

4

Hierarchical Mixed Membership Stochastic Blockmodels for Multiple Networks and Experimental Interventions

Tracy M. Sweet

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Andrew C. Thomas

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Brian W. Junker

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

CONTENTS

4.1	Introduction	95
4.2	Modeling a Single Network	97
4.2.1	Single Network Model Estimation	99
4.2.2	Empirical Example	100
4.2.3	Incorporating Covariates into a MMSBM	101
4.3	Modeling an Ensemble of Networks	103
4.3.1	The Hierarchical Network Framework	103
4.3.2	The Hierarchical Mixed Membership Stochastic Blockmodel	103
4.3.2.1	Examples of HMMSBMs	104
4.3.3	Model Estimation	106
4.3.4	Empirical Examples	107
4.3.5	HMMSBM Extension: Sensitivity Analysis	114
4.4	Discussion	117

4.1 Introduction

A social network represents the relationships among a group of individuals or entities and is commonly illustrated by a graph. The nodes or vertices represent individuals or actors and the edges between them the ties or relationships between two individuals. These edges may be directed, suggesting a sender and receiver of the interaction, or undirected, suggesting reciprocity between the two nodes; a network depicting collaboration is likely to be an undirected graph whereas a network depict-

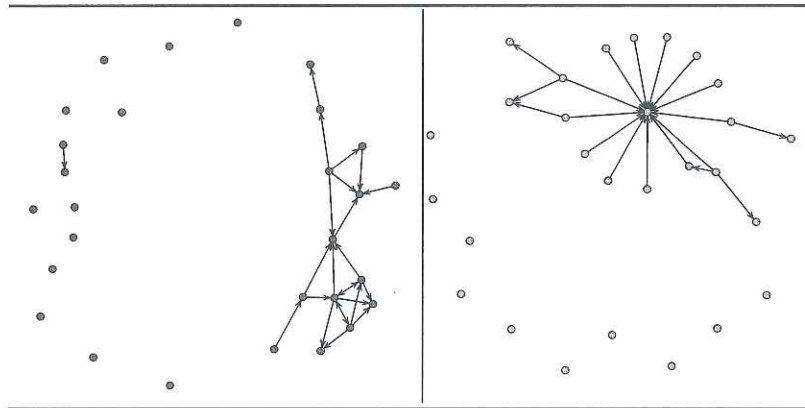


FIGURE 4.1

Two social networks, depicting asymmetric advice-seeking behavior among two groups of teachers, from Pitts and Spillane (2009). Vertices, or nodes, represent individual teachers. Arrows, or directed edges, point from advice seeking teachers to advice providing teachers.

ing advice-seeking would be a directed graph. Figure 4.1 shows advice-seeking ties among teachers regarding two different subjects. Network ties are part of a larger class of observations termed relational data since these data reflect pairwise relationships, such as the presence and direction of pairwise ties. Since relationships are pervasive, it is unsurprising that relational data methodology has applications in a wide variety of fields, including biology (Airoldi et al., 2005), international relations (Hoff and Ward, 2004), education (Weinbaum et al., 2008), sociology (Goodreau et al., 2009), and organizational theory (Krackhardt and Handcock, 2007).

Two prominent quantitative methods for analyzing social networks are descriptive network statistics and statistical modeling. Descriptive network statistics are useful for exploring, summarizing and identifying certain features of networks, which are then used as covariates in other statistical models. Common statistics include density, the total number of ties; degree, the number of ties for any one node; betweenness, the extent that a node connects other nodes; and other observed structural elements such as triangles. Kolaczyk (2009) provides a comprehensive list. Descriptive statistics are inherently aggregate, so using them to represent a network or to compare networks is problematic. For example, Figure 4.1 show two networks with similar density (22 ties among 27 nodes and 19 ties among 28 nodes); however, the structure of the networks is quite different.

Alternatively, a statistical social network model formalizes the probability of

1cm

↑

* put abbr. in after first mention

* observing the entire network and its various structural features. Current methods generally fall into one of three categories, Exponential Random Graph Models, Latent Space Models and Mixed Membership Stochastic Blockmodels; see Goldenberg et al. (2009) for a comprehensive review. An Exponential Random Graph Model (ERGM; Wasserman and Pattison, 1996) represents the probability of observing a particular network as a function of network statistics. The Latent Space Model (LSM; Hoff et al., 2002) assumes each node occupies a position in a latent social space. The probability of a tie between two individuals is modeled as a function of the pairwise distance in this space. Stochastic Blockmodels cluster nodes to one of a fixed number of finite groups, and the probability of a tie between two nodes is determined by the group membership of each node. The Mixed Membership Stochastic Blockmodel (MMSBM; Airoldi et al., 2008) allows nodes to belong to multiple groups so that group membership may vary by node interaction.

(ERGM)
↓ (LSM)
(MMSB)

i.e.

Most modeling methodology for social network focuses on modeling a single network, but in many applications, more than one network may be of interest. The study of multiple networks can be divided into three classes: studying multiple types of ties among nodes of one network (e.g. friendship ties and collaboration ties), studying one network over time, and studying a single measure on multiple isolated networks. There has been a fair amount of work done for the first two cases. Fienberg et al. (1985) showed how loglinear models can be used to model multiple measures on a single network and Pattison and Wasserman (1999) extended this work for the logit forms of p^* models. Longitudinal methods to model a single network over time have been extensively studied. The three categories of models each have known longitudinal extensions: Hanneke et al. (2010) introduce temporal ERGMs which are based on a discrete Markov process; Westveld and Hoff (2011) embed an autoregressive structure in LSMs; and Xing et al. (2010) add a state space model to the MMSBM.

S no comma

↓

Modeling a sample of isolated networks has only recently attracted sustained attention. Motivated by social networks of teachers in education research, Sweet et al. (2012) introduced Hierarchical Network Models (HNM), a class of models for modeling ensembles of networks. The purpose of this paper is to use the HNM framework to formally introduce hierarchical mixed membership stochastic blockmodels (HMMSBM) which extend the MMSBM for use with relational data from multiple isolated networks.

u.c. x 5

In the next section, we formally define the MMSBM for a single network, present a covariate version of a MMSBM and introduce an MCMC algorithm for estimation. In Section 4.3, we present the HNM framework and formally define the HMMSBM. Extending our MCMC algorithm for a single network, we present an algorithm for fitting the HMMSBM that we illustrate with two examples. We conduct a simple simulation study for sensitivity analysis, and conclude with some remarks regarding estimation and utility of these models.

↓

4.2 Modeling a Single Network

A single social network Y among n individuals can be represented by an adjacency matrix of dimension $n \times n$,

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nn} \end{bmatrix}, \quad (4.1)$$

where Y_{ij} is the value of the tie from i to j . These ties might be binary, indicating the presence or absence of a tie, or an integer or real number, the frequency of interaction or strength of a tie. For the purposes of this paper, we restrict ourselves to binary ties.

In many contexts, individuals in the network belong to certain subgroups. In a school faculty network, for example, teachers belong to departments. However, these group memberships are often not directly observed and can only be inferred through the network structure. Figure 4.2 (left) shows an adjacency matrix for networks generated from a stochastic blockmodel in which individuals belong to one of four groups. A black square indicates the presence of a tie between two individuals. Ties within groups are much more likely than ties across groups. Block models are most appropriate for relational data with this structure and a variety of block models have been studied (see Anderson and Wasserman, 1992).

Stochastic blockmodels assign each individual membership to a block or group, and assignment may either be observed or latent. Tie probabilities are then determined based on group membership; often within group probabilities ties to be much higher than across group probabilities, resulting in the block structure shown on the left in Figure 4.2.

Mixed membership stochastic blockmodels (MMSBM) instead allow block membership to be defined for each interaction with a new partner. Rather than assuming individual i is a member of block k for all interactions, the block membership is determined anew for each interaction. Individual i might belong to block k when interacting with individual j but belong to block k' when interacting with individual j' .

We define the MMSBM as a hierarchical Bayesian model (Airoldi et al., 2008),

$$\begin{aligned} Y_{ij} &\sim \text{Bernoulli}(S_{ij}^T B R_{ji}) \\ S_{ij} &\sim \text{Multinomial}(1, \theta_i) \\ R_{ji} &\sim \text{Multinomial}(1, \theta_j) \\ \theta_i &\sim \text{Dirichlet}(\lambda) \\ B_{\ell m} &\sim \text{Beta}(a_{\ell m}, b_{\ell m}), \end{aligned} \quad (4.2)$$

where the group membership probability vector for individual i is θ_i , and specific group memberships are determined through a multinomial distribution. S_{ij} is the

indicating

check }

tend?

v.c. x 4

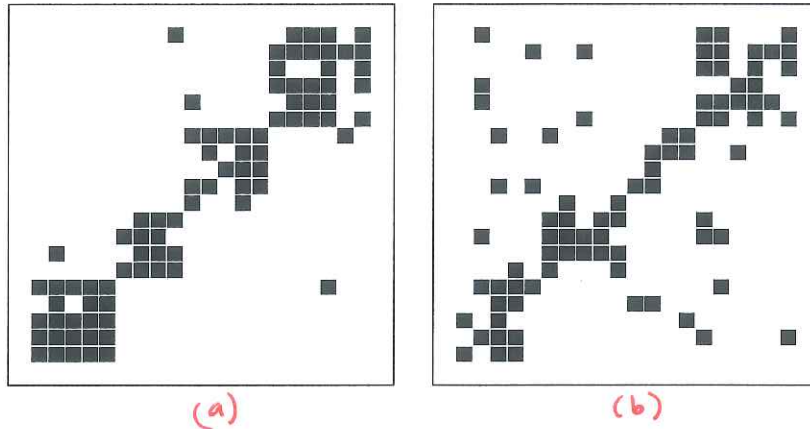


FIGURE 4.2

(a) A network generated from a stochastic blockmodel where group membership is not mixed (left). Each node is assigned a group membership which determines the probability of ties. A network generated from a MMSBM (right). Node membership may vary with each pairwise interaction. Note, the $n \times n$ sociomatrix displays a black box for each tie and white otherwise. (b)

(a) group membership indicator vector of i when initiating interaction with j and R_{ji} is the group membership indicator vector of j when acting in response to i . Notice the stochastic nature S_{ij} and R_{ji} ; each is sampled for every interaction from i to j , allowing individual group memberships to vary. The ties Y_{ij} are determined based on a block dependent probability matrix B , where $B_{\ell m}$ is the probability of a tie from an individual in group ℓ to an individual in group m . (b)

The hyperparameter λ may be fixed and known or estimated as a parameter. The dimension of λ identifies the number of groups (g) and the value of λ determines the shape of the Dirichlet distribution on the g -simplex. The hyperparameters (a_ℓ, b_m) are generally elicited so that within-group tie probabilities are higher than across-group tie probabilities.

4.2.1 Single Network Model Estimation

We developed a Markov Chain Monte Carlo (MCMC; Gelman et al., 2004) algorithm to fit the MMSBM. The joint likelihood of the model can be written as the following product:

$$P(Y|S, R, \theta, \lambda, B)P(S|\theta)P(R|\theta)P(\theta|\lambda)P(B)P(\lambda) = \prod_{i \neq j} P(Y_{ij}|S_{ij}, R_{ji}, \theta_i, \theta_j, \lambda, B) \prod_{i \neq j} P(S_{ij}|\theta_i)P(R_{ji}|\theta_j) \prod_i P(\theta_i|\lambda) \prod_{\ell, m} P(B)P(\lambda). \quad (4.3)$$

The complete conditionals for θ, R, S, B can be written in a closed form, so we

I had to make this smaller to fit.
I'm working to make this big again.

use Gibbs updates for each. Full conditional posterior probability distributions are listed below. Define \dots to represent all other parameters and data in the model, and let ℓ^* represent the group indicated by R_{ji} and m^* represent the group indicated by S_{ij} .

$$\begin{aligned}
 P(\theta_i | \dots) &\propto \text{Dirichlet}(\lambda + \sum_j S_{ij} + \sum_j R_{ij}) \\
 P(S_{ij} | \dots) &\propto \text{Multinomial}(p) \\
 p_k &= \theta_{ik} B_{k\ell^*}^{Y_{ij}} (1 - B_{k\ell^*})^{(1-Y_{ij})} \\
 P(R_{ji} | \dots) &\propto \text{Multinomial}(q) \\
 q_k &= \theta_{ik} B_{m^*k}^{Y_{ij}} (1 - B_{m^*k})^{(1-Y_{ij})} \\
 P(B_{\ell m} | \dots) &\propto \text{Beta}(a_\ell + \sum_{(ij)^*} Y_{ij}, b_m + \sum_{(ij)^*} Y_{ij}), \quad (4.4)
 \end{aligned}$$

(15) where $(ij)^*$ be an (ℓ, m) specific subset of $i = 1, \dots, n$ and $j = 1, \dots, n$ such that $S_{ij} = \ell$ and $R_{ji} = m$. In addition, we incorporate a sparsity parameter ρ (Airoldi et al., 2008). The absence of ties can be attributed to either rarity of interaction across groups or lack of interest in making across group ties. For example, teachers in department in schools may have few collaborative ties outside of their department because they interact less often with teachers outside their department but also because they would rather interact with those who teach the same subjects. The sparsity parameter helps to account for sparsity in the adjacency matrix due to lack of interaction. The probability of tie from group ℓ to m is therefore modeled as $\rho B_{\ell m}$.

If λ is estimated, we use a common parameterization and let $\lambda = \gamma \xi$ where $\gamma = \sum_{k=1}^g \lambda$ and $\sum_{k=1}^g \xi = 1$ (Erosheva, 2003). We can think of γ as a measure of how extreme the Dirichlet distribution is, that is, small values of γ imply greater mass in the corners of the g -simplex. Since ξ sums to 1, it is an indirect measure of the probability of belonging to each group. Equal values of ξ suggest equal sized groups. As defined, γ and ξ are independent and we update each using Metropolis steps.

To update γ , we use a gamma proposal distribution with shape parameter ν_γ and rate parameter selected so that the proposal distribution has a mean at the current value of γ . The value of ν_γ is then tuned to ensure an appropriate acceptance rate. Then the proposed value of γ^{s+1} is accepted with probability $\min\{1, R\}$ where $R = \frac{P(\gamma^{(s+1)} | \dots) P(\gamma^{(s)} | \gamma^{(s+1)})}{P(\gamma^{(s)} | \dots) P(\gamma^{(s+1)} | \gamma^{(s)})}$.

To update ξ , we use a uniform Dirichlet proposal distribution centered at the current value of ξ . Thus, $\xi^{(s+1)} \sim \text{Dirichlet}(\nu_\xi g \xi^{(s)})$ where ν_ξ is the appropriate tuning parameter. The proposed value of $\xi^{(s+1)}$ is accepted with probability $\min\{1, R\}$ where now, $R = \frac{P(\xi^{(s+1)} | \dots) P(\xi^{(s)} | \xi^{(s+1)})}{P(\xi^{(s)} | \dots) P(\xi^{(s+1)} | \xi^{(s)})}$.

hyphen

teach

i.e.

no comma

4.2.2 Empirical Example

To illustrate fitting a single network MMSBM, we use the Monk data of Sampson (1968). While staying with a group of monks as a graduate student, Sampson recorded relational data among the monks at different time periods during his year-long stay. Toward the end of his stay, there was a political crisis which resulted in several monks being expelled and several others leaving.

5 The data we use include relational data from three time periods before the crisis. For each monk, we have their top three selections for whom they most like. These data have been aggregated into a single adjacency matrix, where $Y_{ij} = 1$ if monk i nominated j as one of his top three choices during any of the three time periods. Y_{ii} is undefined.

Based on past work suggesting three subgroups of Monks (Breiger et al., 1975), we fit the following MMSBM:

$$\begin{aligned}
 Y_{ij} &\sim \text{Bernoulli}(S_{ij}^T B R_{ji}) \\
 S_{ij} &\sim \text{Multinomial}(1, \theta_i) \\
 R_{ji} &\sim \text{Multinomial}(1, \theta_j) \\
 \theta_i &\sim \text{Dirichlet}(\gamma)\xi \\
 B_{\ell\ell} &\sim \text{Beta}(3, 1) \\
 B_{\ell m} &\sim \text{Beta}(1, 10), \ell \neq m \\
 \gamma &\sim \text{Gamma}(1, 5) \\
 \xi &\sim \text{Dirichlet}(1, 1, 1),
 \end{aligned}
 \tag{4.5}$$

where $1 - \rho = 1 - \frac{\sum_{ij} Y_{ij}}{N(N-1)}$, and $N = 18$, the number of monks.

5 We sample MCMC chains of length 15000, keeping the last 10000, and retaining 1 out of every 25 steps for a posterior approximation of 401 samples. To assess our fit, we compare the original sociomatrix shown on the left in Figure 4.3 to our fitted model. Using posterior means for each parameter, we illustrate the probability of a tie between two monks by color, with low probabilities in shades of blue and high probabilities in red, orange and yellow (right Figure 4.3).

5 x 2

blue

4.2.3 Incorporating Covariates into a MMSBM

While the MMSBM captures block structure, network ties may also form based on other individual similarity independent of or unrelated to the existing block structure. While teachers in schools may belong to departments, they may also group based on unobserved characteristics. But some ties might also form based on proximity in the school building, teaching the same group of students, or attending new teacher seminars together, independently of the overarching grouping mechanism.

Similarities

belong to groups

We present a simple extension for the MMSBM to include covariates as

new page

$$Y_{ij} \sim \text{Bernoulli}(p_{ij})$$

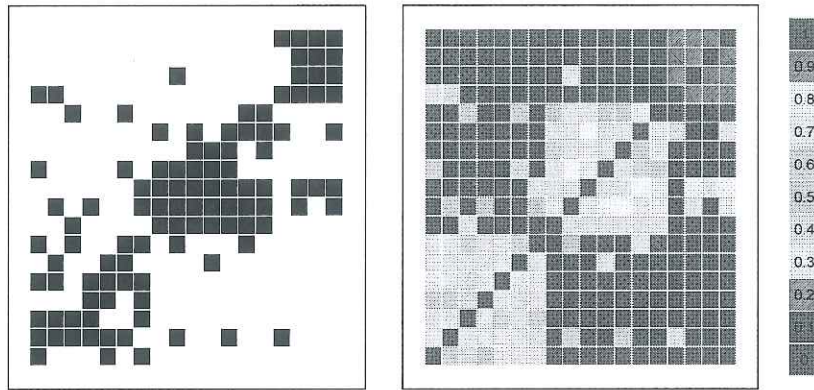


FIGURE 4.3

The original sociomatrix (left) versus the probability of a tie as determined by our model using posterior means (center). In general, estimated tie probabilities mirror the true tie structure. The legend shows increments of 0.1, with all values except 0 being the upper endpoint of the continuous class of colors (right).

$$\begin{aligned}
 p_{ij} &= \frac{\exp\{\text{logit}(S_{ij}^T \text{logit}(B)R_{ji}) + \alpha X_{ij}\}}{1 + \exp\{\text{logit}(S_{ij}^T BR_{ji}) + \alpha X_{ij}\}} \\
 S_{ij} &\sim \text{Multinomial}(1, \theta_i) \\
 R_{ji} &\sim \text{Multinomial}(1, \theta_j) \\
 \theta_i &\sim \text{Dirichlet}(\lambda) \\
 B_{\ell m} &\sim \text{Beta}(a_{\ell m}, b_{\ell m}) \quad \star
 \end{aligned} \tag{4.6}$$

Notice one comma

where X_{ij} is a covariate and α is the coefficient for that covariate.

Model (4.6) can be fit using a MCMC algorithm similar to (4.4), with a more complicated sampling distribution. We use the same Gibbs update for θ , and the same Metropolis updates for γ and ξ , as presented in our standard MMSBM (4.4). We use the following Gibbs updates for S_{ij} and R_{ji} ,

$$\begin{aligned}
 P(S_{ij} | \dots) &\propto \text{Multinomial}(p) \\
 p_k &= \theta_{ik} \frac{\exp\{\text{logit}(B)_{k\ell^*} + \alpha X_{ij}\}^{Y_{ij}}}{1 + \exp\{\text{logit}(B)_{k\ell^*} + \alpha X_{ij}\}} \\
 P(R_{ji} | \dots) &\propto \text{Multinomial}(q) \\
 q_k &= \theta_{jk} \frac{\exp\{\text{logit}(B)_{m^*k} + \alpha X_{ij}\}^{Y_{ij}}}{1 + \exp\{\text{logit}(B)_{m^*k} + \alpha X_{ij}\}}, \tag{4.7}
 \end{aligned}$$

where again ℓ is the group indicated by R_{ji} and m is the group indicated by S_{ij} . We

reparameterize B and use $\text{logit}(B)$ throughout our MCMC algorithm. The entries in B no longer have a direct sampling and we instead use Metropolis–Hastings updates.

To take advantage of random walk updates we reparameterize B as $\text{logit}(B)$, and having an unbounded support for our proposal distributions allows us to update the diagonal and off-diagonal elements using the same proposal distribution. Note that $S_{ij}^T \text{logit}(B) R_{ji}$ and $\text{logit}(S_{ij}^T B R_{ji})$ are equivalent.

Thus, to update an entry of $\text{logit}(B)$, $\text{logit}(B)_{\ell m}^s$, we propose a new entry, $\text{logit}(B)_{\ell m}^{s+1}$, using a normal random walk with mean $\text{logit}(B)_{\ell m}^s$ and variance is determined by a tuning parameter to ensure appropriate acceptance rates. The probability of accepting this new entry is $\min\{1, R\}$ where $R = \frac{P(\text{logit}(B)_{\ell m}^{s+1} | \dots)}{P(\text{logit}(B)_{\ell m}^s | \dots)}$. We illustrate this algorithm in Section 4.3.4.

where
 \int

4.3 Modeling an Ensemble of Networks

4.3.1 The Hierarchical Network Framework

Consider a collection of K networks $\mathbb{Y} = (Y_1, \dots, Y_K)$ where $Y_k = (Y_{11k}, \dots, Y_{n_k n_k k})$. The Hierarchical Network Framework for this collection \mathbb{Y} is given as

$$P(Y|X, \Theta) = \prod_{k=1}^K P(Y_k | X_k, \Theta_k = (\theta_{1k}, \dots, \theta_{pk}))$$

$$(\Theta_1, \dots, \Theta_K) \sim F(\Theta_1, \dots, \Theta_K | W_1, \dots, W_K, \psi), \quad (4.8)$$

where $P(Y_k | X_k, \Theta_k = (\theta_{1k}, \dots, \theta_{pk}))$ is a probability model for network k with covariates X_k .

Notice that this model structure specifies that networks may be independent of each other depending on choice of W but need not be. Additional hierarchical structure can be specified by including additional parameters ψ . Notice also that we purposely omit any within-network dependence assumptions. Thus, this framework allows for a variety of dependence assumptions both across and within networks but is also flexible in that any social network model can be used. For example, Sweet et al. (2012) uses this framework to introduce hierarchical latent space models, a latent space modeling approach for multiple isolated networks.

\int

4.3.2 The Hierarchical Mixed Membership Stochastic Blockmodel

Let Y_{ijk} be a binary tie from node i to node j in network k . The Hierarchical Mixed Membership Stochastic Blockmodel is specified as

$$P(Y|S, R, B, \theta, \gamma)$$

new page

$$= \prod_{k=1}^K \prod_{i \neq j} P(Y_{ijk} | S_{ijk}, R_{jik}, B_k, \theta_k, \gamma_k) P(S_{ijk} | \theta_{ik}) P(R_{jik} | \theta_{jk}) \prod_i P(\theta_{ik} | \lambda_k) \quad (4.9)$$

where S_{ijk} is the group membership indicator vector for person i when sending a tie to person j in network k , and R_{jik} is the group membership indicator vector for j when receiving a tie from i in network k . B_k is the network specific group-group tie probability matrix and θ_{ik} is the group membership probability vector for node i in network k .

This is easily presented as a hierarchical Bayesian model:

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli}(S_{ijk}^T B_k R_{jik}) \\ S_{jik} &\sim \text{Multinomial}(\theta_{ik}, 1) \\ R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\ \theta_{ik} &\sim \text{Dirichlet}(\lambda_k) \\ B_{\ell mk} &\sim \text{Beta}(a_{\ell mk}, b_{\ell mk}). \end{aligned} \quad (4.10)$$

We impose our hierarchical structure by requiring that the parameters come from some common distribution, and in fact, this framework becomes particularly interesting in cases where parameters are shared across networks. We present several examples in the next section.

4.3.2.1 Examples of HMMSBMs

The hierarchical structure of the HMMSBM naturally lends itself to pooling information across networks and we present several extensions of Equation 4.10.

A simple extension is an HMMSBM for experimental data in which the treatment is hypothesized to affect a single parameter. The networks in the treatment condition would be generated from the same model and the control condition networks would be generated from a different model. For example, suppose we examine teacher collaboration networks in high schools. Typically we would expect to see teachers collaborating within their own departments and these departments operating mostly in isolation. But we could imagine an intervention whose aim is to increase collaboration across departments. Now teachers in treatment schools are more likely to have across department ties than teachers in control schools. Such a model is given as

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli}(S_{ijk}^T B_k R_{jik}) \\ S_{jik} &\sim \text{Multinomial}(\theta_{ik}, 1) \\ R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\ \theta_{ik} &\sim \text{Dirichlet}(\lambda_k) \\ B_{\ell mk} &\sim \text{Beta}(a_{\ell mk}, b_{\ell mk}) \\ \lambda_k &= \lambda_0 + T_k(\bar{1} - \lambda_0)(1 - \alpha) \\ \alpha &\sim \text{Uniform}(0, 1), \end{aligned} \quad (4.11)$$

line these up?

In contrast,

where T_k is the indicator for being in the treatment group, $\vec{1}$ is the vector $(1, \dots, 1)$ with length g , and g is the number of groups. The treatment effect α is a proportion of how similar the group membership profiles are to the control group as compared to a uniform distribution on the simplex.

Rather than constraining each network to have a constant network level parameter, e.g. λ_0 , we might instead model network parameters generated from a single distribution, introducing an additional level to the hierarchy. Suppose we are interested in how variable the membership probabilities vectors are across networks, for example we expect teacher collaboration networks to vary depending on the organizational structure in the schools. Then we could estimate the distributional hyperparameters that generate these membership probabilities (θ).

An example of such a model is

$$\begin{aligned}
 Y_{ijk} &\sim \text{Bernoulli}(S_{ijk}^T B_k R_{jik}) \\
 S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1) \\
 R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\
 \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi_k) \\
 B_{\ell mk} &\sim \text{Beta}(a_{\ell mk}, b_{\ell mk}) \\
 \gamma_k &\sim \text{Gamma}(\tau, \beta) \\
 \xi_k &\sim \text{Dirichlet}(c) \\
 \tau &\sim \text{Gamma}(a_\tau, b_\tau) \\
 \beta &\sim \text{Gamma}(a_\beta, b_\beta).
 \end{aligned} \tag{4.12}$$

Thus we allow γ_k to vary by network and then estimate an overall mean and variance as determined by (τ, β) .

Finally, we introduce a covariate MMSBM in Section 4.2.3, which we can easily extend for multiple networks. Consider again our networks of teacher collaboration. A tie-level variable indicating whether two teachers serve on the same committee may be a covariate of interest, such as $X_{ijk} = 1$ if teacher i and j in school k serve on the same committee, and we may want to estimate this effect across all networks. A simple model in which the covariate effect is the same across networks is given as

$$\begin{aligned}
 Y_{ijk} &\sim \text{Bernoulli}(p_{ijk}) \\
 p_{ijk} &= \frac{\exp\{\text{logit}(S_{ijk}^T \text{logit}(B_k) R_{jik}) + \alpha X_{ijk}\}}{1 + \exp\{\text{logit}(S_{ijk}^T B_k R_{jik})\}} \\
 S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1) \\
 R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\
 \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi_k) \\
 B_{\ell mk} &\sim \text{Beta}(a_{\ell mk}, b_{\ell mk}).
 \end{aligned} \tag{4.13}$$

These are merely a few models from the myriad of possibilities. Network-level

no comma

5

generate

hyphen

experiments can affect other parameters in the model; indeed we can include additional hierarchical structure when modeling experimental data. Moreover, observational data may not need the full structure specified above and covariates can be incorporated in other ways as well.

4.3.3 Model Estimation

We use an MCMC algorithm for fitting HMMSMs that is similar to the one used for fitting the single network MMSBM. We first present MCMC steps for fitting the model given in (4.10), and then we will discuss how these steps need to be augmented for models (4.11)–(4.13).

For each network k , we use Gibbs updates for θ_k, S_k, R_k, B_k . The complete conditionals for our Gibbs updates are given as:

$$\begin{aligned}
 P(\theta_{ik} | \dots) &\propto \text{Dirichlet}(\gamma_k \xi_k + \sum_j S_{ijk} + \sum_j R_{ijk}) \\
 P(S_{ijk} | \dots) &\propto \text{Multinomial}(p) \\
 p_h &= \theta_{ikh} B_{h\ell^*}^{Y_{ijk}} (1 - B_{h\ell^*})^{(1 - Y_{ijk})} \\
 P(R_{ji} | \dots) &\propto \text{Multinomial}(q) \\
 q_h &= \theta_{ikh} B_{m^*h}^{Y_{ijk}} (1 - B_{m^*h})^{(1 - Y_{ijk})} \\
 B_{\ell mk} &\propto \text{Beta}(a_{\ell k} + \sum_{(ijk)^*} Y_{ijk}, b_{mk} + \sum_{(ijk)^*} Y_{ijk}) \quad \wedge
 \end{aligned} \tag{4.14}$$

en dash

no indent

where ℓ^*_k is the group membership indicated by R_{jik} and m^*_k is the group membership indicated by S_{ijk} . Again, let $(ijk)^*$ be a specific subset of $i = 1, \dots, n_k$ and $j = 1, \dots, n_k$ such that $S_{ijk} = \ell$ and $R_{jik} = m$. Again, we incorporate a sparsity parameter ρ to account for the absence of ties due to lack of interaction.

For the intervention and covariate examples (4.11) and (4.13) respectively, the additional parameter α uses Metropolis or Metropolis-Hastings updates. For example, if we use a random walk method for proposing new values of α , we accept α^{s+1} with probability $\min\{1, R\}$ where $R = \frac{P(\alpha^{(s+1)} | \dots)}{P(\alpha^{(s)} | \dots)}$. Note that R is a function of all K networks.

For models with additional levels of hierarchy, we can update additional parameters using Metropolis within Gibbs steps. For example, in (4.12), β is updated using Gibbs steps.

$$\beta \propto \text{Gamma}(K\tau + a_\beta, \sum_{k=1}^K \gamma_k + b_\beta),$$

and we use Metropolis updates for γ_k, ξ_k and τ . We update each γ_k using an analogous Metropolis step as for the single network. We use a network-specific tuning

hyphen

parameter $\nu_{\gamma,k}$ which is also the shape parameter and rate parameter of $\frac{\nu_{\gamma,k}}{\gamma_k^{(\xi)}}$, ensuring the proposal distribution has mean at the current value of γ_k . Each ξ_k is updated in the same way using a Dirichlet proposal distribution.

To update τ we use a Gamma proposal distribution not unlike those used for γ_k with shape parameter as the tuning parameter ν_τ and rate parameter such that the proposal distribution has mean of the current value of τ . Then the proposed value of τ^{s+1} is accepted with probability $\min\{1, R\}$ where $R = \frac{P(\tau^{(s+1)}|\dots) P(\tau^{(s)}|\tau^{(s+1)})}{P(\tau^{(s)}|\dots) P(\tau^{(s+1)}|\tau^{(s)})}$.

4.3.4 Empirical Examples

We present two examples to illustrate fitting HMMSBMs and use two simulated datasets, with and without a covariate.

In the first example we demonstrate fitting an HMMSBM similar to the example given in (4.12) where each network has a network-specific Dirichlet hyperparameter, λ_k , used to generate the membership probability vectors. Our goal is to assess parameter recovery on three levels: the hyperparameters of the distribution that generates λ_k , the λ_k themselves and the lower-level parameters, R, S, B that determine the probability of a tie.

We simulate data from 20 networks, each with 20 nodes and 4 groups using the following model to generate our first set of data:

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli}(S_{ijk}^T B_k R_{jik}) \\ S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1) \\ R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\ \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi) \\ \gamma_k &\sim \text{Gamma}(10, 50), \end{aligned} \tag{4.15}$$

where $\xi = (0.25, 0.25, 0.25, 0.25)$. The group-group tie probability matrix is defined as

$$B = \begin{bmatrix} 0.9 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.8 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.7 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.6 \end{bmatrix}.$$

We constrain B to be the same for each network and selected the hyperparameters with which to generate γ_k to ensure small enough values for block structure with low variability. Figure 4.4 shows adjacency matrices for these 20 networks.

We fit the following HMMSBM to these data using the MCMC algorithm described in Section 4.3.3. We let $\xi = (0.25, 0.25, 0.25, 0.25)$ and use a sparsity parameter equal to $\frac{\sum_{ijk} Y_{ijk}}{KN(N-1)}$ for all K networks. The model is given as

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli}(S_{ijk}^T B_k R_{jik}) \\ S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1) \end{aligned}$$

datasets

hyper x 2

and

no 0

no comma

no comma x 2

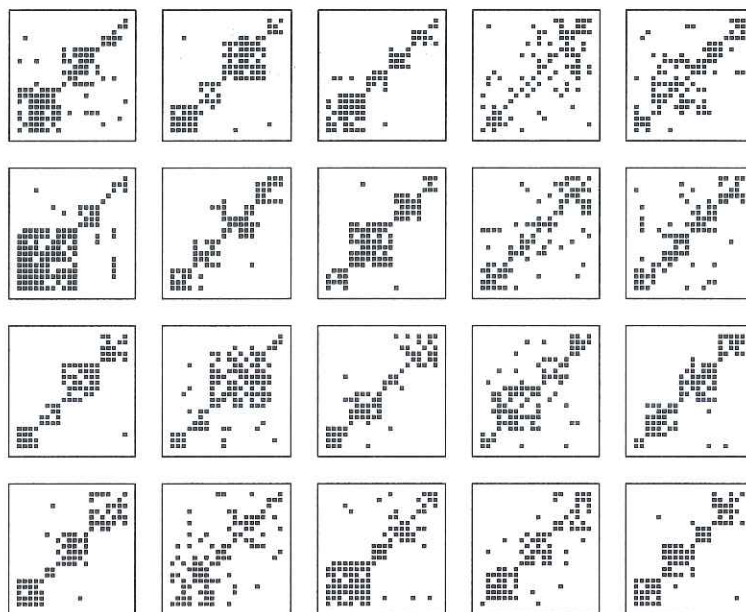


FIGURE 4.4 Networks with 20 nodes generated from a HMMSBM with group membership probabilities from a network-specific Dirichlet parameter $\gamma_k \sim \text{Gamma}(50, 10)$.

$$\begin{aligned}
 R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\
 \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi_k) \\
 B_{\ell\ell k} &\sim \text{Beta}(3, 1) \\
 B_{\ell m k} &\sim \text{Beta}(1, 10), \ell \neq m \\
 \lambda_k &\sim \text{Gamma}(\tau, \beta) \\
 \tau &\sim \text{Gamma}(50, 1) \\
 \beta &\sim \text{Gamma}(10, 1).
 \end{aligned}
 \tag{4.16}$$

no ⊙

Running MCMC chains of length 30,000, we remove the first 5000 steps and then by retaining every 25th step for a posterior sample of 1001. The posterior samples for τ and β are illustrated as \hat{g} densities in Figure 4.5. The vertical line shows the true value for each parameter and gray region indicates the 95% credible interval, suggesting accurate parameter recovery for τ and β . Similar plots for γ_k (see Figure 4.6) for each network k depict the variability in both the true value of γ_k as well as the accuracy of recovery.

⊙ Reword
 the × 2

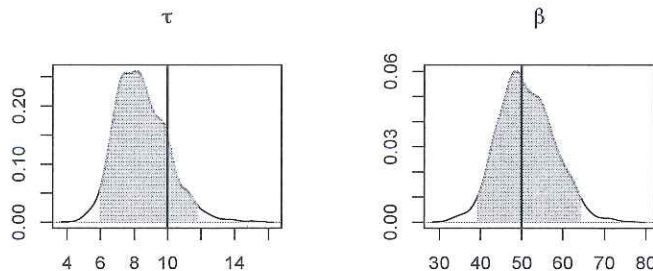


FIGURE 4.5

Posterior Density for τ and β , the hyperparameters for the distribution of γ_k for each network. The vertical line marks the value used to simulate the data, and the 95% equal-tailed credible intervals indicated with gray. Densities show good recovery of the true value of each parameter.

S
 are

We use predicted probability tie matrices to assess recovery of lower-level parameters (Figure 4.7). Ties with high probability are shown as shades of red, orange and yellow and ties with low probability are shown as shades of blue and purple. Visual comparisons to Figure 4.4 reveal that the estimated tie probabilities align with the simulated data; between-pairwise ties are reflected as having higher probability in the fitted model than non-ties. We do note however that ties that exist across groups (those shown outside of the block structure) tend to have smaller estimated probabilities than ties within groups.

hyphen
 S
 hyphen
 S × 2

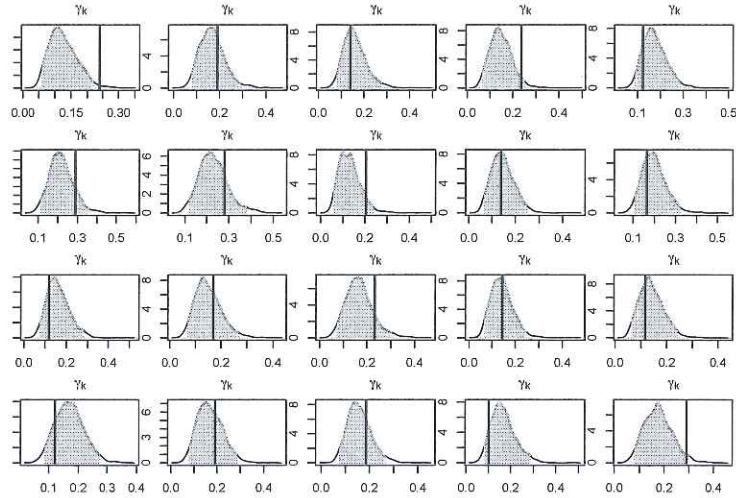


FIGURE 4.6

Posterior densities for γ_k where $k = 1, \dots, 20$. The 95% equal-tailed credible intervals contain the true value of γ_k for all but one of the simulated networks, suggesting good recovery.

The second simulation serves two purposes: to illustrate fitting a MMSBM with covariates and to provide a second example of fitting HMMSBMs. We generate data for 10 networks, each with 15 nodes and 3 groups. We use a single edge-level indicator covariate X_{ijk} , such that $X_{ijk} = 1$ implies that individual i in network k and individual j in network k have the same characteristic and $X_{ijk} = 0$ otherwise. In the context of teacher relationships in school k , for example, X_{ijk} might represent teaching the same grade, serving on the same committee, having ~~their classroom~~ classrooms in the same wing of the building, etc. For these data, we randomly assigned each node to one of 5 groups, and $X_{ijk} = 1$ if nodes belong to the same group. The formal model used to generate these data is:

$$\begin{aligned}
 Y_{ijk} &\sim \text{Bernoulli}(p_{ijk}) \\
 p_{ijk} &= \frac{\exp\{\text{logit}(S_{ijk}^T \text{logit}(B_k) R_{jik}) + 4X_{ijk}\}}{1 + \exp\{\text{logit}(S_{ijk}^T B_k R_{jik})\}} \\
 S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1) \\
 R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\
 \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi_k) \quad \checkmark \\
 B_{\ell k} &\sim \text{Beta}(12, 4) \\
 B_{m k} &\sim \text{Beta}(3, 30) \quad \ell \neq m \quad \checkmark
 \end{aligned}$$

no comma x 2

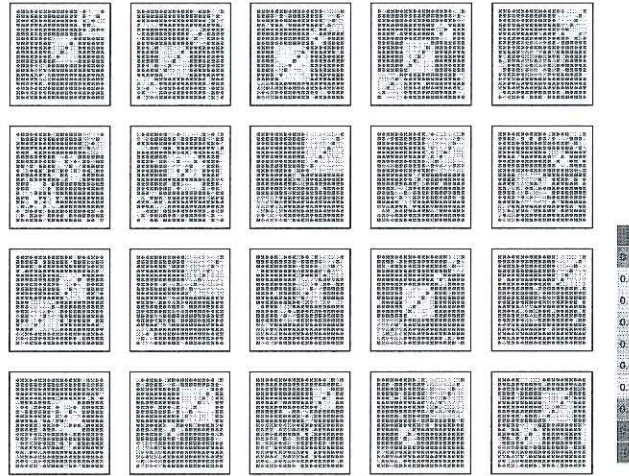


FIGURE 4.7 Tie probability matrix as estimated by posterior means. Visual comparisons to Figure 4.4 suggests accurate estimation of tie probability.

no indent $\gamma_k \sim \text{Gamma}(10, 60)$, (4.17)
 where $\xi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Priors for γ_k were selected to ensure small enough values for block structure with low variability. We use different priors for the diagonal entries of B_k than the off-diagonal entries to model higher within-group tie probabilities. Hyperparameters of these priors were chosen to yield high and low probabilities for the diagonal and off-diagonal entries, respectively, without extreme values of almost 0 or 1.

The adjacency matrices for each of the 10 simulated networks are shown in Figure 4.8. We expect to see more variability in the block structure in these networks as compared to the first simulation study for two reasons. Foremost, we have included a covariate with a strong effect so that there are now many more across-group ties. In addition, we have varied the group-group tie probability matrix B_k by allowing these entries to both differ across networks and be generated (instead of deliberately chosen). As a result, the block structure that we do see varies across networks as the values of the diagonal entries of B_k vary.

We fit the following model on these simulated data:

$$\begin{aligned}
 Y_{ijk} &\sim \text{Bernoulli}(p_{ijk}) \\
 p_{ijk} &= \frac{\exp\{\text{logit}(S_{ijk}^T \text{logit}(B_k) R_{jik}) + \alpha X_{ijk}\}}{1 + \exp\{\text{logit}(S_{ijk}^T B_k R_{jik})\}} \\
 S_{ijk} &\sim \text{Multinomial}(\theta_{ik}, 1)
 \end{aligned}
 \tag{4.18}$$

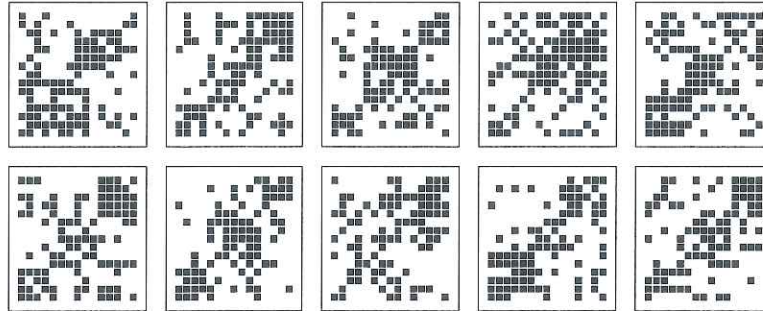


FIGURE 4.8

Networks generated from a single covariate HMMSBM with group membership probabilities from a network-specific Dirichlet parameter $\gamma_k \sim \text{Gamma}(60, 10)$. Despite a strong block structure specified by small values of γ_k , the networks have many across-group ties due to the high value of the regression coefficient of X_{ijk} .

hyphen

$$\begin{aligned}
 R_{jik} &\sim \text{Multinomial}(\theta_{jk}, 1) \\
 \theta_{ik} &\sim \text{Dirichlet}(\gamma_k \xi_k) \\
 B_{\ell\ell k} &\sim \text{Beta}(12, 4) \\
 B_{\ell m k} &\sim \text{Beta}(3, 30) \quad \ell \neq m \\
 \gamma_k &\sim \text{Gamma}(\tau, \beta) \\
 \tau &\sim \text{Gamma}(1, 0.1) \\
 \beta &\sim \text{Gamma}(6, 0.1) \\
 \alpha &\sim \text{Normal}(0, 100),
 \end{aligned}$$

no comma

where $\xi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

We run MCMC chains of length 30000, remove the first 5000 iterations and keep every 25th step. With a posterior sample size of 1001, we assess parameter recovery. We begin with our high-level parameters. Figure 4.9 and Figure 4.10 show the posterior densities for α and τ and β respectively. The true value of each parameter is indicated by a vertical line and 95% credible interval regions are shown in gray.

The estimation of α is accurate, but the estimates for τ and β are much less precise. The posterior distribution for τ is centered at a higher value than the value of $\tau = 10$ used to generate the data. The distribution for β is centered at a value slightly lower than the true value $\beta = 60$. Similarly, the distributions for each γ_k are skewed toward higher values. As shown in Figure 4.11, only 3 of the 10 posterior samples for γ_k contain the true value in their 95% credible interval. We suspect the lack of block structure contributes to these biases even though the covariate was the primary influence for across-group ties in the data generation process.

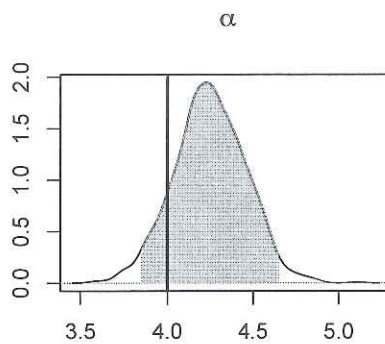
We plot the pairwise probability of a tie in each network in Figure 4.12. Due

insert space

5 x 2
hyphen

5

is

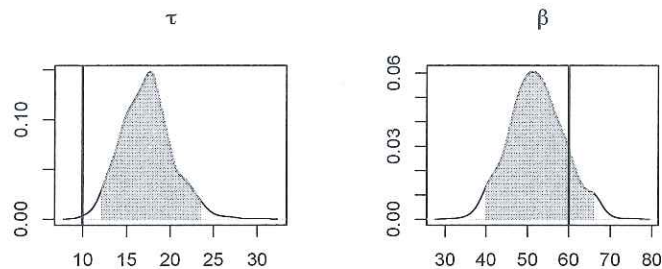
**FIGURE 4.9**

Posterior density for α , the regression coefficient in Equation 4.13. The true value of α is 4 and is displayed as the vertical line. The 95% equal-tailed credible interval is implied by the gray region and suggests good recovery.

Rephrase

[to the lack of block structure, it is make visual comparisons but in fact probabilities align with the data; high pairwise tie probabilities are estimated for pairs with ties and low pairwise tie probabilities otherwise.

I will fix this space!

**FIGURE 4.10**

Posterior density for τ (left) and β (right) with the true values shown as vertical lines. The 95% equal-tailed credible intervals are implied by the gray region. Much of the posterior distribution for τ falls to the right of the true value used to generate the data.

4.3.5 HMMSBM Extension: Sensitivity Analysis

Given the small number of networks used in our simulations, we are interested in the extent to which our prior specification dominates our model fit. Recall from (4.15), we generated data with $\tau = 10$ and $\beta = 50$ and in the model fit illustrated in Section 4.3.4, we used the following prior distributions:

$$\begin{aligned}\tau &\sim \text{Gamma}(10, 1) \\ \beta &\sim \text{Gamma}(50, 1).\end{aligned}$$

line up?
x3

We repeat model estimation twice using a less strong prior and a weak prior, such that both are centered at the true values. The moderate and weak priors used are given as

$$\begin{aligned}\tau &\sim \text{Gamma}(1, 0.1) \\ \beta &\sim \text{Gamma}(5, 0.1)\end{aligned}$$

5

$$\begin{aligned}\tau &\sim \text{Gamma}(0.1, 0.01) \\ \beta &\sim \text{Gamma}(0.5, 0.01).\end{aligned}$$

We first compare the posterior distributions for τ and β . While the posterior distribution for τ and β contain the true values for each fit, the variance of the posterior sample increases as the variance of prior distribution increases (Figure 4.13). We do

()

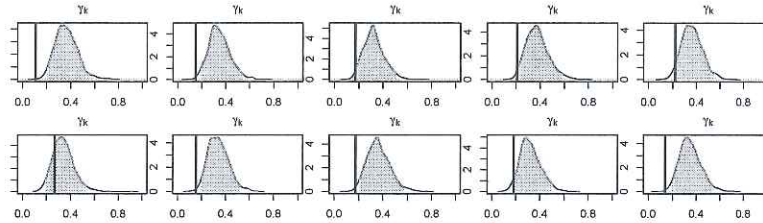


FIGURE 4.11 Posterior densities for γ_k where $k = 1, \dots, 10$. The 95% equal-tailed credible intervals contain the true value of γ_k for only one of the networks and overestimating the value of γ_k .

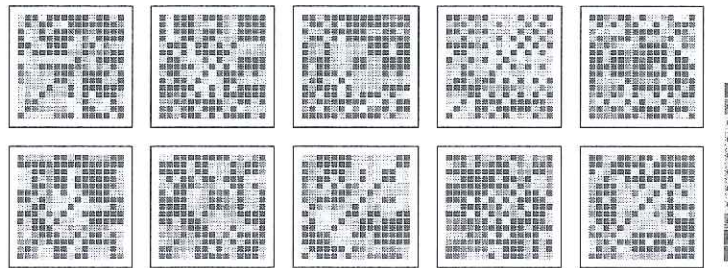


FIGURE 4.12 Tie probability matrix as estimated by posterior means. Ties with high probability are shown as shades of red, orange and yellow and ties with low probability are shown as shades of blue and purple. Pairwise probabilities align well with the original adjacency matrices (Figure 4.8).

note that the scale of the increase is less than the 10-fold increase of the prior distribution variance (50, 500, and 500 for β and 10, 100, and 100 for τ). The posterior mean for τ varies little as the prior changes, 8.5, 8.2, and 10.5 under a strong, less strong and weak prior respectively. The posterior mean for β is much less accurate when the weak prior is used. The respective means are 51.1, 51.8 and 120.3.

To assess the prior distributions, we compare the 95% credible regions posterior distribution for γ_k , $k = 1, \dots, 20$ with the true values (Figure 4.14). We notice the following pattern: If a 95% credible interval γ_k does not cover the true value when the prior distribution is strong, it fails to cover the true value when the prior is moderate or weak. There is little difference in parameter recovery between the strong prior and the less strong prior. The weak prior fit recovers few of the γ_k well, and is strongly biased toward smaller values of γ_k .

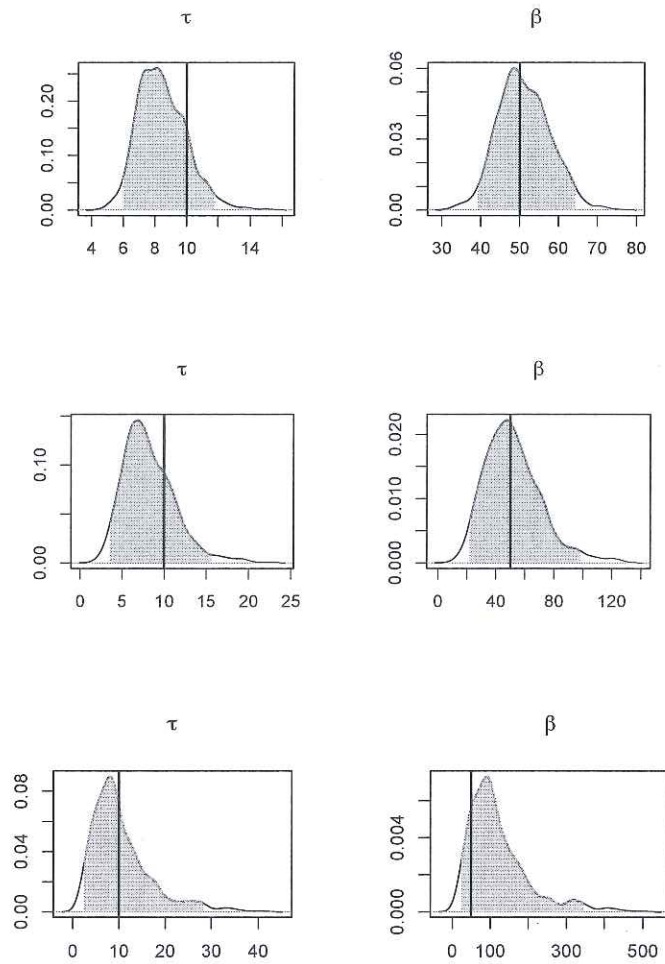
Finally, we are interested in how these differences translate to tie probabilities.

⊙

5

5 x 3

()

**FIGURE 4.13**

A comparison of posterior distribution for τ and β given three different prior gamma distributions. Hyperparameters are $(10,1)$, $(1,0.1)$, $(0.1, 0.01)$ and $(50,1)$, $(5,0.1)$, $(0.5, 0.01)$ for τ and β respectively and plots are shown top to bottom.

space
 5×2

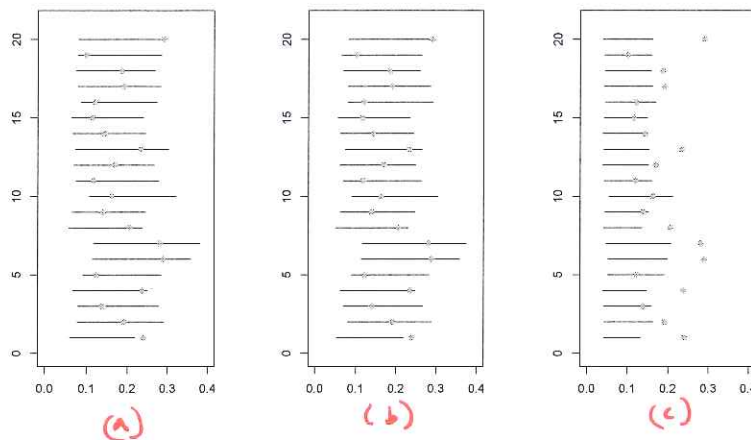


FIGURE 4.14

For each choice of prior specification, strong (left), less strong (middle), and weak (right), 95% credible intervals for γ_k are shown in black and the true value of γ_k is shown in green. Parameter recovery is good when the strong or less strong priors are used but is poor when the weak prior is used.

Figure 4.3.5 shows the adjacency matrix and posterior mean of the pairwise probability of a tie determined for each model fit. We also include a measure of variability, the width of the 95% credible interval for each pairwise tie probability. For brevity, we only show the first 5 networks from the data. The posterior pairwise tie probability varies little across each fit and what is even more surprising is that the 95% credible interval widths also varies little.

Based on this simple sensitivity analysis, we offer several conclusions. The prior specification of the high-level parameters \mathcal{T} and β have moderate influence of mid-level parameters γ_k and very little influence on low-level parameters $R, S,$ and $B,$ even with poor recovery of mid-level parameters. Furthermore, using a prior with much larger variance does not necessarily increase the variability in the low-level parameter estimates.

no common hypern 23

4.4 Discussion

We have presented the Hierarchical Network Models (HNM) framework for modeling ensembles of networks and introduced the Hierarchical Mixed Membership

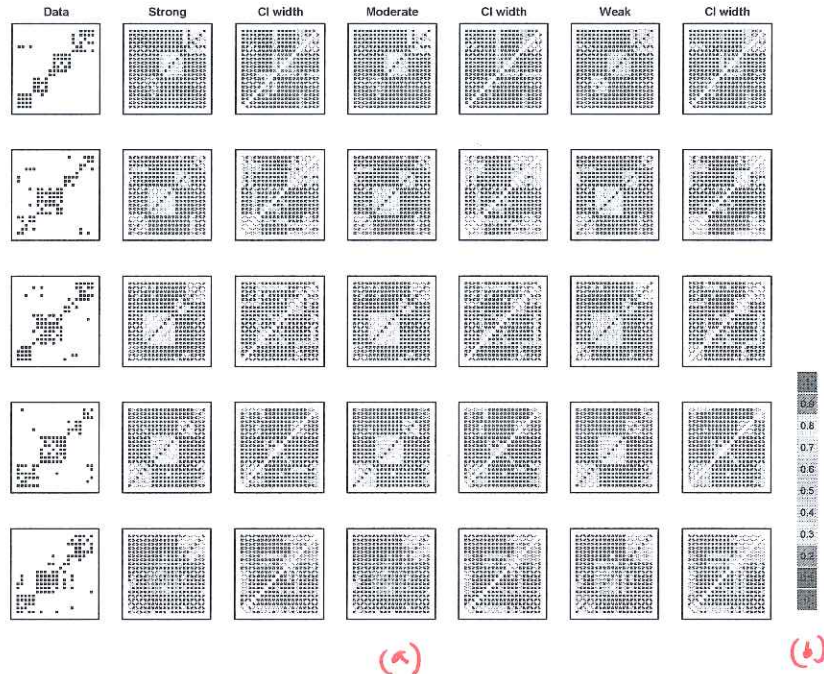


FIGURE 4.15

The adjacency matrix (left) can be compared to the posterior tie probabilities for three model fit that vary by prior distribution specification for τ and β . Priors for each parameter have the same mean but increase in variance by a factor of 10. The width of the 95% credible interval serves as a measure of variability.

Stochastic Blockmodel (HMMSBM) as an example of an HNM for networks with subgroup structure. This fills a substantial methodological void: while both Single and Mixed Membership Stochastic Blockmodels have been used to incorporate grouping structure into models for relational data, very little prior work has focused on jointly modeling an ensemble of networks, and none of that work has focused on block models. In addition we have presented a method for incorporating tie covariates into these models, addressing another void in the literature.

We presented several examples of HMMSBMs to demonstrate both the generality and wide utility of these models. We use two simulated data sets, one with a covariate and one without a covariate, to illustrate model fitting using our MCMC algorithm. Posterior tie probabilities from our fits align well with simulated true ties and non-ties, and in most cases parameters were recovered well. High level parameters, those furthest away from the data, were recovered with less consistency in the simulation study involving tie covariates. Finally, we investigated the effects of prior

no hyphens

datasets

hyphen

true-ties and non-ties ?
true ties and non ties ?

hyphen

specification and found that as expected, high-level parameters were most affected by choice of prior but that priors had little influence on predicted tie probabilities.

With respect to the class of HMMSBMs and modeling fitting, our work reveals several areas for future work. Ties can form perhaps independently of subgroup structure due to common attributes, including covariates to account for this should produce preferable models. An important area for future research is understanding the how covariate effects and block effects interact with each other. Finally, high-level parameter estimates seem to depend strongly on hyperpriors, suggesting that estimation of these parameters is not yet data-dominated. Understanding how this situation improves as more networks (and perhaps larger networks) are added to the ensemble is also clearly important. On the other hand, it appears that priors have little effect on the lower-level tie probabilities.

hyphen

We have illustrated a proof of concept for HMMSBMs and the HNM framework in general. HMMSBMs are appropriate models for ensembles of networks with block structure and can be fit using relatively simple methods. The HNM framework is larger than HMMSBMs alone since most single network statistical models can be extended to model an ensemble of networks. Sweet et al. (2012) introduced Hierarchical Latent Space Models as a class of HNM models, and the authors are currently working on extending work done by Zijlstra, van Duijn and Snijders (2006) and Templin, Ho, Anderson and Wasserman (2003) for hierarchical Exponential Random Graph Models, and relating it to the general HNM framework.

Program for Interdisciplinary Education Research, supported by the Institute for Education Sciences, Department of Education, Grant R305B040063.

{ I'm going to move this around - maybe back to a footnote

↑
model
trans.
keep i

References

- Airoldi, E., Blei, D. M., Fienberg, S., and Xing, E. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research* 9: 1981–2014.
- Airoldi, E., Blei, D. M., Xing, E., and Fienberg, S. (2005). A latent mixed membership model for relational data. In *Proceedings of the 3rd International Workshop on Link Discovery (LINKKDD '05)*. Chicago, Illinois, USA: ACM, 82–89.
- Anderson, C. and Wasserman, S. (1992). Building stochastic blockmodels*. *Social Networks* 14: 137–161.
- Breiger, R., Boorman, S., and Arabie, P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology* 12: 328–383.
- Erosheva, E. (2003). Bayesian estimation of the grade of membership model. *Bayesian Statistics* 7: 501–510.
- Fienberg, S., Meyer, M., and Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association* 80: 51–67.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC press. 2nd edition.
- Goldenberg, A., Zheng, A., Fienberg, S., and Airoldi, E. (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning* 2: 129–133.
- Goodreau, S., Kitts, J. A., and Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46: 103–125.
- Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* 4: 585–605.
- 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.
- Hoff, P. D. and Ward, M. D. (2004). Modeling dependencies in international relations networks. *Political Analysis* 12: 160–175.

- Kolaczyk, E. (2009). *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.
- Krackhardt, D. and Handcock, M. (2007). Heider vs Simmel: Emergent features in dynamic structures. *Statistical Network Analysis: Models, Issues, and New Directions* : 14–27.
- Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52: 169–193.
- Pitts, V. and Spillane, J. (2009). Using social network methods to study school leadership. *International Journal of Research & Method in Education* 32: 185–207.
- Sampson, S. (1968). A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships. Ph.D. thesis, Cornell University, Ithaca, New York, USA.
- Sweet, T. M., Thomas, A. C., and Junker, B. W. (2012). Hierarchical network models for education research. *Journal of Educational and Behavioral Statistics*. doi:10.3102/1076998612458702.
- Templin, J., Ho, M.-H., Anderson, C., and Wasserman, S. (2003). Mixed effects p^* model for multiple social networks. In *Proceedings of the American Statistical Association: Section on Bayesian Statistical Science*. Alexandria, Virginia, USA: American Statistical Association, 4198–4024.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika* 61: 401–425, 10.1007/BF02294547.
- Weinbaum, E., Cole, R., Weiss, M., and Supovitz, J. (2008). Going with the flow: Communication and reform in high schools. In Supovitz, J. and Weinbaum, E. (eds), *The Implementation Gap: Understanding Reform in High Schools*. New York: Teachers College Press, 68–102.
- Westveld, A. and Hoff, P. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics* 5: 843–872.
- Xing, E., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics* 4: 535–566.
- Zijlstra, B., van Duijn, M., and Snijders, T. (2006). The multilevel p_2 model. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 2: 42–47.