij},1,0,\infty)$ if $y_{ij}=1$ and $\mathrm{TN}(\zeta_{ij}|\mu_{ij},1,-\infty,0)$ if $y_{ij}=0$, where $\mu_{ij}=\theta^*_{c_i}+\theta^*_{c_j}+Z_{ij}^T\beta^*$.

2. For $i=1,\ldots,n$: If $m_{-i,z_i}=0$, remove $\beta^*_{z_i}$ from $\beta^*$. Draw $z_i$ according to (3.2):

$$\begin{aligned} \mathrm{P}(z_i=k|\text{rest}) &= a m_{-i,k} \exp\left\{\beta^*_k \sum_{j\neq i:\, z_j=k} (\zeta_{ij}-\theta^*_{c_i}-\theta^*_{c_j}) - \frac{m_{-i,k}}{2}\beta^{*2}_k\right\} \\ &\text{for } k\in z_{-i} \text{ and } \mathrm{P}(z_i\neq z_j \text{ for all } j\neq i|\text{rest}) = a\nu, \end{aligned} \tag{3.2}$$

where $a$ is a normalizing constant that ensures the above probabilities sum to one. If the value of $z_i$ is not in $z_{-i}$, draw $\beta^*_{z_i}\sim \mathrm{N}(0,\sigma^2_\beta)$ and add it to $\beta^*$.

3. Draw $\beta^*\sim\mathrm{N}(P^{-1}\sum_{i<j}(\zeta_{ij}-\theta^*_{c_i}-\theta^*_{c_j})Z_{ij}, P^{-1})$, where $P=\frac{1}{\sigma^2_\beta}I_K+Z^TZ$.

4. Draw $\gamma_1\sim\mathrm{Beta}(\alpha+1,n)$. Then draw $\alpha$ from the mixture: $\pi_\alpha\mathrm{Gamma}(a_\alpha+L, b_\alpha-\log\gamma_1)+(1-\pi_\alpha)\mathrm{Gamma}(a_\alpha+L-1,b_\alpha-\log\gamma_1)$, where $\frac{\pi_\alpha}{1-\pi_\alpha}=\frac{a_\alpha+L-1}{n(b_\alpha-\log\gamma_1)}$.

5. For $i=1,\ldots,n$: If $n_{-i,c_i}=0$, remove $\theta^*_{c_i}$ from $\theta^*$. Draw $c_i$ according to (3.3):

$$\begin{aligned} \mathrm{P}(c_i=\ell|\text{rest}) &= b n_{-i,\ell}\exp\left\{\theta^*_\ell\sum_{j\neq i}(\zeta_{ij}-\theta^*_{c_j}-Z^T_{ij}\beta^*) - \frac{n-1}{2}\theta^{*2}_\ell\right\} \\ &\text{for } \ell\in c_{-i} \text{ and } \mathrm{P}(c_i\neq c_j \text{ for all } j\neq i|\text{rest}) = b\alpha\frac{\sigma_c}{\sigma_\theta}\exp\left\{\frac{\mu^2_{c_i}}{2\sigma^2_c}\right\}, \end{aligned} \tag{3.3}$$

where $\sigma^2_c=\left(n-1+\frac{1}{\sigma^2_\theta}\right)^{-1}$, $\mu_{c_i}=\sigma^2_c\sum_{j\neq i}(\zeta_{ij}-\theta^*_{c_j}-Z^T_{ij}\beta^*)$, and $a$ is a normalizing constants that ensure the above probabilities sum to one. If the value of $c_i$ is not in $c_{-i}$, draw $\theta^*_{c_i}\sim\mathrm{N}(\mu_{c_i},\sigma^2_c)$ and add it to $\theta^*$.

6. For $m=1,\ldots,L$, draw $\theta^*_m\sim N(\mu_m,\sigma^2_m)$, where $\sigma^2_m=\left(\frac{1}{\sigma^2_\theta}+\sum_{\mathcal{S}_m}4+\sum_{\mathcal{P}_m}1\right)^{-1}$ and $\mu_m=\sigma^2_m\left[2\sum_{\mathcal{S}_m}(\zeta_{ij}-Z^T_{ij}\beta^*)+\sum_{\mathcal{P}_m}(\zeta_{ij}-\theta^*_{c_j}-Z^T_{ij}\beta^*)\right]$.

7. Draw $\gamma_2\sim\mathrm{Beta}(\nu+1,n)$. Then draw $\nu$ from the mixture: $\pi_\nu\mathrm{Gamma}(a_\nu+K,b_\nu-\log\gamma_2)+(1-\pi_\nu)\mathrm{Gamma}(a_\nu+K-1,b_\nu-\log\gamma_2)$, where $\frac{\pi_\nu}{1-\pi_\nu}=\frac{a_\nu+K-1}{n(b_\nu-\log\eta)}$.

Algorithm 1: Gibbs sampler for static model.

Initialize $z$, $c$, $\theta^*$ and $\beta^*$ and cycle through the following updates:

1. For $t = 1, \dots, T$, $(i, j) \in \mathcal{S}$, draw $\zeta_{t,ij}$ from $\text{TN}(\zeta_{t,ij}|\mu_{t,ij}, 1, 0, \infty)$ if $y_{t,ij} = 1$ and $\text{TN}(\zeta_{t,ij}|\mu_{t,ij}, 1, -\infty, 0)$ if $y_{t,ij} = 0$, where $\mu_{t,ij} = \theta^*_{c_{it}} + \theta^*_{c_{jt}} + Z_{ij}^T \beta^*$.

2. For $i = 1, \dots, n$: If $m_{-i,z_i} = 0$, remove $\beta^*_{z_i}$ from $\beta^*$. Draw $z_i$ according to (3.4):

$$\text{P}(z_i = k|\text{rest}) = a m_{-i,k} \exp\left\{\beta^*_k \sum_{j \neq i: z_j = k} \sum_t (\zeta_{t,ij} - \theta^*_{c_{it}} - \theta^*_{c_{jt}}) - \tfrac{T m_{-i,k}}{2} \beta^{*2}_k\right\}$$
$$\text{for } k \in z_{-i} \text{ and } \text{P}(z_i \neq z_j \text{ for all } j \neq i|\text{rest}) = a\nu, \tag{3.4}$$

   where $a$ is a normalizing constant that ensures the above probabilities sum to one. If the value of $z_i$ is not in $z_{-i}$, draw $\beta^*_{z_i} \sim \text{N}(0, \sigma^2_\beta)$ and add it to $\beta^*$.

3. Draw $\beta^* \sim \text{N}(P^{-1} \sum_{i<j} Z_{ij} \sum_t (\zeta_{t,ij} - \theta^*_{c_{it}} - \theta^*_{c_{jt}}), P^{-1})$, where $P = \frac{1}{\sigma^2_\beta} I_K + Z^T Z$.

4. As in Step 4 of Algorithm 1.

5. For $t = 1, \dots, T$, $i = 1, \dots, n$: If $n_{-it,c_{it}} = 0$, remove $\theta^*_{c_{it}}$ from $\theta^*$. Draw $c_{it}$ according to (3.5):

$$\text{P}(c_{it} = \ell|\text{rest}) = b n_{-it,\ell} \exp\left\{\theta^*_\ell \sum_{j \neq i} (\zeta_{t,ij} - \theta^*_{c_{jt}} - Z_{ij}^T \beta^*) - \tfrac{n-1}{2} \theta^{*2}_\ell\right\}$$
$$\text{for } \ell \in c_{-i} \text{ and } \text{P}(c_{it} \neq \text{any value in } c_{it}|\text{rest}) = b\alpha \frac{\sigma_c}{\sigma_\theta} \exp\left\{\frac{\mu^2_{c_{it}}}{2\sigma^2_c}\right\}, \tag{3.5}$$

   where $\sigma^2_c = \left(n - 1 + \frac{1}{\sigma^2_\theta}\right)^{-1}$ and $\mu_{c_{it}} = \sigma^2_c \sum_{j \neq i} (\zeta_{t,ij} - \theta^*_{c_{jt}} - Z_{ij}^T \beta^*)$, and $a$ is a normalizing constants that ensure the above probabilities sum to one. If the value of $c_{it}$ is not in $c_{-it}$, draw $\theta^*_{c_{it}} \sim \text{N}(\mu_{c_{it}}, \sigma^2_c)$ and add it to $\theta^*$.

6. For $m = 1, \dots, L$, draw $\theta^*_m \sim N(\mu_m, \sigma^2_m)$, where $\sigma^2_m = \left(\frac{1}{\sigma^2_\theta} + \sum_t \sum_{\mathcal{S}_{t,m}} 4 + \sum_t \sum_{\mathcal{P}_{t,m}} 1\right)^{-1}$, $\mu_m = \sigma^2_m \big[2 \sum_t \sum_{\mathcal{S}_{t,m}} (\zeta_{t,ij} - Z_{ij}^T \beta^*) + \sum_t \sum_{\mathcal{P}_{t,m}} (\zeta_{t,ij} - \theta^*_{c_{jt}} - Z_{ij}^T \beta^*)\big]$.

7. As in Step 7 of Algorithm 1.

Algorithm 2: Gibbs sampler for dynamic model I.

Initialize $z$, $c$, $\theta^*$ and $\beta^*$ and cycle through the following updates:

1. For $t = 1, \dots, T$, $(i, j) \in \mathcal{S}$, draw $\zeta_{t,ij}$ from $\text{TN}(\zeta_{t,ij}|\mu_{t,ij}, 1, 0, \infty)$ if $y_{t,ij} = 1$ and $\text{TN}(\zeta_{t,ij}|\mu_{t,ij}, 1, -\infty, 0)$ if $y_{t,ij} = 0$, where $\mu_{t,ij} = \eta y_{t-1,ij}\mathbb{1}\{t > 1\} + \theta^*_{c_i} + \theta^*_{c_j} + Z_{ij}^T\beta^*$.

2. For $i = 1, \dots, n$: If $m_{-i,z_i} = 0$, remove $\beta^*_{z_i}$ from $\beta^*$. Draw $z_i$ according to (3.6):
$$\text{P}(z_i = k|\text{rest}) = am_{-i,k}\exp\left\{\beta^*_k \sum_t \sum_{j\neq i: z_j = k}(\tilde{\zeta}_{t,ij} - \theta^*_{c_i} - \theta^*_{c_j}), -\frac{Tm_{-i,k}}{2}\beta^{*2}_k\right\}$$
$$\text{for } k \in z_{-i} \text{ and } \text{P}(z_i \neq z_j \text{ for all } j \neq i|\text{rest}) = a\nu, \tag{3.6}$$
where $b$ is a normalizing constant that ensures the above probabilities sum to 1. If the value of $z_i$ is not in $z_{-i}$, draw $\beta^*_{z_i} \sim \text{N}(0, \sigma^2_\beta)$ and add it to $\beta^*$.

3. Draw $\beta^* \sim \text{N}(P^{-1}\sum_{i<j} Z_{ij}\sum_t(\tilde{\zeta}_{t,ij} - \theta^*_{c_i} - \theta^*_{c_j}), P^{-1})$, where $P = \frac{1}{\sigma^2_\beta}I_K + TZ^TZ$.

4. As in Step 4 of Algorithm 1.

5. For $i = 1, \dots, n$: If $n_{-i,c_i} = 0$, remove $\theta^*_{c_i}$ from $\theta^*$. Draw $c_i$ according to (3.7):
$$\text{P}(c_i = \ell|\text{rest}) = bn_{-i,\ell}\exp\left\{\theta^*_\ell \sum_t \sum_{j\neq i}(\tilde{\zeta}_{t,ij} - \theta^*_{c_j} - Z_{ij}^T\beta^*) - \tfrac{T(n-1)}{2}\theta^{*2}_\ell\right\}.$$
$$\text{for } \ell \in c_{-i} \text{ and } \text{P}(c_i \neq c_j \text{ for all } j \neq i|\text{rest}) = b\alpha\frac{\sigma_c}{\sigma_\theta}\exp\left\{\frac{\mu^2_{c_i}}{2\sigma^2_c}\right\}, \tag{3.7}$$
where $\sigma^2_c = \left(T(n-1) + \frac{1}{\sigma^2_\theta}\right)^{-1}$, $\mu_{c_i} = \sigma^2_c\sum_t\sum_{j\neq i}(\tilde{\zeta}_{t,ij} - \theta^*_{c_j} - Z^T_{ij}\beta^*)$ and $b$ is a normalizing constant that ensures the above probabilities sum to 1. If the value of $c_i$ is not in $c_{-i}$, draw $\theta^*_{c_i} \sim \text{N}(\mu_{c_i}, \sigma^2_c)$ and add it to $\theta^*$.

6. For $m = 1, \dots, L$, draw $\theta^*_m \sim \text{N}(\mu_m, \sigma^2_m)$, where $\sigma^2_m = \left(\frac{1}{\sigma^2_\theta} + \sum_{\mathcal{S}_m} 4T + \sum_{\mathcal{P}_m} T\right)^{-1}$ and $\mu_m = \sigma^2_m\left(2\sum_{\mathcal{S}_m}(\tilde{\zeta}_{t,ij} - Z^T_{ij}\beta^*) + \sum_{\mathcal{P}_m}(\tilde{\zeta}_{t,ij} - \theta^*_{c_j} - Z^T_{ij}\beta^*)\right)$.

7. As in Step 7 of Algorithm 1

8. Draw $\eta \sim \text{N}(\mu_\eta, \sigma^2_{\eta,1})$, where $\sigma^2_{\eta,1} = \left(\frac{1}{\sigma^2_\eta} + \sum_{t\geq 2}\sum_{i<j} y^2_{t-1,ij}\right)^{-1}$ and $\mu_\eta = \sigma^2_{\eta,1}\sum_{t\geq 2}\sum_{i<j} y_{t-1,ij}(\zeta_{t,ij} - \theta^*_{c_i} - \theta^*_{c_j} - Z^T_{ij}\beta^*)$.

Algorithm 3: Gibbs sampler for dynamic model II.