A BAYESIAN ANALYSIS OF WEIGHTED STOCHASTIC BLOCK MODELS WITH APPLICATIONS IN BRAIN FUNCTIONAL CONNECTOMICS

Christopher Bryant

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

Joseph Ibrahim

Hongtu Zhu

Yun Li

Mengjie Chen

Yen-Yu Ian Shih

ABSTRACT

CHRISTOPHER BRYANT: A Bayesian Analysis of Weighted Stochastic Block Models with Applications in Brain Functional Connectomics
(Under the direction of Joseph Ibrahim and Hongtu Zhu)

The network paradigm has become a popular approach for modeling complex systems, with applications ranging from social sciences to genetics to neuroscience and beyond. Often the individual connections between network nodes are of less interest than network characteristics such as its community structure - the tendency in many real-data networks for nodes to be naturally organized in groups with dense connections between nodes in the same (unobserved) group but sparse connections between nodes in different groups. Characterizing the structure of networks is of particular interest in the study of brain function, especially in the context of diseases and disorders such as Alzheimer's disease and attention deficit hyperactivity disorder (ADHD), where disruption of functional brain networks has been observed.

The stochastic block model (SBM) is a probabilistic formulation of the community detection problem that has been utilized to estimate latent communities in both binary and weighted networks, but as of yet not in brain networks. We build a flexible Bayesian hierarchical framework for the SBM to capture the community structure in weighted graphs, with a focus on the application in functional brain networks.

First, we fit a version of the SBM to Gaussian-weighted networks via an efficient Gibbs sampling algorithm. We compare results from simulated networks to several existing estimation methods and then apply our approach to estimate the community structures in the functional resting brain networks of 185 subjects from the ADHD-200 sample.

Next, we extend this probabilistic framework and our efficient estimation algorithm to capture the shared latent structure in groups of networks; we perform simulation studies and then apply this extended model to the same sample of brain networks from the ADHD-200 sample.

Finally, we adapt this model to allow for more complex latent structures and incorporate a regression component to test for differences in the latent functional brain structure between study groups. After examining the ability of this approach to capture the latent structures in simulated networks, we apply this method once again to the same set of functional brain networks to assess the differences between ADHD subtypes and healthy control subjects in latent functional brain structure.

ACKNOWLEDGMENTS

I'd like to give thanks to the National Cancer Institute for supporting several years of my research through the training grant *Biostatistics for Research in Genomics and Cancer*, NCI grant 5T32CA106209-07 (T32). Also, thanks to my advisors, Dr. Joseph Ibrahim and Dr. Hongtu Zhu, for their support through my time at UNC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In high-dimensional data, it is often of interest to understand the relationships between a number of variables. Networks are commonly used to model complex systems with up to millions of variables; some that are studied extensively include biological networks such as gene regulatory networks, protein-protein interactions, and functional brain networks; social networks of friends; links between webpages on the internet; contact between individuals in epidemiological studies.

While graph theory (graph and network are interchangeable terms) has been studied for decades, recently work in this area has expanded rapidly with technology bringing "big data" to the forefront of many fields. Some types of networks are directly observed, for example co-authorship networks and social networks, while others have to be estimated from observed data. Numerous algorithmic approaches have been used to estimate the structure and the parameters of directed acyclic graphs (DAGs), in which edges (signifying connections or associations) between nodes (representing the random variables) have explicit direction and no self-loops are allowed. Bayesian networks give a flexible framework for modeling dependencies in a DAG through factorization of the joint probability distribution via the Bayes theorem. Correlation and partial correlation are among the methods used to estimate undirected graphs, where edges indicate association but not causality. Edges between pairs of nodes can be binary, indicating the presence or absence of a connection, or weighted - for example, counts of the number of connections between nodes or their correlation over time. See Kolaczyk (2009) for a more thorough introduction to networks.

The field of network science has focused on answering questions about the structure

and organization of these complex systems. Measures and features such as connectivity (a fully connected graph has an edge between each pair of nodes), degree distribution (the degree of a node is the number of edges connecting to it), motifs (repeating sub-graphs within a graph or set of graphs), clustering coefficient (measuring the tendency of nodes to form small clusters), and community structure (also called graph clustering some places) describe both global and local characteristics of networks, and there is particular interest in understanding these aspects of functional brain networks. These and other measures give macro-level information on the organization of high-dimensional complex systems.

1.1 Brain Connectomics

The network paradigm is especially useful in studying the structure and function of the brain. Recently, a push has been made to characterize the connection matrix of the brain, called the "connectome" (Sporns et al. 2005, Cao et al. 2014): the Human Connectome Project (`http://www.humanconnectomeproject.org`) aims to map functional and structural brain connectivity and its variability in healthy adults; the Developing Connectome Project (`http://www.developingconnectome.org/project/`) has the same goal for the developing brain; the BRAIN Initiative (`http://www.braininitiative.nih.gov`) is a large Federally-funded program with the goal of moving beyond a macro-level understanding of the brain and mapping the individual neuronal connections, much in the same way the human genome was mapped in the early 2000s. The existing literature has largely focused on various topological measures, such as degree distribution, clustering coefficient, network diameter, and modularity (Bullmore and Sporns 2012, Simpson et al. 2013); the modular organization of brain networks has been associated with cognitive function (Crossley et al. 2013).

There is also tremendous interest in uncovering the relationships between brain connectivity and predictors such as disease status and behavior. Disruption in brain networks has

been observed in neurological disorders such as depression (Zhang et al. 2011), Alzheimer's disease (Wang et al. 2013, Dickerson and Sperling 2009), and ADHD (Cocchi et al. 2012, Qiu et al. 2011, Ahmadlou et al. 2012, Konrad and Eickhoff 2010, Castellanos and Proal 2012). In ADHD, researchers have found evidence of the importance of functional connectivity in certain brain subnetworks in attention-related tasks (Rosenberg et al. 2015).

1.2 Community Detection

The community network structure, in which there exist groups of nodes that have dense connections within the group and sparse connections between groups, has been observed in numerous real life networks (Zhang et al. 2008, Girvan and Newman 2002, Fortunato 2010, Lancichinetti et al. 2010, Yang and Leskovec 2014), including functional brain networks (Simpson et al. 2013). Closely tied to the concept of community detection is network modularity, which measures the strength of a network's community structure - for binary graphs the quantity is calculated as the difference between the fraction of the edges that fall within the defined communities and the expected fraction if edges were randomly distributed, holding the degree-distribution of the graph constant. The greater the modularity of a network, the more dense connections are within-community compared to out-of-community; this concept can be extended in various ways to account for edge weights (Newman 2004). Fortunato (2010) gives an in-depth review of the various approaches to the community detection problem. These approaches largely fall into one of a few categories: deterministic algorithms that maximize modularity or optimize other graphical quantities (such as betweenness centrality - the number of shortest paths between all other pairs of nodes that pass through a given node, used to find the nodes that connect distinct communities (Girvan and Newman 2002)) with respect to community membership; spectral methods that focus on the eigenvectors of the adjacency matrix or the related Laplacian matrix ($L = D - A$, where D is a diagonal matrix with the degrees of each node and A is the adjacency matrix); and statistical approaches

including generative models based on the joint probability distributions of the edges or on hypothesis testing (see Wilson et al. (2014)).

Random graph models offer a statistical framework for modeling graphs probabilistically and making inference on various aspects of the graphical structure. Paul Erdős and Alfréd Rényi (Erdös and Rényi 1959) and Edgar Gilbert (Gilbert 1959) independently proposed variants of what is now known as the Erdős-Rényi model, in which binary edges appear independently with probability p. Other models have been created to capture characteristics of particular observed networks, such as the average path length between two nodes, the clustering coefficient, and the degree distribution. Exponential random graph models (ERGM) offer a more flexible framework based on the exponential family of distributions, where the probability of an observed graph y is given by:

$$P(Y = y|\eta) = \frac{\exp\{\eta^\top t(y)\}}{c(\eta)}$$

Here η is the natural parameter of the exponential family and $t(y)$ is a vector of statistics related to network or nodal characteristics, and $c(\eta)$ is the normalizing constant. ERGMs have been used to model graphs parametrically (Robins et al. 2007a;b) including in studying functional brain networks (Simpson et al. 2011), but they typically cannot capture community structure (Fronczak et al. 2013).

1.3 Stochastic Block Model

The stochastic block model has been the focus of much of the statistical work on the community detection problem. White et al. (1976) introduced the (deterministic) blockmodel for social networks, where rows and columns of the adjacency matrix A (where $A_{ij} = 1$ if node i and node j are connected and $A_{ij} = 0$ if they are not) are reordered until the matrix consists of blocks along the diagonal. Figure 1 shows the observed adjacency matrix (for a perfect blockmodel) on left, with the reorganized block matrix on right.

Figure 1.1: Blockmodel

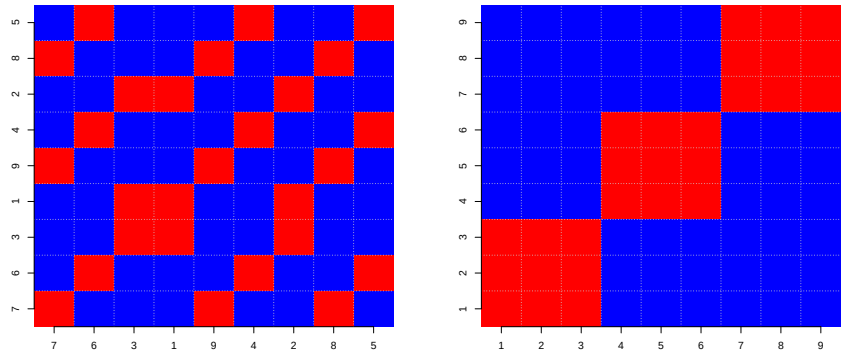


Figure 1.1. Observed adjacency matrix on left, reorganized into blocks on right.

A stochastic model for directed graphs (digraphs) was proposed by Holland and Leinhardt (1981). In what they called the p_1 model, each node has a parameter governing the probability of incoming ties and a parameter governing the probability of outgoing ties, with each dyad in the directed graph (X_{ij}, X_{ji}) assumed to be independent. This model was effectively combined with the deterministic blockmodel into the stochastic blockmodel (SBM) by Holland et al. (1983), to capture a greater variety of graphical structures. In this early formulation, blocks of nodes are known *a priori*, and edges within a given combination of blocks are stochastically equivalent. That is, the probability of an edge between any pair of nodes from blocks B_1 and B_2 is a constant, $p_{B_1B_2}$.

Work on the SBM picked up more than a decade later, when Snijders and Nowicki (1997) proposed a framework for *a posteriori* blockmodeling in undirected graphs with two blocks. It it more typical in practice that the blocks that form the structure of the observed network are not known *a priori* and must be learned from the data; indeed, it is often the primary question of interest. Snijders and Nowicki proposed maximum likelihood estimation, only feasible for small graphs on the order of n=20 nodes, and a Bayesian approach for larger graphs. They extended the model to an arbitrary number of classes, directed edges, and non-binary edges (Nowicki and Snijders 2001); this formulation is the basis for most recent work

on the SBM. Essentially the SBM is a mixture of Erdős-Rényi graphs, with the probability of an edge between two nodes determined by their latent classes.

Snijders and Nowicki propose the general form of the SBM for the undirected graph $Y = (Y_{ij})_{1 \le i \ne j \le n}$ as follows (no self loops $\implies Y_{ii} = 0 \ \forall \ i$):

- The parameters of the model are given by the vector of latent class probabilities $\theta = (\theta_1, \ldots, \theta_Q)$ and class-specific edge probabilities $\eta = (\eta_{kl})$, where Q is the number of latent classes (which they call the colors of the nodes) and $1 \le k \le l \le Q$.
- The latent classes Z_i, for each node $i = 1, \ldots, n$ are i.i.d. generated with $P(Z_i = k) = \theta_k$.
- The edges are i.i.d. conditional on class-specific probabilities: $Y_{ij} | Z_i, Z_j \sim \text{Bernoulli}(\eta_{Z_i Z_j})$.

This leads to the joint probability distribution for (Y, Z):

$$p(y, z; \theta, \eta) = \theta_1^{n_1} \cdots \theta_Q^{n_Q} \prod_{1 \le k \le l \le Q} \eta_{kl}^{e_{kl}} (1 - \eta_{kl})^{n_{kl} - e_{kl}},$$

where $n_k = \sum_{i=1}^{n} I(Z_i = k)$ gives the number of nodes in latent class k and

$$e_{kl} = \frac{1}{1 + \delta_{kl}} \sum_{1 \le i \ne j \le n} y_{ij} I(z_i = k) I(z_j = l)$$

denotes the number of edges in the graph with one node in class k and the other in class l. Here

$$n_{kl} = \begin{cases} n_k n_l, & \text{if } k \ne l \\ \binom{n_k}{2}, & \text{if } k = l \end{cases}$$

and $\delta_{kl} = 1$ for $k = l$ and $\delta_{kl} = 0$ for $k \ne l$. This model is the basis for more recent extensions to directed graphs, signed graphs (with +, 0, - edges), weighted graphs, and beyond.

1.3.1 Identifiability

As in all mixture models, the SBM is only identifiable up to the ordering of the latent labels. Because the likelihood is invariant to permutations of the class labels, multiple parameter sets can lead to exactly the same value of the likelihood. This problem leads to the phenomenon frequently called label switching or swapping, which is well known in the literature of Bayesian mixture modeling. Within Markov Chain Monte Carlo (MCMC) samplers, when using symmetric priors the order of the latent labels will often switch between successive samples, which leads to nonidentifiability of both the labels and the class-specific parameters. One simple approach is to impose artificial identifiability constraints, say $\pi_1 < \pi_2 < \ldots < \pi_Q$, but this is an ineffective approach when the latent class probabilities are relatively similar since it is impossible to tell the difference between natural variation and label switching. Similarly, informative priors can be used to break the symmetry in the posterior distribution, but often the ordering is impossible to know *a priori*.

Stephens (2000) proposed a class of relabelling algorithms for post-processing MCMC samples that attempt to minimize the posterior expected loss under a particular loss function. However, for the SBM *a posteriori* relabelling is not usually a valid solution to non-identifiability when the parameters are also being sampled, because they are sampled based on the uncorrected labels. Bayesian approaches to estimating the SBM typically have to incorporate label switching algorithms into the MCMC sampler (such as Cron and West (2011), Mena and Walker (2014)). An alternative approach is to focus on functions of the parameters that are invariant to label order (Nowicki and Snijders 2001) or to parameterize the model in such a way that it is identifiable.

1.3.2 Extensions

Considerable work has been done to extend the SBM to better fit the characteristics of observed graphs. For example, in many social networks there are hub nodes that are more

connected both within its community and to other communities. Since the classical SBM assumes all nodes in a community share the same degree distribution, to capture this degree-heterogeneity Karrer and Newman (2011) proposed the degree-corrected SBM. This model utilizes an additional parameter for each node, governing the node's degree. In addition to better fitting graphs with known degree-heterogeneity, the degree-corrected SBM can sometimes fit comparably well as the uncorrected SBM but with fewer clusters of nodes; that is, the degree-corrected SBM can sometimes explain a graph more succinctly than the uncorrected SBM (Herlau et al. 2014). Yan et al. (2014) introduced an algorithm for model selection between the degree-corrected and uncorrected SBM.

Sometimes it may be known that a particular nodal attribute can affect the structure of a graph, and other times it may be of interest to uncover which attributes influence graph structure. A number of versions of the SBM have incorporated nodal information via regression in the latent space (for example Tallberg (2005), Handcock et al. (2007), Mariadassou et al. (2010), Choi et al. (2012), Vu et al. (2013)).

While the standard SBM assumes that communities are distinct, several extensions have generalized the model to capture more complex latent structures. Hoff et al. (2002) generalized the SBM to what they called latent space modeling, which allows a more general latent structure than that of the blockmodel. Airoldi et al. (2008) introduce the mixed-membership SBM for directed graphs, where nodes can have membership in multiple latent classes dependent on the pair of nodes. The overlapping SBM developed by Latouche et al. (2011) allows clusters of nodes to overlap, and the probability of edges between two nodes depends on how many latent clusters in which they share membership. The community-affiliation graph model presented by Yang and Leskovec (2012) allows for both overlapping and hierarchically-nested communities, the latter of which is also the aim of the method proposed by Lyzinski et al. (2015).

The affiliation model reduces the number of parameters of the SBM to two by constraining

the probabilities of edges to:

$$p_{ql} = \begin{cases} p_{in}, & \text{if } q = l \\ p_{out}, & \text{if } q \neq l. \end{cases}$$

This model is parsimonious and can capture both assortative (where the connections are denser within communities than between them) and disassortative (where the connections are denser between communities than within them) (Ambroise and Matias 2012).

1.3.3 Weighted SBM

While most of the work done on the SBM has been motivated by the study of social networks, and therefore has been focused on modeling the structure of binary graphs, many approaches are easily altered to fit weighted graphs. Newman (2004) generalized a deterministic algorithm for community detection in binary graphs, based on edge-betweenness, to networks with positive weights, and modularity-maximizing community detection methods have also been extended to weighted graphs (Traag and Bruggeman 2009). Within the realm of statistical models for the SBM, Nowicki and Snijders' formulation allows for an arbitrary set of possible values for edges (Nowicki and Snijders 2001). Frequently, in practice, researchers have thresholded weighted graphs such as correlation or partial correlation matrices to reduce them to binary graphs and then apply existing approaches for community detection. Clearly this thresholding is discarding potentially useful information and it may induce an artificial structure to the graph (Fortunato 2010, Simpson et al. 2013).

Mariadassou et al. (2010) introduced a version of the SBM for weighted graphs via the mixed model:

$$Z_i \sim \text{i.i.d. Multinomial}(\alpha),$$

where Z_i denotes the latent class of node i and $\alpha = (\alpha_1, \ldots, \alpha_Q)$, with $\sum_{q=1}^{Q} \alpha_q = 1$ and

$$X_{ij} | i \in q, j \in l, \theta_{ql} \sim f(\cdot, \theta_{ql}),$$

for nodes i, j and latent groups q, l. Here $f(\cdot)$ is a distribution from the exponential family and $\{\theta_{ql}\}_{q,l=1,\ldots,Q}$ are class-specific edge parameters. There has been particular focus on implementation for discrete-valued graphs, such as from the Poisson distribution (see McDaid et al. (2013), Vu et al. (2013), Herlau et al. (2014)), but also for real-valued graphs via the Gaussian distribution and others from the exponential family. Ambroise and Matias (2012) extended the affiliation model to graphs with weighted edges, reducing the number of parameters of the weighted SBM.

1.3.4 Estimation approaches

Approximate maximum-likelihood estimates can be calculated by using an Expectation-Maximization (EM) algorithm (Dempster et al. 1977), but because the expectation step involves summing over all possible combination of the latent variables, this is computationally intractable for any practically-sized graph (Daudin et al. 2008). Snijders and Nowicki (1997) present this ML estimation scheme for small graphs with just two latent classes, and they suggest Bayesian estimation for larger problems.

Because of the computational demands of estimating the SBM, variational methods have also been used to approximate the likelihood or posterior distribution (in the case of Bayesian estimation). While the exact form may vary, typically this involves approximating the target distribution $\pi(z, \theta)$ by a factorizable distribution $q(z, \theta) = q_Z(z)q_\theta(\theta)$; this approach is common in statistical physics and machine learning (Aicher et al. 2015). Within the class of functions $q(\cdot)$ where the parameters of the model and the latent variables are conditionally independent, given the data (and therefore can be factorized as just mentioned), the optimal *variational* distribution q is the one that minimizes the Kullback-Leibler divergence between $q(\cdot)$ and $\pi(\cdot)$ (Attias 2000). Hofman and Wiggins (2008) note that EM is a limiting case of variational Bayes (VB) when the distributions of parameters are collapsed to their modes.

Daudin et al. (2008) introduced a variational method for approximating ML estimates of

the binary SBM, which Mariadassou et al. (2010) extended to weighted graphs. Amini et al. (2013) utilized a pseudo-likelihood which is computationally tractable for solving with an adaptation of the standard EM algorithm. Ambroise and Matias (2012) computed consistent estimates for the SBM by moment estimators and also by use of EM on a tractable composite likelihood; they showed these estimates were consistent and asymptotically normally distributed. Côme and Latouche (2015) introduce what they call the exact integrated complete data likelihood, a version of which had been proposed by Daudin et al. (2008) as a way to compare latent groupings with different numbers of classes; Côme and Latouche develop a greedy algorithm to use this exact method for estimation as well as model selection, but theoretical properties are as of now unestablished.

Bayesian estimation is more commonly used than frequentist methods for estimating all versions of the SBM. Nowicki and Snijders (2001) originally proposed a Gibbs sampling algorithm, which was extended by Tallberg (2005) to include regression to allow for variability in degree within each community, dependent on nodal attributes. In an empirical Bayes approach, Suwan et al. (2014) use a Gibbs sampler which involves Metropolis-Hastings (MH) steps within; Herlau et al. (2014) also utilize a Gibbs sampler with MH within it to sample their nonparametric Bayesian formulation of the degree-corrected SBM; Peng and Carvalho (2013) use a Gibbs sampler with data-augmentation for sampling regression coefficients in their formulation of the degree-corrected SBM; McDaid et al. (2013) create an MCMC algorithm with several MH steps. Variational approximation of the posterior is sometimes preferred to standard MCMC sampling due to the ability to handle far larger graphs; these have been used for the standard binary SBM (Hofman and Wiggins 2008) and also the mixed-membership SBM (Airoldi et al. 2008), the weighted SBM (Aicher et al. 2015), and the more general random graph model of Schweinberger and Handcock (2015).

There are also non-statistical approaches to estimating the latent structure of the SBM, as mentioned previously. In addition to modularity-maximizing algorithms, particular focus

has been on applying spectral algorithms to graphs. For the K-block SBM, these algorithms involve calculating the spectral decomposition of either the adjacency matrix (Sussman et al. 2012) or the Laplacian ($L = D - A$, where D is a diagonal matrix of node degrees), then clustering the K largest eigenvectors using a method such as k-means. The statistical properties of these algorithms have been well-studied (see the next section); under certain conditions the eigenvectors of the graph Laplacian (adjacency matrix) asymptotically converge to a population graph Laplacian (adjacency matrix) (Rohe et al. 2011).

1.3.5 Consistency and Model Selection

Consistency of the binary SBM has been well-studied with respect to what is known as detectability or, sometimes in the context of the community detection problem, the resolution limit. Fortunato and Barthélemy (2007) show that modularity optimization methods can fail to detect communities that are smaller than a value which depends on the total network size and the connectedness of separate communities. Abbe et al. (2014) give a partial list of methods for the binary SBM that have established bounds for two-community recovery, based on the within-community edge probability p and the between-community edge probability q; the EM approach of Snijders and Nowicki (1997) yields the optimal bound $p - q = \Omega(1)$ (i.e. $|p - q| > 0$ and the number of nodes $n \to \infty$). Abbe et al. find a sharp threshold for exact recovery with two groups (misclassified nodes $\to 0$): if we let $\alpha = pn/log(n)$ and $\beta = qn/log(n)$ be constant with $\alpha > \beta$, then exact recovery of the communities with high probability is only possibly if $\frac{\alpha+\beta}{2} - \sqrt{\alpha\beta} > 1$; maximum likelihood achieves this optimal threshold. Other authors have addressed this question for specific algorithms: maximum likelihood (Abbe et al. 2014, Celisse et al. 2012, Choi et al. 2012, Zhao et al. 2012), Bayesian estimation (Decelle et al. 2011, Mariadassou and Matias 2015), spectral methods (Chin et al. 2015, Chen and Hero III 2015, Lei and Rinaldo 2015, Rohe et al. 2011, Sussman et al. 2012), other algorithmic approaches (Zhao et al. 2012, Gao et al. 2015), and general

results applicable to all algorithms (Mossel et al. 2014). While the thresholds determined for various estimation methods and various conditions (e.g. growing number of classes, sparse graphs) differ in precise form, the common asymptotic behavior is a threshold past which recovery is impossible, a region in which recovery is highly probable, and then a region in which a given method can recover the block structure exactly. This has apparently not been studied yet for weighted SBM, but as in all mixture models a similar behavior would be expected - the closer together the mixture distributions are, the more difficult the classification problem will be.

In the pursuit of uncovering the correct latent block structure for the SBM, the issue of selecting the correct number of groups arises. For many of the estimation methods for the SBM, this is assumed known (and then often assessed *a posteriori*), though in practice it rarely is. While criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) have been suggested in places, Biernacki et al. (2000) state that the regularity conditions under which the BIC approximates the integrated likelihood (the marginal likelihood of the model of interest) do not hold for the problem of determining the number of components in a mixture model. They proposed the Integrated Completed Likelihood (ICL) criterion, which has commonly been used in some form for the SBM (Daudin et al. 2008, Mariadassou et al. 2010, Matias and Miele 2015); Côme and Latouche (2015) introduce a similar expression, which they call the integrated complete data log likelihood. Specifically for variational Bayesian estimation of the SBM, Latouche et al. (2012) developed the Integrated Likelihood Variational Bayes criterion. Some other methods for SBM model selection include: McDaid et al. (2013) put a prior on the number of clusters and estimate the best number within their Bayesian formulation (see also Herlau et al. (2014) for a similar approach); Lei (2014) developed a goodness-of-fit test statistic and its asymptotic distribution based on random matrix theory; Wang and Bickel (2015) have recently proposed an approach for model selection based on likelihood-ratio testing.

CHAPTER 2: LCN: A RANDOM GRAPH MIXTURE MODEL FOR COMMUNITY DETECTION IN FUNCTIONAL BRAIN NETWORKS

2.1 Introduction

Community structure has been studied in brain networks based on functional magnetic resonance imaging (fMRI) studies (Bullmore and Sporns 2009), but most of the focus with regards to statistical modeling for fMRI networks has been on defining brain nodes and network estimation techniques, and further downstream the analyses typically have involved network summary measures and the use of ERGMs (Simpson et al. 2011). The stochastic block model and other similar random graph models have apparently not been utilized in studying brain connectivity, despite the strengths of this approach in this application. The SBM not only captures the important community structure of a network but also simultaneously quantifies its modularity, all within a statistically-principled framework. In this chapter we present a Bayesian formulation of the affiliation SBM, followed by several extensions in the next chapters to settings where there are multiple networks and questions about the population of networks, such as in brain connectivity studies with collections of fMRI networks.

Our motivating data are resting-state functional magnetic resonance images (fMRI) from the ADHD-200 sample, which is downloadable from `http://fcon\_1000.projects.nitrc.org/indi/adhd200`. For each of 185 subjects, we estimate resting-state brain networks via Pearson correlation across 172 fMRI images per subject. See Figure 2.1 for the networks of two randomly selected subjects, which have been visualized with the BrainNet Viewer (`http://www.nitrc.org/projects/bnv/`) (Xia et al. 2013). Our question of interest is how

these networks are organized into functionally-overlapping groups of brain regions, in both healthy children and those with ADHD.

Figure 2.1: Functional networks and estimated latent structure from 2 subjects

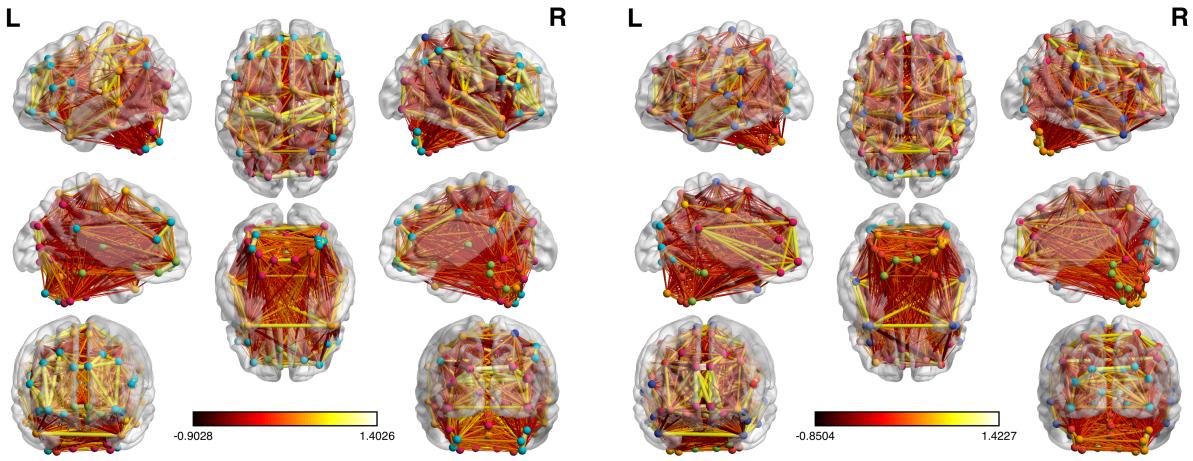


Figure 2.1. Functional brain networks for subject 1 (L) and subject 2 (R). There are 116 brain regions in each image. Different colored nodes indicate different estimated latent classes, but colors are not comparable between subjects.

We fit the affiliation SBM to each of these weighted networks as a Bayesian hierarchical random graph model and estimate the latent structure, including the graphical parameters governing the edge weights. We also compare the performance of our approach, which we call latent class network (LCN) estimation, to the variational EM method of Ambroise and Matias (2012) for the same model and to the Bayesian implementation of the weighted SBM in the R package *hergm* (R Core Team 2015, Schweinberger and Luna 2015) for several simulations schemes; we find that our efficient Gibbs sampling algorithm using conjugate priors is more accurate in classifying nodes into the correct groups than both of these methods and is faster than the other Bayesian approach.

2.2 Methodology

2.2.1 Random Graph Mixture Model

Let $\mathbf{Y} = (Y_{ij})$ denote an observed undirected graph with n nodes, where Y_{ij} denotes the weighted edge value between node i and node j. We assume that the n nodes each fall into one of Q latent classes, with the unobserved class label of node i given by the random vector $Z_i = (Z_{i1}, \ldots, Z_{iQ})$, where $Z_{iq} = 1$ indicates that node i is in the q–th group. Following the version of the SBM in Ambroise and Matias (2012), our RGMM consists of:

- (i) A latent class model for characterizing the class label Z_i for each node $i = 1, \ldots, n$.
- (ii) A measurement model for characterizing the conditional distribution of Y_{ij} given $\{Z_i, Z_j\}$.

We assume that the latent classes $\{Z_i\}$ are independently and identically distributed as Multinomial random variables with the probability vector $\pi = (\pi_1, \ldots, \pi_Q)$ such that $0 \leq \pi_q \leq 1$ and $\sum_q \pi_q = 1$. The measurement model is a two-component mixture model: we assume that Y_{ij} conditional on $\{Z_i\}_{1 \leq i \leq n}$ are independent and the conditional distribution of Y_{ij} given $Z_{ic} \cdot Z_{jd} = 1$ is given by

$$p_{cd} f(\cdot; \theta_{cd}) + (1 - p_{cd}) \delta_0(\cdot) \text{ for } i, j = 1, \ldots, n, \tag{2.1}$$

where $f(\cdot; \theta_{cd})$ is a prefixed probability distribution with an unknown parameter vector θ_{cd} and $\delta_0(\cdot)$ denotes the Dirac measure at zero accounting for non-present edges. By assuming that the edge values are conditionally independent given the latent classes of the nodes, the (marginal) dependencies of the graph are fully determined by the latent community structure.

Furthermore, we impose the affiliation SBM by reducing the $Q \cdot (Q + 1)$ parameters in

$\{p_{cd}\}_{1\leq c\leq d\leq Q}$ and $\{\theta_{cd}\}_{1\leq c\leq d\leq Q}$ to:

$$p_{cd} = \begin{cases} p_{in} & \text{if } c = d, \\ p_{out} & \text{if } c \neq d, \end{cases} \quad \text{and } \theta_{cd} = \begin{cases} \theta_{in} & \text{if } c = d, \\ \theta_{out} & \text{if } c \neq d. \end{cases} \tag{2.2}$$

Use of this parameterization allows us to avoid the typical problem of label switching/swapping in Bayesian mixture modeling. When non-symmetric priors are used for the group proportions, the nonidentifiability of the order of the latent classes of nodes can lead to the class labels changing between successive MCMC samples and make posterior inference difficult. The affiliation SBM does not have class-specific parameters, so the sampler arbitrarily sets the order in the initialization step and then sampling proceeds without label swapping.

This framework is flexible and can model directed graphs by utilizing a bivariate distribution for $f(Y_{ij}, Y_{ji})$ and allowing $p_{cd} \neq p_{dc}$ and $\theta_{cd} \neq \theta_{dc}$. Here we focus on Gaussian-weighted edges, such that $\theta = (\theta_{in}, \theta_{out}) = (\mu_{in}, \tau_{in}, \mu_{out}, \tau_{out})$, but we can easily incorporate different distributions for the edge distribution $f(\cdot)$. We can also adapt the model for more complex latent structures such as the overlapping SBM and correlated latent groups by alterations within this hierarchical formulation.

2.2.2 Prior distributions

Priors are chosen to preserve conjugacy to allow for efficient MCMC estimation as follows:

$$\begin{aligned} &\pi|Q \sim \text{Dirichlet}(a_1, \ldots, a_Q), && p_{in}, p_{out} \sim \text{Uniform}(0, 1), \\ &\mu_{in}|\tau_{in} \sim N(\mu_{0,in}, \tfrac{\sigma^2_{0,in}}{\tau_{in}}), && \mu_{out}|\tau_{out} \sim N(\mu_{0,out}, \frac{\sigma^2_{0,out}}{\tau_{out}}), \\ &\tau_{in} \sim \text{Ga}(\alpha_{0,in}, \beta_{0,in}), && \tau_{out} \sim \text{Ga}(\alpha_{0,out}, \beta_{0,out}), \end{aligned}$$

where $\text{Ga}(a, b)$ is a gamma distribution with shape a and rate b. To achieve relatively flat priors, we set the hyperparameters to be: $a_1, \ldots, a_Q = 1$, $\mu_{0,in} = \mu_{0,out} = 0$, $\sigma^2_{0,in} = \sigma^2_{0,out} = 10$, and $\alpha_{0,in} = \alpha_{0,out} = \beta_{0,in} = \beta_{0,out} = 0.01$.

2.2.3 Estimation

We utilize a Gibbs sampler for posterior computation, with all full conditional posterior distributions given in the Appendix. The Gibbs sampler involves sampling from a series of conditional distributions while each of the components is updated in turn. Our Gibbs sampler starts as follows:

- Initialize $\pi_q^{(0)} = 1/Q$ for $q = 1, \ldots, Q$.
- Sample $Z_i^{(0)}$ from Dirichlet$(\pi^{(0)})$ for $i = 1, \ldots, n$.
- Initialize $p_{in}^{(0)} = p_{out}^{(0)} = \frac{1}{2}, \mu_{in}^{(0)} = \mu_{out}^{(0)} = 0, \tau_{in}^{(0)} = \tau_{out}^{(0)} = 1$.

Then for $t = 1, \ldots, N$, we sequentially update all parameters as follows:

- Sample $\pi^{(t)}$ from $P(\pi|Q, Y, Z^{(t-1)})$.
- For $i = 1, \ldots, n$, sample $Z_i^{(t)}$ from

$$P(Z_i|Q, Y, Z_{1:(i-1)}^{(t)}, Z_{(i+1):n}^{(t-1)}, \pi^{(t)}, p_{in}^{(t-1)}, p_{out}^{(t-1)}, \mu_{in}^{(t-1)}, \mu_{out}^{(t-1)}, \tau_{in}^{(t-1)}, \tau_{out}^{(t-1)}).$$

- Sample $\mu_{in}^{(t)}$ from $P(\mu_{in}|Q, Y, Z^{(t)}, \tau_{in}^{(t-1)})$ and $\mu_{out}^{(t)}$ from $P(\mu_{out}|Q, Y, Z^{(t)}, \tau_{out}^{(t-1)})$.
- Sample $\tau_{in}^{(t)}$ from $P(\tau_{in}|Y, Z^{(t)}, \mu_{in}^{(t)})$ and $\tau_{out}^{(t)}$ from $P(\tau_{out}|Y, Z^{(t)}, \mu_{out}^{(t)})$.
- Sample $p_{in}^{(t)}$ from $P(p_{in}|Y, Z^{(t)})$ and $p_{out}^{(t)}$ from $P(p_{out}|Y, Z^{(t)})$.

To improve sampling performance, we run multiple MCMC chains and use the Integrated Completed Likelihood (ICL) criterion to automatically select the chain that maximizes ICL

(Mariadassou et al. 2010, Daudin et al. 2008). For a graph with n nodes, the ICL criterion is given by:

$$\text{ICL}_Q \quad = \quad \max_{\theta} \log \mathbb{P}(Y, \tilde{Z}|Q, \theta) \; - \; \frac{1}{2}[P_Q \; \cdot \; \log\left(\frac{n(n-1)}{2}\right) \; + \; (Q \; - \; 1) \; \cdot \; \log(n)], \quad (2.3)$$

where $\tilde{Z}$ denotes the predictions for the latent Z and P_Q denotes the number of independent parameters. In this case, we have $\theta = (p_{in}, p_{out}, \mu_{in}, \mu_{out}, \tau_{in}, \tau_{out})$ and $P_Q = 6$. Moreover, we plug in the univariate mode of each parameter into ICL_Q. This amounts to maximizing the observed data likelihood when comparing two MCMC chains with Q and n fixed.

To achieve better sampling performance for large graphs, we propose using spectral clustering to estimate the initial value of the latent structure $Z^{(0)}$; we can use the k-means clustering algorithm (Hartigan and Wong 1979) to cluster all n nodes into Q groups according to the first Q eigenvectors of a graph. Moreover, the diagnostic tools in the *coda* R package (Plummer et al. 2006) can be used to assess posterior convergence.

2.3 Simulations

We carried out simulations to examine the finite sample performance of the LCN RGMM in detecting the community structure of simulated networks and quantify their network modularity.

2.3.1 Simulation setup

We simulated networks as follows: for a given Q^*, π was randomly generated from Dirichlet$(a_1, \ldots, a_{Q^*})$, and then each Z_i for $i = 1, \ldots, n$ was independently generated from Multinomial$(\pi_1, \ldots, \pi_{Q^*})$. The data Y_{ij} were generated from a mixture of zero-valued edges, randomly drawn from either Bernoulli$(1 - p_{in})$ or Bernoulli$(1 - p_{out})$ distributions and either Normal$(\mu_{in}, \tau_{in}^{-1})$ or Normal$(\mu_{out}, \tau_{out}^{-1})$, depending on whether nodes i and j are in

the same latent class. We set hyper-parameters $\sigma^2_{0,in}$ and $\sigma^2_{0,out}$ to one. The parameters $p_{in}, p_{out}, \mu_{in}, \mu_{out}, \tau_{in},$ and τ_{out} were fixed at various values in order to examine the finite sample performance of LCN and the associated MCMC algorithm as modularity measures change.

We considered six schemes and simulated 200 independent graphs for each scheme. Simulation schemes are listed in Table 2.1. Scheme 1 is an example of a relatively easy community detection problem with $p_{in} >> p_{out}$ and $\mu_{in} >> \mu_{out}$. Scheme 2 is a much harder problem with decreased distance between mixture distributions and fully dense graphs (no zero edges). Schemes 3 and 5 were designed to test performance when the number of latent groups is misspecified. Scheme 4 represents a scenario with a large number of nodes. Scheme 6 is a scenario with a relatively large number of smaller latent groups.

For each graph, we ran two independent chains of the Gibbs sampler and then used ICL to choose the best chain as described previously. We also compared our method with two competing methods including the approximating method of Ambroise and Matias based on a composite likelihood (Ambroise and Matias 2012) and the Bayesian implementation of the original SBM of Nowicki and Snijders in the *hergm* R package (Nowicki and Snijders 2001, R Core Team 2015, Schweinberger and Luna 2015). To deal with the label switching phenomenon seen in the *hergm* output, MCMC samples were relabeled with the use of the loss function from Peng and Carvalho (2013), which is included in the R function *hergm.postprocess*.

2.3.2 Results

Classification is typically accurate under all of the simulation schemes, as shown via box plots of the misclassification rates in Figure 2.2, though expectedly less so with more similar mixture distributions. The most probable classes were estimated from the 10,000 MCMC samples for each simulation, and the misclassification rate was estimated as the sum

Table 2.1: Simulation schemes

Sim	n	$Q*$	Q (est)	p_{in}	p_{out}	μ_{in}	μ_{out}	τ_{in}	τ_{out}
1	50	3	3	0.8	0.3	1	-1	1	1
2	50	3	3	1	1	0.5	-0.3	0.2	0.4
3	50	3	5	0.8	0.3	1	-1	1	1
4	500	3	3	0.8	0.3	0.5	-0.5	1	1
5	100	5	10	0.8	0.3	0.5	-0.5	1	1
6	50	10	10	0.8	0.3	0.5	-0.5	1	1

Table 2.1. 200 datasets were simulated from each of these schemes, then analyzed using 2 MCMC chains, and the chain with the greatest ICL was selected.

of false positives (nodes estimated to be in the same community when they are not) and false negatives (nodes estimated to be in different communities when they are in the same) divided by the total number of possible latent connections ($n \cdot (n-1)/2$). Most misclassification occurred in MCMC chains that did not converge to the true posterior distribution, which is seen in the tails of the box plots - many of these incorrectly estimated a single latent class containing all the nodes.

All the methods do well for the "easy" community detection problem (Scheme 1). Our method outperformed the other two methods for the selection of the true number of groups when more groups were specified (Schemes 3 and 5). The approximating method of Ambroise and Matias fares well with a large number of nodes (Scheme 4), but it is not as accurate for smaller graphs (Schemes 2 and 6). The Bayesian method (Nowicki and Snijders 2001, R Core Team 2015, Schweinberger and Luna 2015) is approximately exact, but it involves a computationally intensive algorithm for solving the label switching problem, which adds another level of error in estimating the latent structure, especially in the difficult Scheme 2.

In our estimation method, when the MCMC chain converges to the true distribution, estimation of the other parameters is accurate. Figure 2.3 shows the absolute deviation from between the posterior median and the true parameter value, scaled by the magnitude of the

Figure 2.2: Misclassifcation rates by simulation scheme

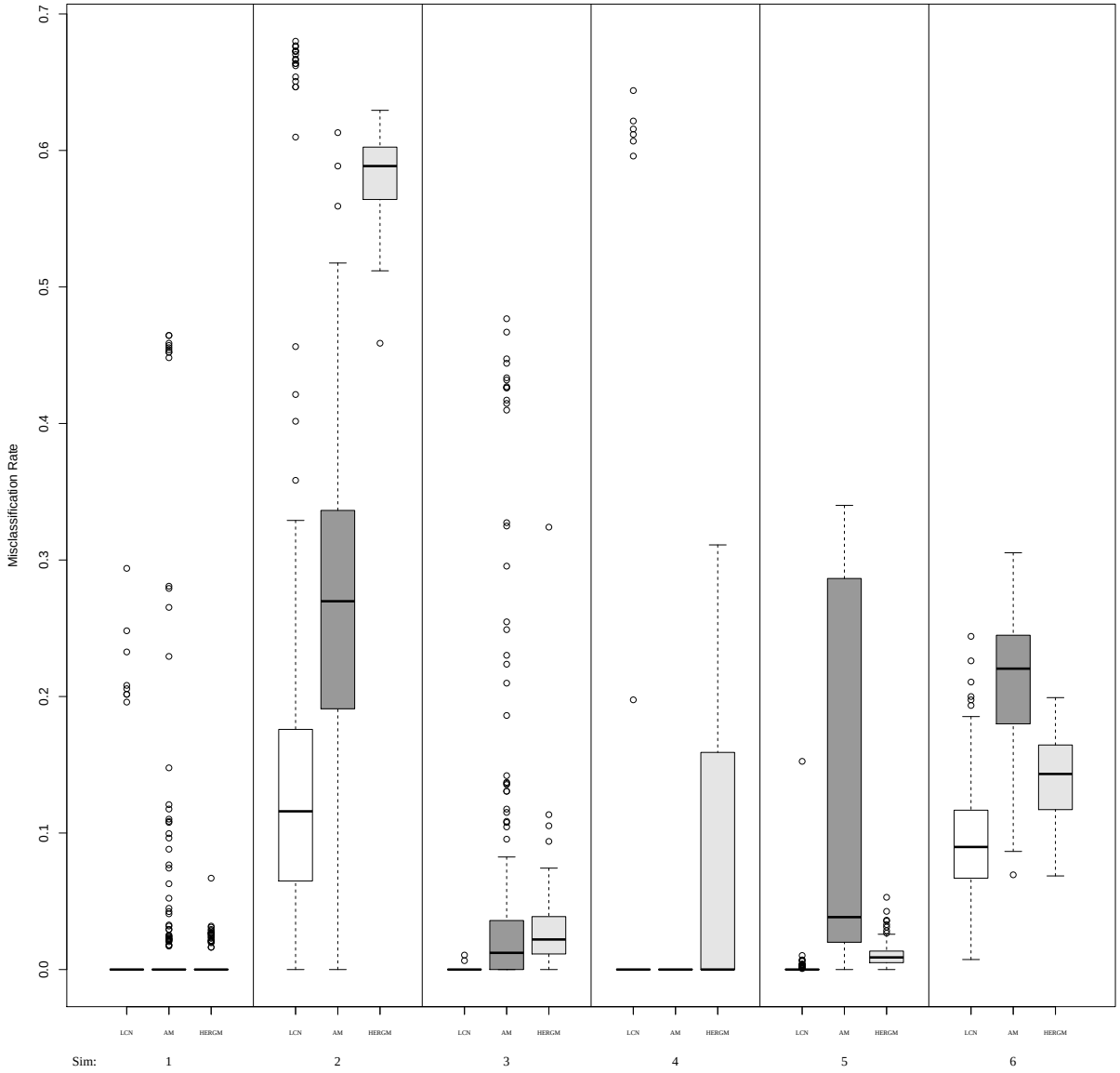


Figure 2.2. Boxplots of misclassification rates by simulation scheme. The 6 schemes, each with 200 simulated datasets, are listed in Table 2.1. Misclassification rate is defined as the sum of the false positives and false negatives divided by the total number of possible node pairs. AM is the implementation of SBM by Ambroise and Matias and HERGM is the implementation of the SBM in the *hergm* R package.

parameter. For the edge parameters $(\mu_{in}, \mu_{out}, \tau_{in}, \tau_{out})$ estimated in both our formulation and the parameterization used in (Ambroise and Matias 2012), our approach typically has

less estimation error. In the appendix, Figure A.1.1 gives coverage of the 95% highest posterior density (HPD) regions for the edge parameters, and Figure A.1.2 shows the median HPD width across the 200 simulations in each scheme. Coverage is near 95% for most edge parameters, except for p_{in} and p_{out} in Scheme 2 - in which the parameters are on the boundary of the parameter space. For large graphs, the decreasing HPD widths indicate efficient estimation of the edge parameters (see Schemes 4 and 5 in Figure A.1.2).

2.4 ADHD-200 resting-state fMRI networks

2.4.1 Functional brain networks

The resting state fMRI scans were acquired using a Siemens Allegra 3T scanner for six minutes (voxel size = $3 \times 3 \times 4$mm, slice thickness = 4mm, number of slices=33, TR=2s, TE=15ms, flip angle=90°, field of view=240mm). The Athena pipeline was applied for data preprocessing and the images were band-pass filtered within a frequency range of $(0.009, 0.08)$hz. The automated anatomical labeling (AAL) template (Tzourio-Mazoyer et al. 2002) was used to split patients' brains into 116 non-overlapping regions of interest (ROIs); blood-oxygen-level dependent (BOLD) contrast signals were averaged within each region for each of 172 time points, and a Pearson correlation matrix was estimated for each subject's 116 ROI × 172 time point matrix. Subsequently, the elements in each 116 x 116 matrix were transformed to approximate normality via the Fisher transformation, $z = 0.5 \times \ln(\frac{1+r}{1-r})$. Additionally, the Fisher-transformed correlation matrices were thresholded at ± 0.1 (which corresponds to $r \sim \pm 0.1$) to allow for some level of sparsity.

2.4.2 Results

We applied our RGMM to each subject's weighted network as follows: two parallel MCMC chains of our Gibbs sampling algorithm were run for each of Q = 3, 6, 9, and 12, and then ICL was used to choose the best of the 8 chains, which allowed for anywhere between 1

Figure 2.3: Scaled absolute deviation between posterior median and true value

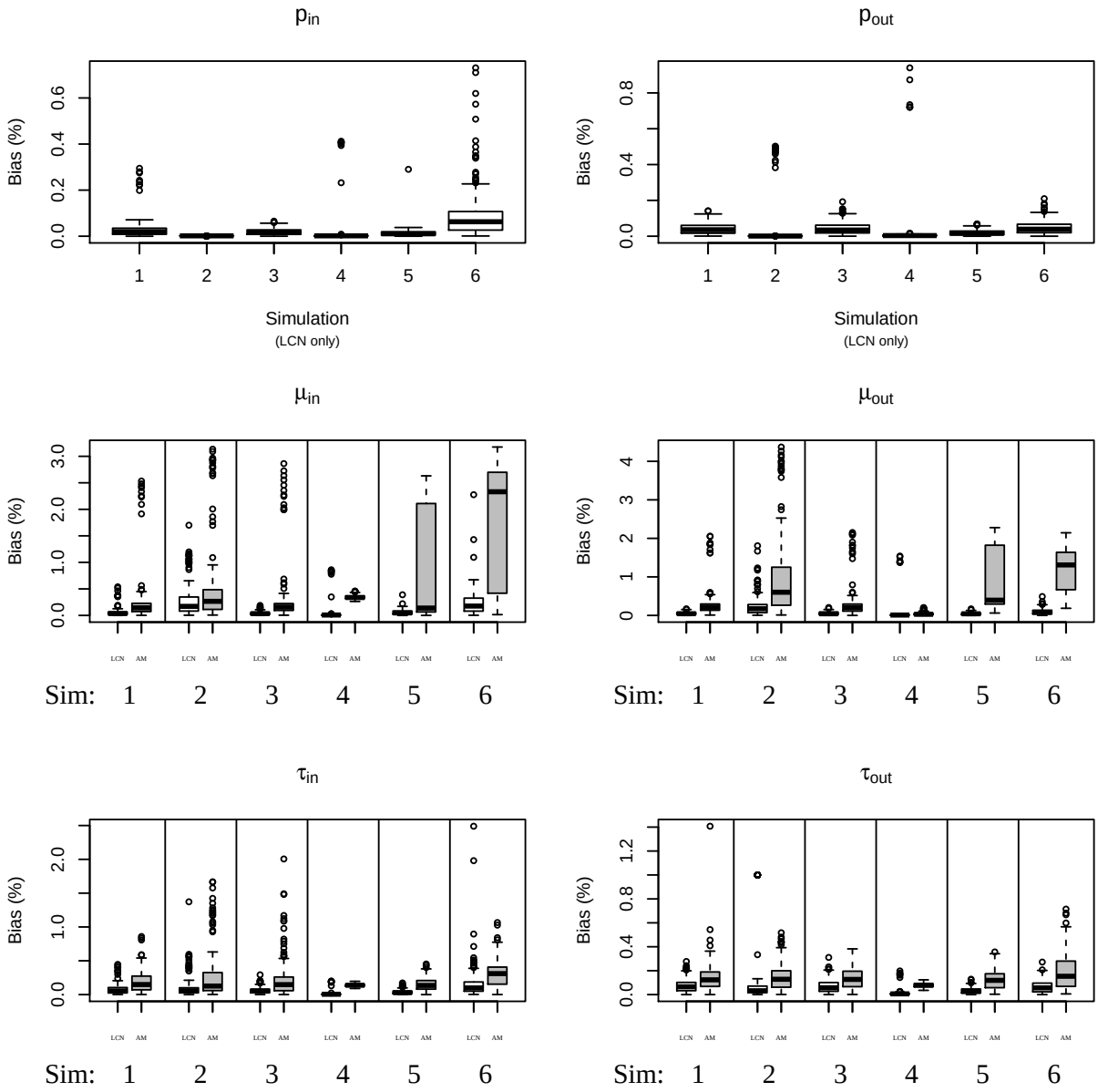


Figure 2.3. Absolute deviation between the posterior median of each parameter and the true value, scaled by the true value, from each of the 6 schemes listed in Table 2.1. For the edge parameters $\mu_{in}, \mu_{out}, \tau_{in}, \tau_{out}$, results from the Bayesian random graph model (LCN - on the left of each panel) are compared to the method of Ambroise and Matias (AM - on right).

and 12 latent classes for each subject. Figure 2.1 shows the estimated latent classes of the 116 ROIs for two randomly selected subjects as the color of nodes in the networks; subject

1 (L) has 7 latent classes of regions and subject 2 (R) has 8 latent classes of regions. To assess the overlap of the community structures of the two subjects, the adjusted Rand Index (Hubert and Arabie 1985) between the two clusterings was estimated to be 0.182, which is significantly different from zero (which would indicate no overlap at all) via permutation testing (10,000 permutations of the class labels, $p < 0.001$). So, while the latent community structure for these two subjects is different, there is significant overlap between them, which suggests that there may be a shared latent structure and individual deviations from this structure.

Figure 2.5 shows the estimated number of latent classes across the 185 subjects, with values ranging from 2 to 10. In Figure 3.5, the overlap of the latent structures of all 185 subjects is shown; the node pairs in red are those that are in the same latent class in most networks. The node pairs that appear in the same latent class in a large proportion of the subjects can be considered as the shared structure, while other node pairs that more frequently are in different latent classes indicate deviations from this shared structure. Also, the posterior distributions of the modularity parameters appear to vary across many of the subjects, indicating heterogeneity in latent community structure even beyond the latent class membership of the 116 ROIs. See Figure 2.4 for posterior samples of the modularity parameters of the two subjects from Figure 2.1.

2.5 Discussion

We have developed the weighted affiliation SBM as a Bayesian RGMM. Our RGMM utilizes an intuitive hierarchical parametric framework that accurately captures the affiliation community structure in simulated data. The benefits of using this fully Bayesian framework include incorporation of prior data, the ability to characterize the entirety of the posterior distribution, as well as the validity of estimates and accurate classification with smaller graphs. Additionally, this approach yields estimates of the modularity of the network as

Figure 2.4: Modularity estimates from 2 subjects

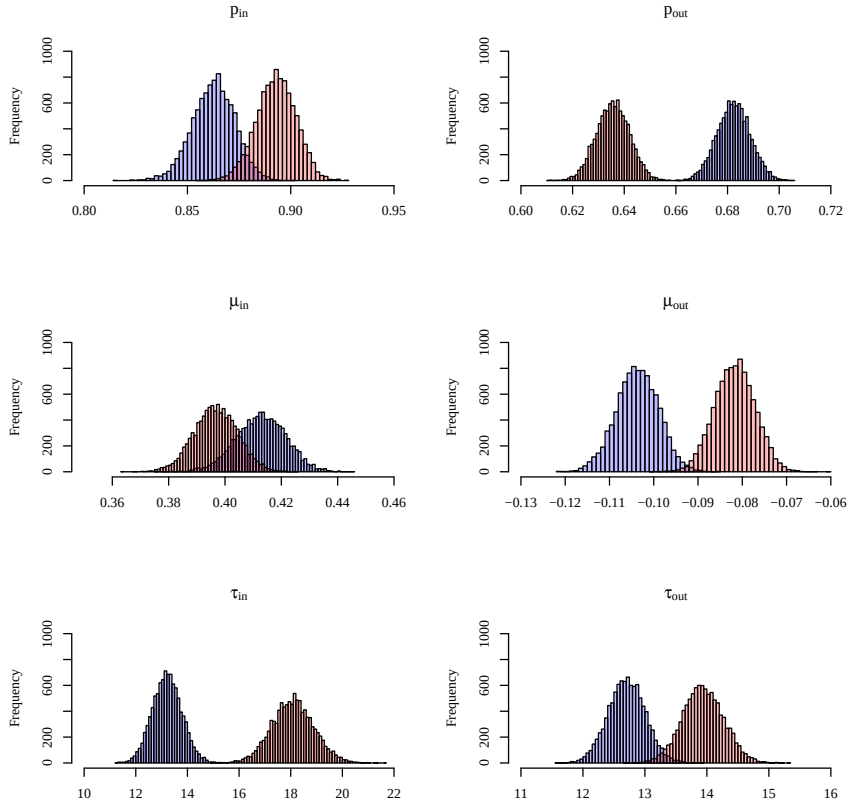


Figure 2.4. Posterior estimates of modularity parameters (Sparsity: p_{in} and p_{out}, edge weights: $\mu_{in}, \mu_{out}, \tau_{in}, \tau_{out}$) for subject 1 (Blue) and subject 2 (Red). First 100 samples were dropped, 9900 samples of each parameter shown.

parameters in the model. For highly modular graphs, in which nodes in one latent class have considerably more connections and different weights as compared to nodes in different classes, our estimation method performs well with minimal misclassification and accurate estimates of the parameters.

Within the 185 functional brain networks from the ADHD-200 sample, subjects were estimated to have between 2 and 10 latent classes of brain regions, but considerable overlap in the latent structure is seen between some subjects. The commonalities between subjects appear to include some level of symmetry in the latent classes across the left and right hemispheres, as well as the functional overlap in the regions of the occipital lobe (see the

Figure 2.5: Number of latent classes selected

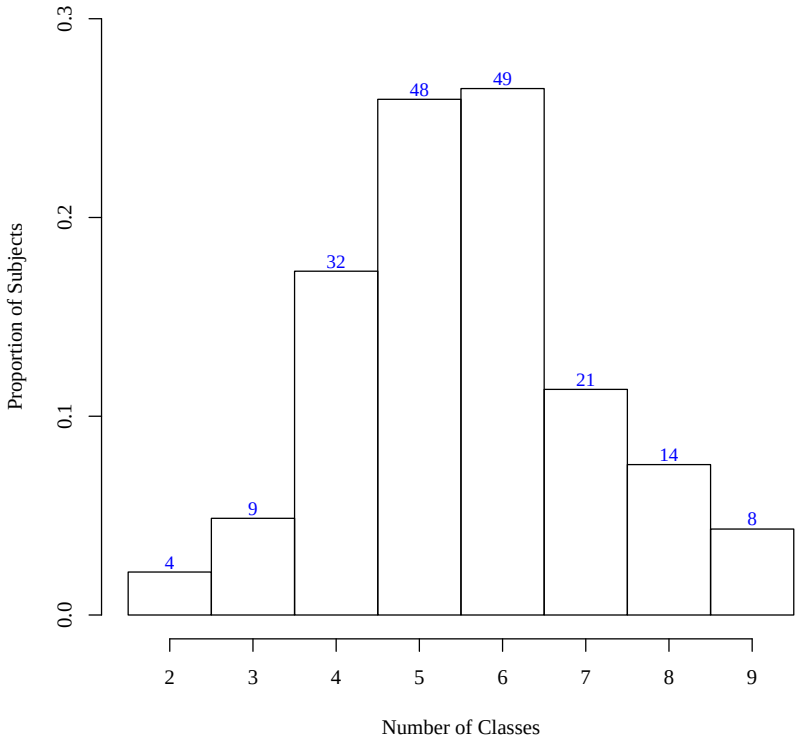


Figure 2.5. Number of latent classes of brain regions selected across 185 subjects from the ADHD-200 sample.

red region near the center of the diagnoal of Figure 3.5). Other relationships between brain regions appear to be conserved, in some cases within all subjects and in others more so within a particular diagnosis group (see Figure A.1.3 in the appendix for group-specific overlap of latent classes).

This framework allows the flexibility to utilize different distributions for the edge weights, detect overlapping communities, and estimate the community structure in directed graphs, all by straightforward alterations to the model. The following chapter is focused on extending this model to allow for groups of subjects that share a common structure, which appear

Figure 2.6: Overlap of communities

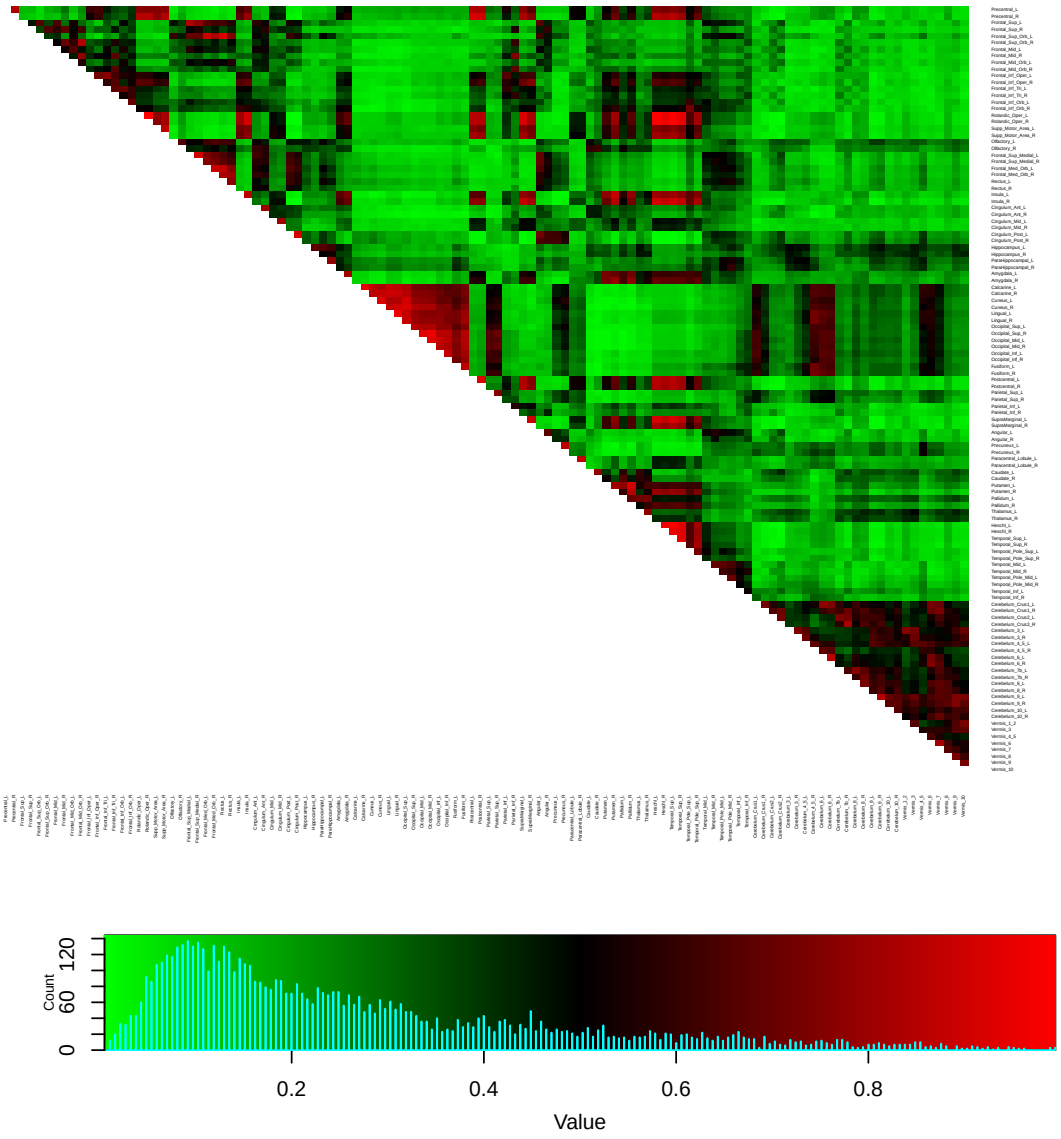


Figure 3.5. Overlap of the latent class structure across 185 subjects from the ADHD-200 sample. Each element of the matrix is the proportion of all 185 subjects in whom the corresponding two nodes fall in the same estimated latent class.

plausible based on our analyses of the resting-state fMRI networks from the ADHD-200 sample.

CHAPTER 3: A RANDOM GRAPH MIXTURE MODEL FOR ESTIMATING THE SHARED LATENT STRUCTURE IN GROUPS OF BRAIN NETWORKS

3.1 Introduction

While we can apply the weighted SBM to a single network and estimate the latent block structure underlying the observed configuration of edge values, in fMRI studies there are often multiple subjects and questions about a population of interest. We have previously observed heterogeneity in brain structure, but it is also apparent that some relationships between brain regions and some larger-scale structural characteristics are conserved within groups of individuals. In Figure 3.1 we see two groups of three graphs each, where the same latent block structure is generating the graphs in each group. When we observe a sample of weighted brain networks, it would be of interest to uncover these shared characteristics to better understand functional brain organization.

Here we extend the weighted SBM to model a collection of weighted graphs by fitting a random graph mixture model to classify the subjects into groups with similar graphical parameters. We adapt our Bayesian hierarchical model from the previous chapter and estimate the latent groups of subjects as well as the latent block structure for each individual's weighted network via an efficient Gibbs sampling algorithm. In Section 2, we present our hierarchical model and our estimation scheme. Then, in Section 3, we show the ability of this approach to accurately capture groups of subjects and classes of nodes in simulation studies. In Section 4, we apply our method to the resting-state fMRI networks mentioned previously, and finally in Section 5 we present a brief discussion of our results.

Figure 3.1: Shared Latent Graphical Structure

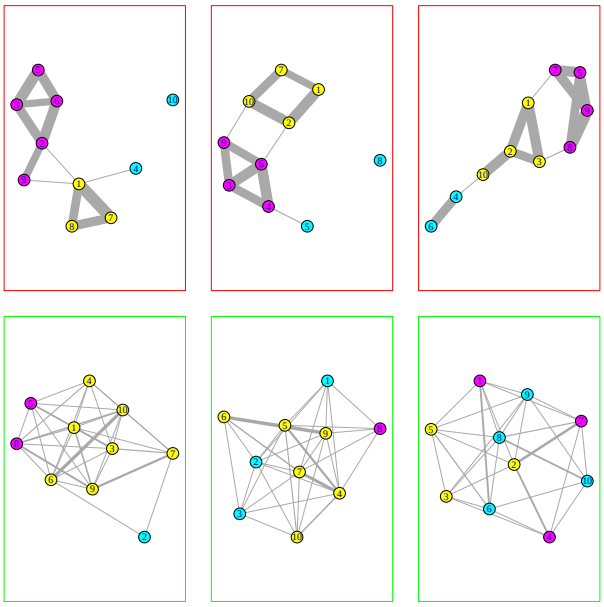


Figure 3.1. Two groups of networks with the same basic structure in each group.

3.2 Methodology

3.2.1 Hierarchical model

Let $Y_{n\times n(m)} = (Y_{ij})_m$ denote the observed undirected graph with n nodes for subject m, where $Y_{ij(m)}$ is the weighted edge value between node i and node j for that subject. We require that the n nodes are in common for all subjects $m = 1, \ldots, M$ (e.g. the same brain regions or the same genes) and assume that the M subjects fall into one of W latent groups of subjects that share the same parameters governing the latent block structure of an individual graph. We denote the latent group label for subject m by the random vector $V_m = (V_{m1}, \ldots, V_{mW})$, where $V_{mw} = 1$ indicates that subject m is in the w^{th} group. Then for each subject m the n nodes each fall into one of Q latent classes, with the unobserved class label of node i given by the random vector $Z_{i(m)} = (Z_{i1(m)}, \ldots, Z_{iQ(m)})$, where $Z_{iq(m)} = 1$ indicates that node i is in the q^{th} group for subject m. Our model consists of:

- (i) A latent class model for the group label V_m for each subject $m = 1, \ldots, M$.

- (ii) A latent class model for the class label $Z_{i(m)}$ for each node $i = 1, \ldots, n$ within each subject $m = 1, \ldots, M$.

- (iii) A measurement model for characterizing the conditional distribution of $Y_{ij(m)}$ given $\{V_m, Z_{i(m)}, Z_{j(m)}\}$.

We assume that the latent group labels $\{V_m\}$ are independently and identically distributed as Multinomial random variables with probability vector $\xi = (\xi_1, \ldots, \xi_W)$, such that $0 \leq \xi_w \leq 1$ and $\sum_w \xi_w = 1$. Subsequently, we assume that, conditional on the latent group for subject m, the latent class labels $\{Z_i\}_m$ are independently distributed as Multinomial random variables with class-specific and node-specific probability vectors $\pi_{i(w)} = (\pi_{i1(w)}, \ldots, \pi_{iQ(w)})$. Then, the conditional distribution of $Y_{ij(m)}$ given $V_m = w$ and $\{Z_{i(m)}, Z_{j(m)}\}$ is given by:

$$\left[p_{in(w)} f(\cdot; \theta_{in(w)}) + (1 - p_{in(w)})\delta_0(\cdot)\right]^{\sum_{q=1}^{Q} Z_{iq(m)} Z_{jq(m)}} \left[p_{out(w)} f(\cdot; \theta_{out(w)}) + (1 - p_{out(w)})\delta_0(\cdot)\right]^{1 - \sum_{q=1}^{Q} Z_{iq(m)} Z_{jq(m)}} \tag{3.1}$$

for $i, j = 1, \ldots, n$, and $m = 1, \ldots, M$. Here $f(\cdot; \theta_{(w)})$ is a probability distribution for the edge values with an unknown parameter vector $\theta_{(w)}$ and $\delta_0(\cdot)$ denotes the Dirac measure at zero accounting for non-present edges. We assume that the edge values are conditionally independent given the latent classes of the nodes, and so the dependencies of each graph and between graphs are fully determined by the latent community structures.

We focus on Gaussian-weighted edges, but this framework can accommodate different edge-distributions and various latent structures. Here we have proposed a more complicated latent structure than the standard SBM by allowing each node to have its own generating Dirichlet distribution, which is shrunken towards a common distribution for each group when we assume that each of these distributions has the same hyperparameters ($\alpha_{i(w)} = \alpha_w \ \forall \ i, w$); for a more parsimonious blockmodel we could assume that $\pi_{i(w)} = \pi_{(w)}$ for all $i = 1, \ldots, n$.

3.2.2 Prior distributions

Without loss of generality, we let $Y_{ij(m)}|\theta_{cd(w)} \sim N(\mu_{cd(w)}, \tau^{-1}_{cd(w)})$. Priors were chosen to preserve conjugacy and allow for efficient MCMC estimation:

$$\xi|a \sim \text{Dirichlet}(a_1, \ldots, a_W),\ \pi_{i(w)}|Q, \alpha_{(w)} \sim \text{Dirichlet}(\alpha_{1(w)}, \ldots, \alpha_{Q(w)}),$$
$$p_{in(w)} \sim \text{Beta}(\alpha_{p_{in(w)}}, \beta_{p_{in(w)}}),\ p_{out(w)} \sim \text{Beta}(\alpha_{p_{out(w)}}, \beta_{p_{out(w)}}),$$
$$\mu_{in(w)}|\tau_{in(w)} \sim N(\mu_{0,in(w)}, \frac{\sigma^2_{0,in(w)}}{\tau_{in(w)}}),\ \mu_{out(w)}|\tau_{out(w)} \sim N(\mu_{0,out(w)}, \frac{\sigma^2_{0,out(w)}}{\tau_{out(w)}}),$$
$$\tau_{in(w)} \sim \text{Ga}(\alpha_{0,in(w)}, \beta_{0,in(w)}),\ \tau_{out(w)} \sim \text{Ga}(\alpha_{0,out(w)}, \beta_{0,out(w)}),$$

where $\text{Ga}(a, b)$ is a gamma distribution with scale a and shape b. We specify flat priors by setting the hyperparameters to $a_1, \ldots, a_W = 1$, $\alpha_{1(w)}, \ldots, \alpha_{Q(w)} = 1$, $\mu_{0,in(w)} = \mu_{0,out(w)} = 0$, $\sigma^2_{0,in(w)} = \sigma^2_{0,out(w)} = 10$, $\alpha_{0,in(w)} = \alpha_{0,out(w)} = \beta_{0,in(w)} = \beta_{0,out(w)} = 0.01$, and $\alpha_{p_{in(w)}} = \beta_{p_{in(w)}} = \alpha_{p_{out(w)}} = \beta_{p_{out(w)}} = 1$, for $w = 1, \ldots, W$.

3.2.3 Estimation

We use a Gibbs sampling algorithm to obtain samples from the posterior distribution for the latent class multinomial variables (at the subject level and the node level) and the associated graphical parameters. To assist with convergence, we propose utilizing spectral clustering to initialize the MCMC chain. First, we initialize $V^{(0)}$ by clustering the eigenvalues from all the graphs, and then we initialize $Z^{(0)}_{(m)}$ for each subject $m = 1, \ldots, M$ by using a traditional spectral clustering algorithm and clustering the eigenvectors of that subject's graph. The algorithm is as follows:

- Initialize $\xi^{(0)}_w = \frac{1}{W}$ for $w = 1, \ldots, W$.
- Initialize $V^{(0)}$ by using k-means to find W clusters among the first W eigenvalues from each subject's graph.
- Initialize $\pi^{(0)}_{q(w)} = \frac{1}{Q}$ for $q = 1, \ldots, Q$ and $w = 1, \ldots, W$.

- Initialize $Z_{(m)}^{(0)}$ by applying k-means to find Q clusters among the first Q eigenvectors for $m = 1, \ldots, M$.
- Initialize $p_{in(w)}^{(0)} = p_{out(w)}^{(0)} = \frac{1}{2}, \mu_{in(w)}^{(0)} = \mu_{out(w)}^{(0)} = 0, \tau_{in(w)}^{(0)} = \tau_{out(w)}^{(0)} = 1$.

Then for $t = 1, \ldots, N$, we sample the parameters:

- Sample $\xi^{(t)}$ from $P(\xi|V^{(0)})$
- For $m = 1, \ldots, M$:
  - For $i = 1, \ldots, n$, sample $Z_{i(m)}^{(t)}$ from

$$P\big(Z_{i(m)}|W, Q, Y, V_m = w, Z_{1:[i-1](m)}^{(t)}, Z_{[i+1]:n(m)}^{(t-1)}, \pi_{(w)}^{(t-1)}, p_{in(w)}^{(t-1)}, p_{out(w)}^{(t-1)}, \mu_{in(w)}^{(t-1)}, \mu_{out(w)}^{(t-1)}, \tau_{in(w)}^{(t-1)}, \tau_{out(w)}^{(t-1)}\big)$$

  - Sample $V_m^{(t)}$ from

$$P(V_m|W, Q, Y, \xi^{(t)}, Z_{(m)}^{(t)}, \pi^{(t-1)}, p_{in}^{(t-1)}, p_{out}^{(t-1)}, \mu_{in}^{(t-1)}, \mu_{out}^{(t-1)}, \tau_{in}^{(t-1)}, \tau_{out}^{(t-1)})$$

- For $w = 1, \ldots, W$:
  - For $i = 1, \ldots, n$, sample $\pi_{i(w)}^{(t)}$ from $P(\pi_{i(w)}|Q, V^{(t)}, Z^{(t)})$
  - Sample $\mu_{in(w)}^{(t)}$ from $P(\mu_{in(w)}|Y, V^{(t)}, Z^{(t)}, \tau_{in(w)}^{(t-1)})$ and $\mu_{out(w)}^{(t)}$ from $P(\mu_{out(w)}|Y, V^{(t)}, Z^{(t)}, \tau_{out}^{(t-1)})$
  - Sample $\tau_{in(w)}^{(t)}$ from $P(\tau_{in(w)}|Y, V^{(t)}, Z^{(t)}, \mu_{in}^{(t)})$ and $\tau_{out}^{(t)}$ from $P(\tau_{out}|Y, V^{(t)}, Z^{(t)}, \mu_{out}^{(t)})$
  - Sample $p_{in(w)}^{(t)}$ from $P(p_{in(w)}|Y, V^{(t)}, Z^{(t)})$ and $p_{out(w)}^{(t)}$ from $P(p_{out(w)}|Y, V^{(t)}, Z^{(t)})$

The Gibbs sampler is typically able to collapse to the correct number of groups of subjects and the correct number of classes of nodes for each subject, when the posterior distribution is distinctly unimodal (i.e. in simulations). For more difficult, high-dimensional problems where the posterior may be multimodal, we propose using a variation of the the Integrated Classification Likelihood (ICL) (Biernacki et al. 2000) to assist in model selection. Modifying

the version of ICL used for a single graph (Daudin et al. 2008, Mariadassou et al. 2010), we define

$$ICL_{W,Q} = \max_{\theta} \log \mathbb{P}(Y, \tilde{V}, \tilde{Z}|W, Q, \theta) - \tfrac{1}{2}\{P_{W,Q} \cdot \log[M \cdot \tfrac{n(n-1)}{2}] + nW(Q-1) \cdot \log(n) + (W-1) \cdot \log(M)\},$$

where $P_{W,Q}$ is the length of θ, n is the number of nodes, M is the number of subjects, Q is the number of classes of nodes, and W is the number of groups of subjects. The first penalty term ($P_{W,Q} \cdot \log[Mn(n-1)]$) penalizes the number of edge parameters in the model, the second term ($nW(Q-1) \cdot \log(n)$) penalizes the number of latent classes of nodes, and the third term ($(W-1) \cdot \log(M)$) penalizes the number of latent groups of subjects.

3.3 Simulations

3.3.1 Simulation setup

To assess performance, groups of weighted networks were simulated as follows:

- For a given W, ξ is randomly generated from Dirichlet$(a_1, \ldots, a_W)$.
- Each V_m, $m = 1, \ldots, M$, is independently generated from Multinomial$(\xi_1, \ldots, \xi_W)$.
- For a given Q, each $\pi_{i(w)}$, $w = 1, \ldots, W$, is independently generated from Dirichlet$(\alpha_{1(w)}, \ldots, \alpha_{Q(w)})$.
- Each $Z_{i(m)}$, $i = 1, \ldots, n$ and $m = 1, \ldots, M$, is independently generated from Multinomial$(\pi_{i1(w)}, \ldots, \pi_{iQ(w)})$, dependent on the value of V_m.
- The data $Y_{ij(m)}$ are generated from a mixture of zero-valued edges, randomly drawn from either Bernoulli$(1 - p_{in(w)})$ or Bernoulli$(1 - p_{out(w)})$ distributions and either Normal$(\mu_{in(w)}, \tau^{-1}_{in(w)})$ or Normal$(\mu_{out(w)}, \tau^{-1}_{out(w)})$, depending on whether nodes i and j are in the same latent class (hyper-parameters $\sigma^2_{0,in(w)}$ and $\sigma^2_{0,out(w)}$ were set to one) and depending on the value of V_m.

The parameters $p_{in(w)}, p_{out(w)}, \mu_{in(w)}, \mu_{out(w)}, \tau_{in(w)}, \tau_{out(w)}$ were fixed at various values for $w = 1, \ldots, W$ to examine performance differences as modularity measures and the differences between groups change. For each of 5 simulation setups (listed in Table 3.1), 200 separate graphs were simulated, then analyzed with two chains of the Gibbs sampler; $ICL_{W,Q}$ was used to choose the best chain as described previously.

Table 3.1: Simulation schemes

Sim	n	W*	W (est)	Q*	Q (est)	p_{in}	p_{out}	μ_{in}	μ_{out}	τ_{in}	τ_{out}
1	50	2	2	2	2	(1,1)	(1,1)	(0.1,2.1)	(-0.1,1.9)	(1,1)	(1,1)
2	50	2	2	2	2	(1,1)	(1,1)	(0.5,0.6)	(-0.5,-0.4)	(1,1)	(1,1)
3	50	2	4	2	2	(0.8,0.7)	(0.3,0.2)	(0.1,0.2)	(-0.1,0)	(1,1)	(1,1)
4			2		4						
5			4		4						

Table 3.1. 200 datasets were simulated from each simulation scheme, then analyzed using 2 MCMC chains, and the chain with the greatest $ICL_{W,Q}$ was selected. For simulations 3-5, the same 200 datasets were analyzed with different input for W and Q, denoted by W(est) and Q(est).

3.3.2 Results

For each simulation, we have collected two MCMC chains of 10,000 samples each and estimated the maximum a posteriori (MAP) latent groups of subjects (V) and latent classes of nodes (Z), and then we have selected the MCMC chain with the greatest $ICL_{W,Q}$. After determining the pairwise connections implied by the estimated latent classes, we calculate the misclassification rate as the sum of the false positives (where a pair is estimated to be in the same latent class but is not) and the false negatives (where a pair is estimated to be in different classes when they are in the same class), divided by the total number of pairwise comparisons. Figure 3.2 shows the misclassification rates for each latent variable; simulation scheme 1 has easily distinguishable groups of subjects but difficult classification of nodes (i.e. no block structure), while simulation scheme 2 has easy classification of nodes (i.e. strong

block structure) but little difference between groups of subjects; for simulation schemes 3-5, we have simulated 200 datasets and misspecified W and Q to examine the sampler's ability to collapse to the correct number of groups of subjects and classes of nodes. Interestingly, estimation is more difficult only when W is misspecified but Q is not. As with other mixture models, the similarity of the mixture distributions is a primary determining factor of the ability of this approach to successfully recover the multinomial classes; the other main factor in performance is difficult convergence for this high-dimensional posterior distribution. In practice, issues with convergence can be dealt with by running parallel chains until several of them are in sufficient classification agreement or until diagnostics confirm convergence has been achieved.

3.4 Application to ADHD Data

The data consist of the same 185 fMRI correlation networks from the previous chapter. We aim to capture any latent groups of subjects that share a similar large-scale latent structure, as well as that latent structure for each subject, including the modularity parameters.

We ran our algorithm four times on our sample of 185 weighted functional brain networks, with two parallel MCMC chains for each of $\{W = 5,\ Q = 5\}$ and $\{W = 10,\ Q = 10\}$. We used $ICL_{W,Q}$ to select the "best" chain, which was found to have $W = 3$ groups of subjects, $Q = 5$ classes of nodes for 178 subjects and $Q = 4$ classes of nodes for 7 subjects. The parallel chains with $\{W = 5,\ Q = 5\}$ had over 90% agreement in the pairwise latent connections between subjects and over 93% agreement across all pairwise latent connections between subjects' brain regions, with at least 75% agreement for each subject.

Figure 3.3 shows the latent structures estimated for one randomly chosen subject from each latent group, out of 19,500 MCMC samples after a burn-in of 1,000 samples. While the latent classes are able to vary at the subject level, even within one latent group, the shared block structure governed by the edge parameters is distinct for the three estimated

Figure 3.2: Misclassifcation rates by simulation scheme

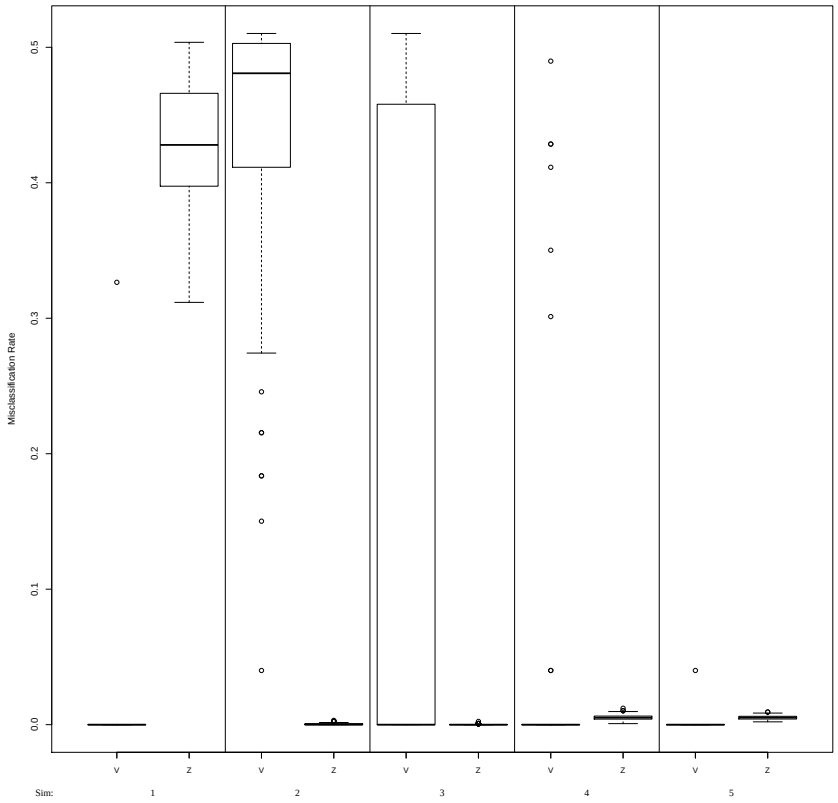


Figure 3.2. Boxplots of misclassification rates by simulation scheme. The 5 schemes, each with 200 simulated data sets, are listed in Table 3.2. Misclassification rate is defined as the sum of the false positives and false negatives divided by the total number of possible subject or node pairs. V denotes the subjects and Z denotes the nodes within each subject.

groups of subjects. Figure 3.4 gives the posterior samples for the edge parameters, showing that the three latent groups of subjects have differing modular brain structures. Group one, for example, has more positive correlation within a class of nodes and more negative correlation between classes of nodes than the other two groups, indicating a stronger block structure. Table 3.2 shows the overlap in the estimated latent groups of subjects with their ADHD diagnoses, and there is no clear relationship between diagnosis and the classification from the model (Fisher's exact test: p=0.22), indicating that there are functional differences amongst these subjects that may not be captured by the ADHD diagnoses. Figure 3.5 gives

the proportion of all 185 subjects in which each pair of brain regions is estimated to be in the same latent class, showing functional relationships that are conserved in most subjects (the node pairs which are in the same latent class across nearly all subjects, in red, and the node pairs which are in different latent classes across nearly all subjects, in green), and others that differ between subgroups of subjects. Figure B.1.1 in the Appendix shows the differences in these proportions between pairs of latent groups.

Figure 3.3: Three estimated latent groups.

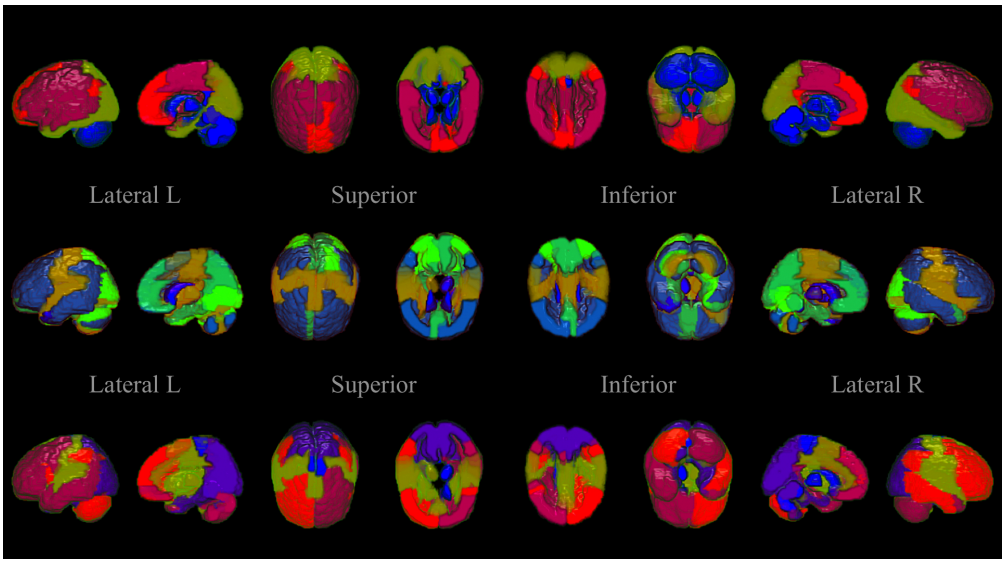


Figure 3.3. The estimated latent community structure for individual subjects from latent group 1 (Top), group 2 (Middle) and group 3 (Bottom). There are 116 brain regions in each image. Different colors indicate different latent classes, but colors are not comparable between subjects.

Figure 3.4: Posterior estimates of modularity parameters.

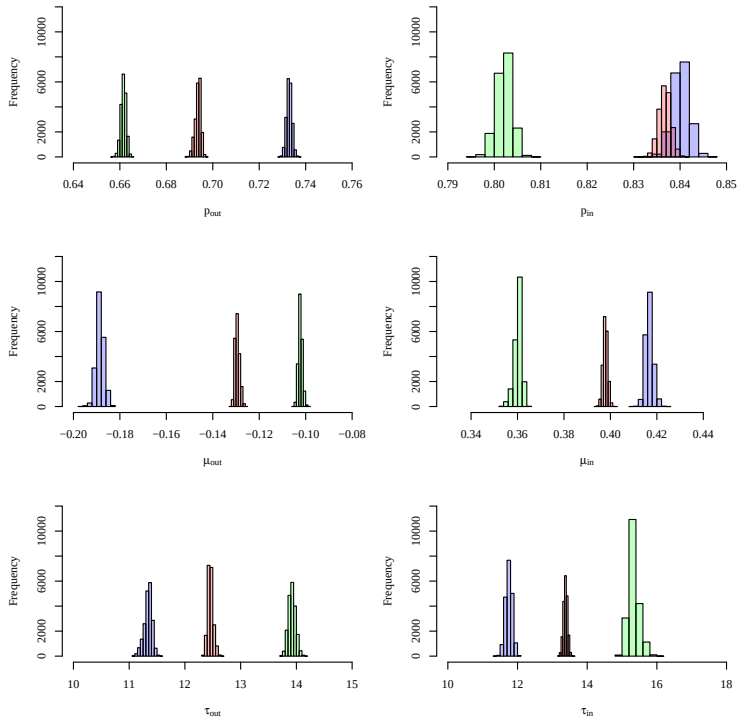


Figure 3.4. Posterior samples of edge parameters for the three latent groups (1=blue, 2=red, 3=green). A burn-in of 1,000 samples was discarded and the remaining 19,500 posterior samples for each group are included.

Table 3.2: Estimated latent group by diagnosis group

	Diagnosis			
Latent group	Control	Combined	Inattentive	Total
1	16	12	8	36
2	41	28	19	88
3	34	22	5	61
Total	91	62	32	185

Table 3.2. Maximum a posteriori latent group estimated from 19,500 MCMC samples, by ADHD diagnosis group.

Figure 3.5: Overlap of communities

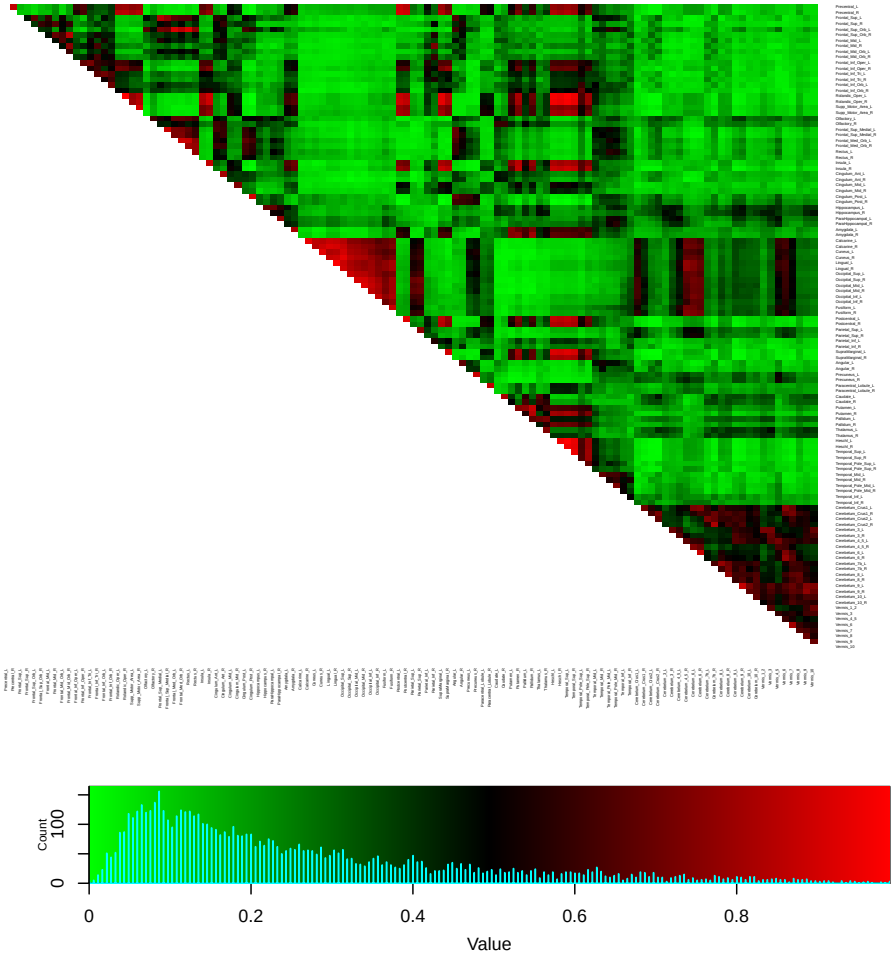


Figure 3.5. Overlap of the latent class structure across 185 subjects from the ADHD-200 sample. Each element of the matrix is the proportion of all 185 subjects in whom the corresponding two nodes fall in the same estimated latent class.

3.5 Discussion

We have extended the weighted stochastic block model to accommodate multiple networks, for situations in which we have the same network nodes defined for multiple subjects as we see in fMRI studies. The results from our simulation studies suggest this approach allows for accurate classification of both subjects and nodes within each subject, as well as estimation of the associated modularity parameters that govern the latent graphical structure. Use of our modified classification criterion and parallel MCMC chains can be combined with typical Bayesian convergence diagnostics to alleviate potential convergence issues that can arise in real data problems where the likelihood is not clearly unimodal. Our efficient estimation scheme, utilizing conjugate priors, allows for use on moderate-sized collections (the number of subjects M in the low hundreds) of moderate-sized networks (n in the hundreds or low thousands), such as in our motivating dataset where we have $M = 185$ subjects and $n = 116$ nodes in common across all subjects. In truly high-dimensional problems (large n, M, W, or Q), variational approximation would be necessary to avoid excessive computation times.

In applying this method to 185 fMRI networks from the ADHD-200 Sample, we uncover 3 groups of subjects, each of which has 4 or 5 estimated classes of brain regions. Our results indicate differences in modularity between these distinct groups of subjects, as well as heterogeneity in the latent connections between brain regions. The latent groups of subjects that we have uncovered do not precisely align with the ADHD diagnoses, which suggests that there are functional brain structures that involve other factors than ADHD. The following chapter is focused on adapting this hierarchical formulation to accommodate regression in the latent space, in order to assess whether differences between diagnosis groups, for example, are associated with differences in the latent network structure.

CHAPTER 4: A BAYESIAN HIERARCHICAL FRAMEWORK FOR INFERENCE ON MULTIPLE WEIGHTED NETWORKS WITH APPLICATIONS IN BRAIN IMAGING

4.1 Introduction

In the context of functional connectivity studies, the questions of interest may be focused on examining differences among subjects in the structure of their functional brain networks. There has been particular interest in understanding the effects of ADHD on functional connectivity, as previously mentioned in the first chapter. While network summary measures have been compared between small samples of ADHD patients and small control samples, most of these approaches have been ad hoc in their statistical methodology. In this chapter, our aim is to estimate the latent functional brain structures and examine the effects of covariates such as gender and the differences between ADHD and control groups in their latent functional brain structures.

We propose a novel adaptation of the SBM to estimate the latent structure in collections of weighted networks, with a regression component to assess the effects of covariates. Our model is formulated as a Bayesian hierarchical random graph mixture model (BHRGMM), and, as before, we utilize an efficient Markov Chain Monte Carlo (MCMC) algorithm with conjugate priors to obtain samples from the posterior distribution. We conduct simulation studies to examine the ability of our method to accurately capture the latent network structure across subjects as well as its accuracy in capturing differences in this structure explained by covariates. In the following section, we present the hierarchical model, the prior distributions, and the sampling algorithm, and then in Section 3 we show some results from simulation studies. In Section 4 we give results from analyzing the same 185 functional

brain networks from the previous chapters, and in Section 5 we present a brief discussion of our results.

4.2 Methodology

4.2.1 Hierarchical random graph model

Let the weighted edge between nodes i and j in subject m's network be denoted by $Y_{ij(m)}$. We assume all subjects $m = 1, \ldots, M$ have the same network nodes, such as in a brain network or a gene network. Then the distribution of each $Y_{ij(m)}$ depends on whether nodes i and j are in the same latent group for subject m, which is given by the binomial variable $Z_{ij(m)}$. We propose our hierarchical model as follows:

- (i) A latent model characterizing whether each pair of nodes, (i, j), where n is the total number of nodes in each graph and $1 \leq i < j \leq n$, are in the same functional group, indicated by $Z_{ij(m)} = 1$. This model incorporates regression in the latent space to account for the effects of covariates on the latent graphical structure.

- (ii) A measurement model for each weighted edge value $Y_{ij(m)}$, conditional on the value of $Z_{ij(m)}$.

This latent structure allows for more complexity than the standard SBM, where each node is a member of only one class (which is a subset of this formulation). We utilize probit regression by introducing another latent variable $Z^*_{ij(m)}$ and assume that

$$Z^*_{ij(m)} | x_m, \beta_{ij} \sim \mathrm{N}(x_m^\top \beta_{ij}, 1) \text{ and } Z_{ij(m)} = \begin{cases} 1, & \text{if } Z^*_{ij(m)} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where x_m denotes the covariate vector for subject m and β_{ij} denotes the parameter vector governing the latent connections between nodes i and j. We assume that the latent variables

$\{Z_{ij(m)}\}$ are conditionally independent, given covariates $\{x_m\}$ and regression coefficients β_{ij}. Furthermore we assume that the edge values $\{Y_{ij(m)}\}$ are independent conditional on the latent variables $\{Z_{ij(m)}\}$, and the distribution of $Y_{ij(m)}$ given $Z_{ij(m)}$ is given by

$$\Big[p_{in(m)} \cdot f(\cdot, \theta_{in(m)}) + (1 - p_{in(m)}) \cdot \delta_0(\cdot)\Big]^{Z_{ij(m)}} \Big[p_{out(m)} \cdot f(\cdot, \theta_{out(m)}) + (1 - p_{out(m)}) \cdot \delta_0(\cdot)\Big]^{1 - Z_{ij(m)}}, \tag{4.2}$$

where $\{p_{in(m)}, p_{out(m)}\}$ are unknown parameters governing the sparsity of subject m's graph, $\{f(\cdot, \theta_{in(m)}), f(\cdot, \theta_{out(m)})\}$ are arbitrary probability distributions with unknown parameter vectors $\{\theta_{in(m)}, \theta_{out(m)}\}$ and $\delta_0(\cdot)$ denotes the Dirac measure at zero accounting for non-present edges. The dependencies of the graphs are determined by the latent structure Z, and the latent structure is determined by the covariates.

Without loss of generality, we focus on Gaussian-weighted edges, but this framework can easily accommodate other edge distributions, as well as directed graphs by utilizing a bivariate distribution. We can also allow for more a complex latent structure via multinomial probit regression instead of binomial probit regression in the latent space.

4.2.2 Prior distributions

Conjugate prior distributions are chosen to allow for efficient MCMC sampling:

$$\beta_{ij} \sim \text{MVN}(b_{0,ij}, \Lambda_{0,ij}^{-1}), \text{where } x_m^\intercal = (1, x_{m1}, \ldots, x_{mp}) \text{ and } \beta_{ij} = (\beta_{ij,0}, \ldots, \beta_{ij,p})^\intercal,$$

$$p_{in(m)} \sim \text{Beta}(\alpha_{p_{in}}, \beta_{p_{in}}),\ p_{out(m)} \sim \text{Beta}(\alpha_{p_{out}}, \beta_{p_{out}}),$$

$$\mu_{in(m)} \sim \text{N}(\mu_{0(in)}, \sigma^2_{0(in)}),\ \mu_{out(m)} | \mu_{in(m)} \sim \text{TN}(\mu_{0(out)}, \sigma^2_{0(out)}, \mu_{in(m)}, \infty),$$

$$\tau_{in(m)} \sim \text{Ga}(\alpha_{0,in}, \beta_{0,in}),\ \tau_{out(m)} \sim \text{Ga}(\alpha_{0,out}, \beta_{0,out}),$$

where $b_{0,ij}$ indicates the prior mean and $\Lambda_{0,ij}$ the prior precision for the parameter vector β_{ij}, $\text{Ga}(a, b)$ is a gamma distribution with shape a and rate b, and $\text{TN}(\cdot, \cdot, a, \infty)$ indicates the left truncated normal distribution, with the truncation point given by a.

4.2.3 Estimation

We propose a Gibbs sampler for computation of the posterior distribution, with full conditional posterior distributions shown in the Appendix. The Gibbs sampler draws from these full conditional distributions to update each of the components sequentially. Our sampling algorithm is initiated by setting: $\beta_{ij}^0 = 0 \ \forall \ i, j$; $p_{in(m)} = p_{out(m)} = 0.5$, $\mu_{in(m)} = 1$, $\mu_{out(m)} = -1$, $\tau_{in(m)} = \tau_{out(m)} = 1$ for all m, and then we fill $Z_{ij(m)}^{*(0)}$ with random values from a standard normal distribution for all i, j, m. Then for $t = 1, \ldots, N$, we update the parameters as follows:

- Sample $Z_{ij(m)}^{*(t)}$ from $P(Z_{ij(m)}^{*} | p_{in(m)}^{(t-1)}, p_{out(m)}^{(t-1)}, \mu_{in(m)}^{(t-1)}, \mu_{out(m)}^{(t-1)}, \tau_{in(m)}^{(t-1)}, \tau_{out(m)}^{(t-1)}, \beta_{ij}^{(t-1)}, Y_{ij(m)})$ for all $1 \leq i < j \leq n$ and $m = 1, \ldots, M$.

- Sample $\beta_{ij}^{(t)}$ from $P(\beta_{ij} | Z_{ij}^{*(t)}, b_{0,ij}, \Lambda_{0,ij}, X) \ \forall \ 1 \leq i < j \leq n$

- Sample $p_{in(m)}^{(t)}$ from $P(p_{in(m)} | Z_{(m)}^{*(t)}, \alpha_{p_{in}}, \beta_{p_{in}}, Y_{(m)})$ and $p_{out(m)}^{(t)}$ from $P(p_{out(m)} | Z_{(m)}^{*(t)}, \alpha_{p_{out}}, \beta_{p_{out}}, Y_{(m)})$

- Sample $\mu_{in(m)}^{(t)}$ from $P(\mu_{in(m)} | Z_{(m)}^{*(t)}, \tau_{in(m)}^{(t-1)}, \mu_{0,in}, \sigma_{0,in}^2, Y_{(m)})$ and $\mu_{out(m)}^{(t)}$ from $P(\mu_{out(m)} | Z_{(m)}^{*(t)}, \tau_{out(m)}^{(t-1)}, \mu_{0,out}, \sigma_{0,out}^2, Y_{(m)})$

- Sample $\tau_{in(m)}^{(t)}$ from $P(\tau_{in(m)} | Z_{(m)}^{*(t)}, \mu_{in(m)}^{(t)}, \alpha_{0,in}, \beta_{0,in}, Y_{(m)})$ and $\tau_{out(m)}^{(t)}$ from $P(\tau_{out(m)} | Z_{(m)}^{*(t)}, \mu_{out(m)}^{(t)}, \alpha_{0,out}, \beta_{0,out}, Y_{(m)})$

In practice, to ensure convergence we can run multiple parallel chains and then combine the samples if they have converged to the same posterior distribution, which can be assessed by the diagnostic tools in the *coda* R package (Plummer et al. 2006). The method was written in C++ and implemented in R using the Rcpp and RcppEigen (Eddelbuettel et al. 2011, Bates and Eddelbuettel 2013) packages.

4.3 Simulations

We performed simulation studies to examine the finite-sample performance of our estimation algorithm. For each of the four simulation schemes, listed in Table 4.1, we have simulated 100 datasets and then used our sampling algorithm to obtain 20,000 samples from the posterior distribution, following an initial burn-in of 1,000 samples. Subsequently, we examined classification accuracy and the ability of our method to capture differences between two groups in the latent relationships between nodes.

4.3.1 Simulation setup

For each scheme, we have simulated weighted networks for 100 subjects with 50 nodes in each network. We focus on the scenario when there are two groups of subjects, defined for example by diagnosis or treatment, which leads to a single covariate in our latent probit model. Additionally, we have set the hyperparameters b_0 and b_1 to zero for 100 of the $\{i, j\}$ pairs, to be considered "null" effects. We fixed the other hyperparameters and sampled from the appropriate prior distributions for each subject's parameters, after which we generated each subject's weighted edges from the model presented in Eq. (4.2).

Table 4.1: Simulation schemes

Scheme	b_0	b_1	$\Lambda_{0,11} = \Lambda_{0,22} =$	$\mu_{0,in}$	$\mu_{0,out}$
1	-1	2	5	2	-2
2	-0.5	1	10	2	-2
3	-1	2	10	1	-1
4	-0.5	1	10	1	-1
5	-0.25	0.5	10	1	-1
6	-0.25	0.5	10	0.5	-0.5

Table 4.1. 100 datasets were simulated from each of these schemes.

For all simulations, we used the same flat priors, setting the hyperparameters as: $b_{0(ij)} = (0, \ldots, 0)$

and $\Lambda_{0(ij)} = \begin{pmatrix} 0.01 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & 0.01 \end{pmatrix}$ $\forall\ 0 \leq i < j \leq n$; $\mu_{0,in} = \mu_{0,out} = 0, \sigma^2_{0,in} = \sigma^2_{0,out} = 10$; $\alpha_{p_{in}} = \alpha_{p_{out}} = 1, \beta_{p_{in}} = \beta_{p_{out}} = 1$; $\alpha_{0,in} = \alpha_{0,out} = 1, \beta_{0,in} = \beta_{0,out} = 0.2$.

4.3.2 Results

In each simulation scheme, our estimation algorithm does well at recovering the true latent relationships between nodes. To determine the best threshold for the posterior probability to determine that a pair of nodes is in the same class, we examine the average misclassification rates for each simulation scheme in Figure 4.1(L), across different cutoff values for $P(Z^*_{ij(m)} > 0|\ldots)$.

Figure 4.1: Misclassifcation rates by simulation scheme

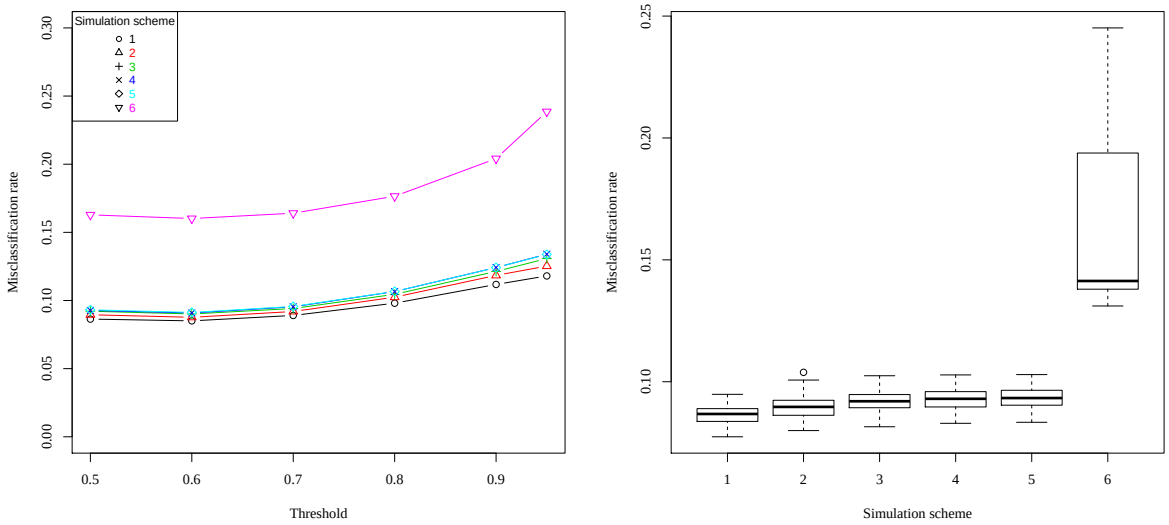


Figure 4.1. (L) Average misclassification rates by simulation scheme, for different thresholds of $P(Z^*_{ij(m)} > 0|\ldots)$. (R) Boxplots of misclassification rates by simulation scheme when using $P(Z^*_{ij(m)} > 0|\ldots) = 0.5$ as the threshold. The 6 schemes, each with 100 simulated datasets, are listed in Table 4.1. Misclassification rate is calculated as the sum of the false positives and false negatives divided by the total number of possible node pairs.

The threshold of 0.5 for the posterior probability $P(Z^*_{ij(m)} > 0|\ldots)$ to determine if $Z_{ij(m)} = 1$, corresponding to the maximum *a posteriori* (MAP) estimate of $Z_{ij(m)}$, allows for nearly as accurate classification as the 0.6 threshold but is more easily interpretable, so we proceed

using 0.5 as the cutoff. Figure 2.2(R) shows boxplots of the misclassification rates for each set of 100 simulations. The classification is likely affected by the β_{ij} with $\{b_0, b_1\}$ set to zero, where we would not expect to capture the latent relationships with $P(Z_{ij(m)} = 1) \approx 0.5$ for all subjects.

To assess the ability of our method to capture the effects of covariates on the estimated latent structure, we calculated the average coverage of the 95% credible intervals (the proportion of the parameters where the true value fell between the 0.025 and 0.975 quantiles) across all $\beta_{ij(1)},\ 1 \le i < j \le n$ and all 100 simulations in each scheme. We found 91% coverage, on average, for each of simulation schemes 1 – 5 and 82% coverage for the much more difficult scheme 6.

4.4 Application to ADHD Data

We applied our Bayesian hierarchical random graph model to the same functional connectivity networks seen in the previous two chapters, to assess whether covariates are associated with latent functional brain structure. In our latent probit regression model, gender, age, and diagnosis group are included as covariates. Table 4.2 shows the distribution of gender across diagnosis groups. A Chi-squared test for independence leads us to reject the null hypothesis and conclude that gender and study group are associated (p=0.0001). One subject had a missing gender value; to include this subject, the mean of the gender indicator variable (the proportion of males) was substituted as the imputed value into the design matrix. Figure 4.2 shows that the age distribution is similar for both genders but differs somewhat between diagnosis groups.

We analyze the set of functional networks with two parallel MCMC chains, using flat priors by setting the hyperparameters to: $b_{0(ij)} = (0, \ldots, 0)$ and $\Lambda_{0(ij)} = \begin{pmatrix} 0.1 & & 0.01 \\ \vdots & \ddots & \vdots \\ 0.01 & & 0.1 \end{pmatrix}$ for all $0 \le i < j \le n$; $\mu_{0,in} = \mu_{0,out} = 0,\ \sigma^2_{0,in} = \sigma^2_{0,out} = 5;\ \ \alpha_{p_{in}} = \alpha_{p_{out}} = 1, \beta_{p_{in}} = \beta_{p_{out}} = 1;$ $\alpha_{0,in} = \alpha_{0,out} = 1, \beta_{0,in} = \beta_{0,out} = 0.2$. We collect 5,000 samples in each chain, after a burn-in of

Table 4.2: Gender by diagnosis group

	Control	Combined	Inattentive
Female	46	10	12
Male	45	51	20

Table 4.2. Gender breakdown by diagnosis group.

Figure 4.2: Age distribution

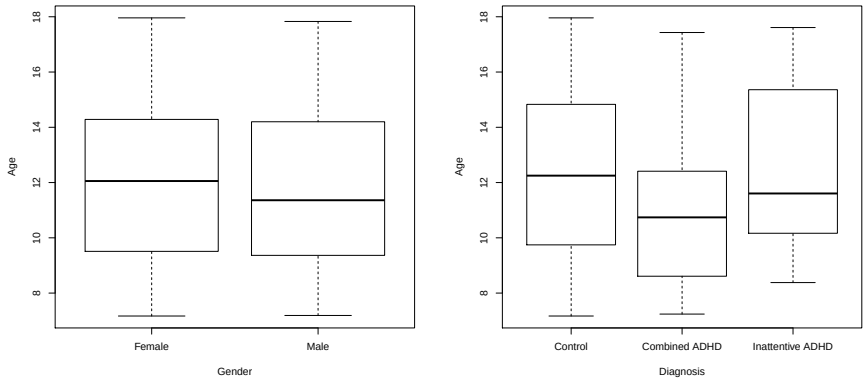


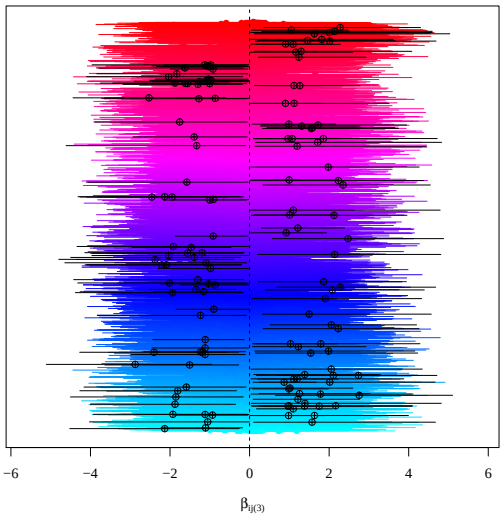
Figure 4.2. Age distribution by gender (L) and by diagnosis (R).

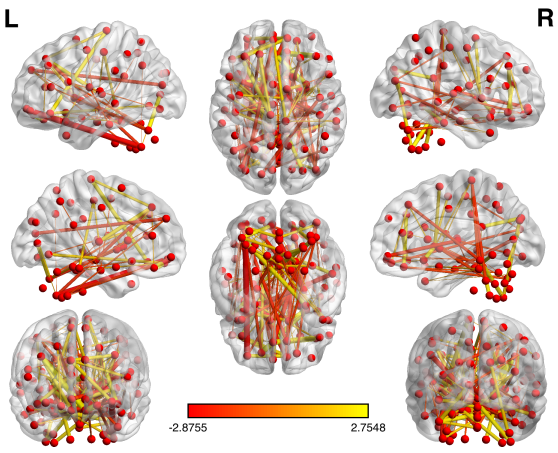
500 samples in each, and combine the samples after checking to make sure the two chains have converged to the same posterior distribution by visual inspection of the posterior samples.

Figure 4.3 shows the parameter estimates comparing the latent structures of the ADHD groups to that of the control group. In the left panel are the 99% credible intervals for each β_{ij}, with those in black indicating parameters where zero was not contained in the interval, and in the right panel are the median estimates for just those "significant" β_{ij}, which has been visualized with the BrainNet Viewer (`http://www.nitrc.org/projects/bnv/`) (Xia et al. 2013). The inattentive subtype was found to have decreased probability of latent connections between brain regions, after adjusting for age and gender (see Figure 4.3(b)). In the appendix, Figure C.1.1 shows the effects of gender and age on the probability of latent connections. There appears to be a tremendous effect of age on latent brain structure, after

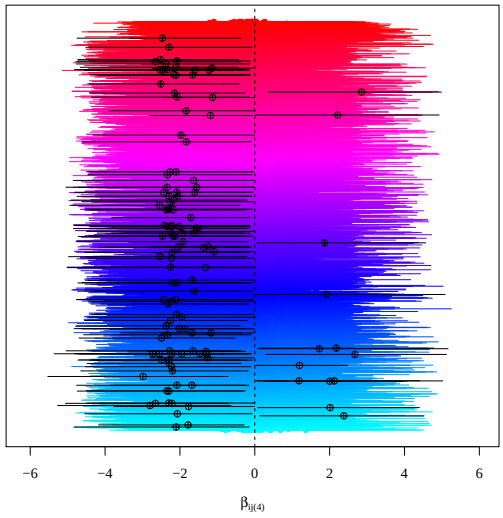
adjusting for gender and ADHD diagnosis.

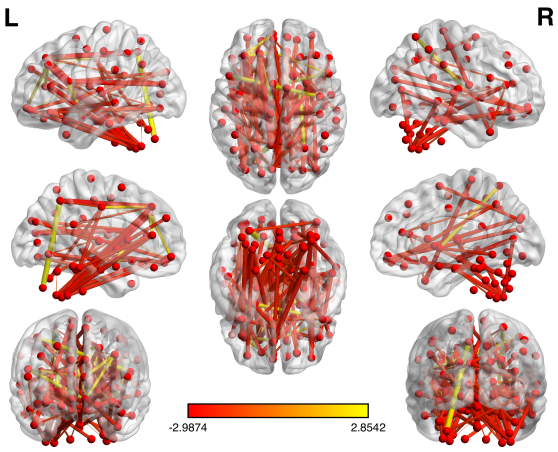
Figure 4.3: Parameter estimates: ADHD diagnoses





(a) Combined subtype vs. control





(b) Inattentive subtype vs. control

Figure 4.3. Coefficient estimates comparing the probability of a latent connection between all $n \cdot (n-1)/2$ pairs of nodes. (L) β_{ij} posterior medians and 99% credible intervals and (R) the posterior medians just for the "significant" estimates where zero was not contained in the credible interval.

Figure 4.4 gives the posterior medians and the 95% credible intervals for the modularity parameters of all subjects $1, \ldots, M$, highlighting the heterogeneity of latent functional brain structure across the children in this study.

4.5 Discussion

In this chapter we have presented a novel adaptation of the SBM, to allow for a more complex latent structure than the typical block structure and additionally capture the effects of covariates in the latent space. This framework makes sense in settings where the data consist of a collection of networks that share the same nodes, such as in an fMRI study or in the context of a genetic array. Our approach addresses, particularly, the common question in neuroscience literature about differences in functional brain networks by going beyond the typical assessment of network summary measures. By using an efficient Gibbs sampling algorithm with conjugate priors, we are able to accommodate moderately large networks, on the order of hundreds of nodes. In our simulations, we were able to accurately capture both the latent structure across subjects and differences in the latent structure between two groups, as in a study comparing treatments or diagnosis groups.

We applied our method to 185 functional brain networks from the ADHD-200 study and found numerous differences in latent functional brain structure that can be explained by covariates. There appear to be fundamental differences in the latent structures of ADHD subtypes as compared to the control group, after adjusting for gender and age, as well as differences between genders after adjusting for diagnosis and age. In particular, the inattentive ADHD subtype appears to have significantly less connectivity across many regions of the brain as compared to the control group, which is in line with what some other researchers have found (Rosenberg et al. 2015). We also found a strong effect of age in latent functional brain structure, which is unsurprising given all that is known about brain development in children. This probabilistic approach could also be useful in applying to studies of other

Figure 4.4: Modularity estimates

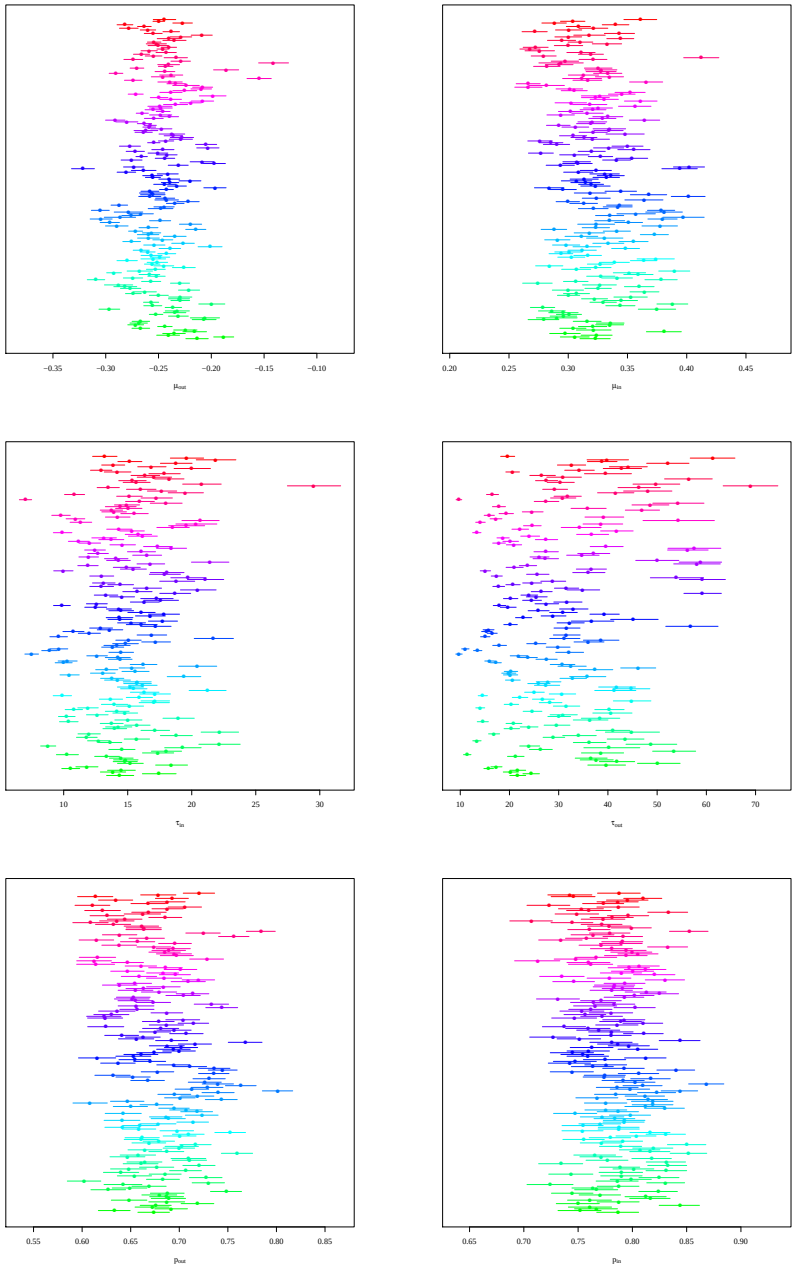


Figure 4.4. Posterior medians and 95% credible intervals for modularity parameters (Sparsity: $p_{in(m)}$ and $p_{out(m)}$, edge weights: $\mu_{in(m)}, \mu_{out(m)}, \tau_{in(m)}, \tau_{out(m)}$) for all subjects.

brain diseases such as Alzheimer's disease, to better understand the underlying structure that influences correlations in brain activity.

APPENDIX A: CHAPTER 2

A.1 Figures

Figure A.1.1: Coverage of 95% highest posterior density regions by simulation scheme

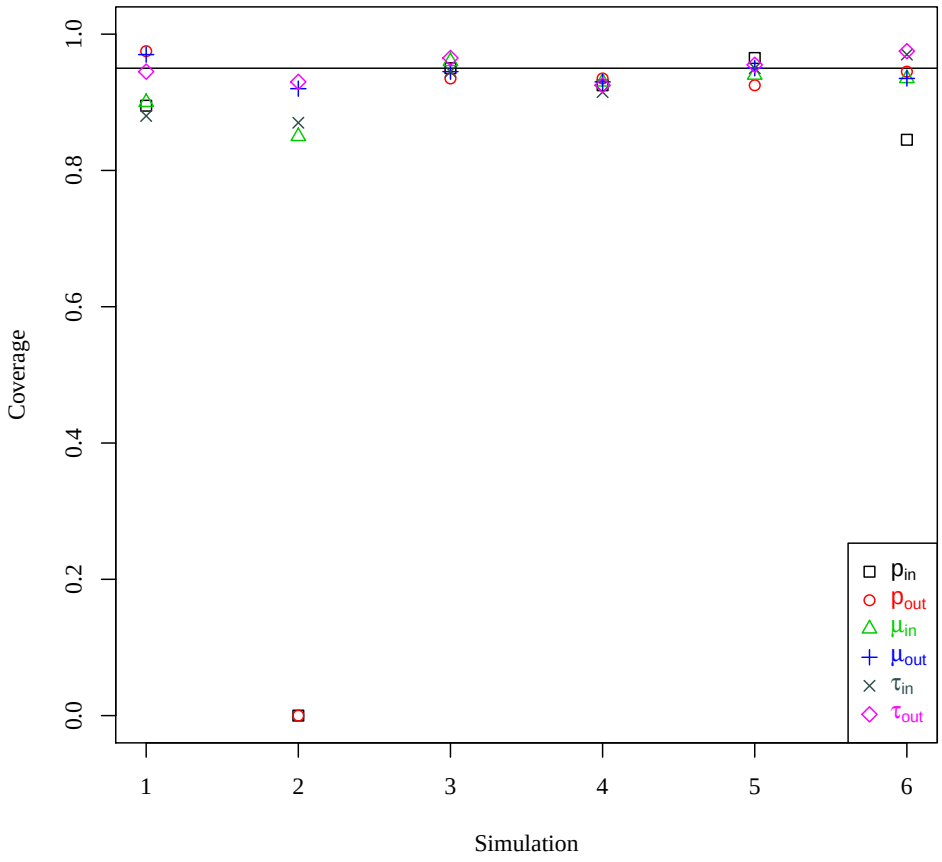


Figure A.1.1. Percent of the 95% HPD intervals containing the true value, across 200 simulations in each scheme. Simulation schemes are listed in Table 2.1. The horizontal line indicates 95%.

Figure A.1.2: Median 95% highest posterior density region width by simulation scheme

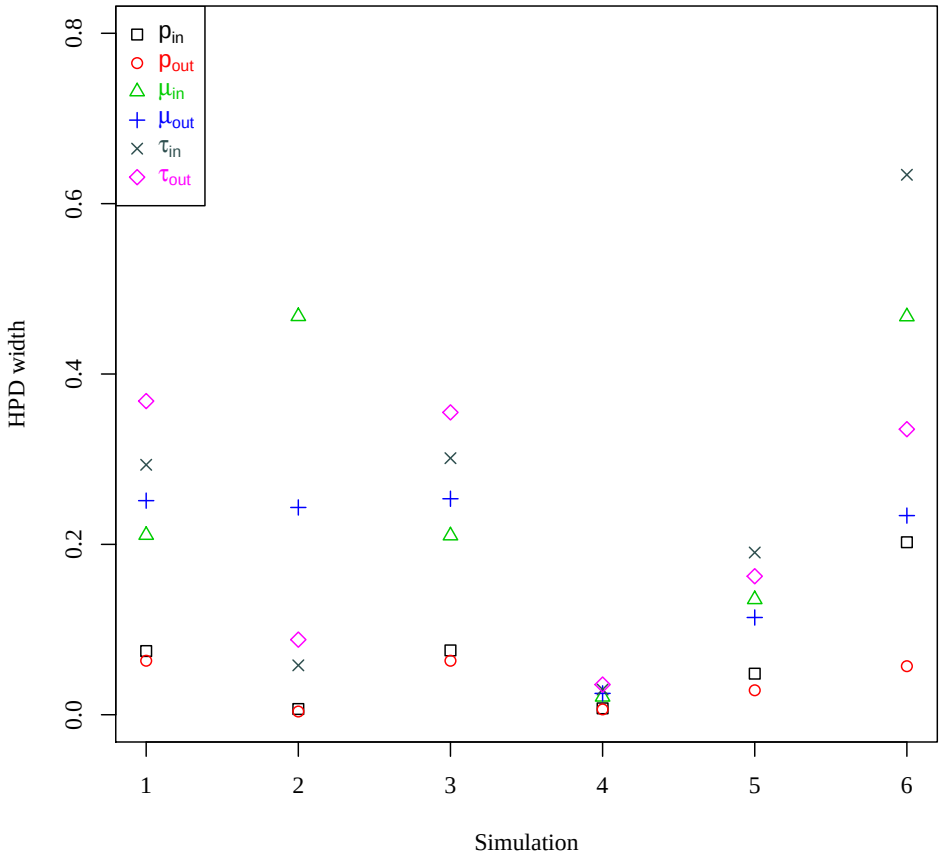


Figure A.1.2. Median widths of the 95% HPD intervals across 200 simulations in each scheme. Simulation schemes are listed in Table 2.1.

Figure A.1.3: Overlap of communities: by study group

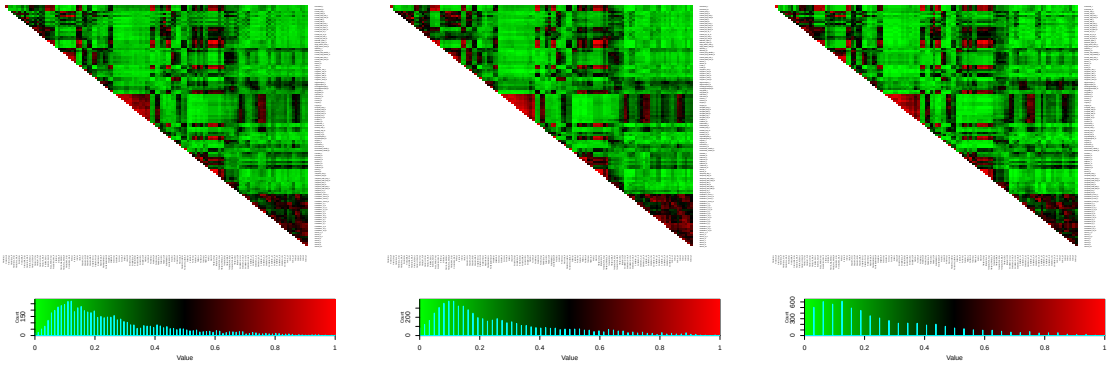


Figure A.1.3. Overlap of the latent class structure across 91 control subjects (Top), 62 combined-subtype subjects (Middle), and 32 inattentive-subtype subjects (Bottom) from the ADHD-200 sample. Each element of the matrix is the proportion of the corresponding subjects in whom the corresponding two nodes fall in the same estimated latent class.

A.2 Distributions

A.2.1 Prior and sampling distributions

Prior and sampling distributions are listed as follows:

$$P(Z_i|\pi,Q)=\pi_1^{Z_{i1}}\cdots\pi_Q^{Z_{iQ}},\ P(\pi|Q)=\prod_{q=1}^{Q}\pi_q^{a_q-1},$$

and $P(Y|Z,\pi,Q,\theta,p)$ is given by

$$\begin{aligned}
&\prod_{1<i<j<n}P(Y_{ij}|\theta_{in},\theta_{out},p_{in},p_{out},Z_i,Z_j,Q)\\
=&\prod_{1<i<j<n}\prod_{q,l}P(Y_{ij}|\theta_{in},\theta_{out},p_{in},p_{out},Z_{iq}*Z_{jl}=1,Q)\\
=&\prod_{1<i<j<n}\Big(\prod_{q=l}\Big[\big(p_{in}f(Y_{ij};\theta_{in})\big)^{1(Y_{ij}\neq 0)}(1-p_{in})^{1(Y_{ij}=0)}\Big]^{Z_{iq}Z_{jl}}\Big)\\
&\Big(\prod_{q\neq l}\Big[\big(p_{out}f(Y_{ij};\theta_{out})\big)^{1(Y_{ij}\neq 0)}(1-p_{out})^{1(Y_{ij}=0)}\Big]^{Z_{iq}Z_{jl}}\Big)\\
=&\prod_{1<i<j<n}\Big(\prod_{q=l}\Big[\big(p_{in}(\frac{\tau_{in}}{2\pi})^{1/2}\exp\{-\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\}\big)^{1(Y_{ij}\neq 0)}(1-p_{in})^{1(Y_{ij}=0)}\Big]^{Z_{iq}Z_{jl}}\Big)\\
&\Big(\prod_{q\neq l}\Big[\big(p_{out}(\frac{\tau_{out}}{2\pi})^{1/2}\exp\{-\frac{\tau_{out}}{2}(Y_{ij}-\mu_{out})^2\}\big)^{1(Y_{ij}\neq 0)}(1-p_{out})^{1(Y_{ij}=0)}\Big]^{Z_{iq}Z_{jl}}\Big)\\
=&\prod_{A}\Big(p_{in}(\frac{\tau_{in}}{2\pi})^{1/2}\exp\{-\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\}\Big)\\
&\prod_{B}\Big(p_{out}(\frac{\tau_{out}}{2\pi})^{1/2}\exp\{-\frac{\tau_{out}}{2}(Y_{ij}-\mu_{out})^2\}\Big)\prod_{C}(1-p_{in})\prod_{D}(1-p_{out}),
\end{aligned}$$

where A,B,C, and D satisfy

$$A=\{i<j,q:Z_{iq}Z_{jq}=1,Y_{ij}\neq 0\},\ B=\{i<j,q\neq l:Z_{iq}Z_{jl}=1,Y_{ij}\neq 0\};$$

$$C=\{i<j,q:Z_{iq}Z_{jq}=1,Y_{ij}=0\},\ \text{and}\ D=\{i<j,q\neq l:Z_{iq}Z_{jl}=1,Y_{ij}=0\}.$$

Moreover, we set

$$P(p_{in})=1(0<p_{in}<1),\ P(p_{out})=1(0<p_{out}<1),$$

$$P(\mu_{in}|\tau_{in}) = \left(\frac{\tau_{in}}{2\pi\sigma_{0,in}^2}\right)^{1/2} \exp\{\frac{\tau_{in}}{2\sigma_{0,in}^2}(\mu_{in}-\mu_{0,in})^2\},$$
$$P(\tau_{in}) = \beta_{0,in}^{\alpha_{0,in}} \frac{1}{\Gamma(\alpha_{0,in})} \tau_{in}^{\alpha_{0,in}-1} \exp\{-\beta_{0,in}\tau_{in}\},$$
$$P(\mu_{out}|\tau_{out}) = \left(\frac{\tau_{out}}{2\pi\sigma_{0,out}^2}\right)^{1/2} \exp\{\frac{\tau_{out}}{2\sigma_{0,out}^2}(\mu_{out}-\mu_{0,out})^2\},$$
$$P(\tau_{out}) = \beta_{0,out}^{\alpha_{0,out}} \frac{1}{\Gamma(\alpha_{0,out})} \tau_{out}^{\alpha_{0,out}-1} \exp\{-\beta_{0,out}\tau_{out}\}.$$

A.2.2 Full conditional posterior distributions

The full conditional distributions are derived as follows. First, we have

$$\begin{aligned}
&P(Z_i|\pi, Q, p_{in}, p_{out}, \theta, Y) \propto P(Y|\theta, p, \pi, Z, Q) P(Z_i|\pi, Q) \\
&\propto \prod_{1\le i<j\le n} \left(\prod_{q=l} \left[\left(p_{in} \left(\frac{\tau_{in}}{2\pi}\right)^{1/2} \exp\{-\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\} \right)^{1(Y_{ij}\neq 0)} (1-p_{in})^{1(Y_{ij}=0)} \right]^{Z_{iq}Z_{jl}} \right) \\
&* \left(\prod_{q\neq l} \left[\left(p_{out} \left(\frac{\tau_{out}}{2\pi}\right)^{1/2} \exp\{-\frac{\tau_{out}}{2}(Y_{ij}-\mu_{out})^2\} \right)^{1(Y_{ij}\neq 0)} (1-p_{out})^{1(Y_{ij}=0)} \right]^{Z_{iq}Z_{jl}} \right) \\
&* \left(\pi_1^{Z_{i1}} \cdots \pi_Q^{Z_{iQ}} \right).
\end{aligned}$$

Therefore, the full conditional distribution of Z_i given all others is proportional to

$$\left[\prod_{j\neq i} D_{in}(i,j)^{Z_{j1}} D_{out}(i,j)^{\sum_{q\neq 1} Z_{jq}} \right]^{Z_{i1}} \cdots \left[\prod_{j\neq i} D_{in}(i,j)^{Z_{jQ}} D_{out}(i,j)^{\sum_{q\neq Q} Z_{jq}} \right]^{Z_{iQ}} \left(\pi_1^{Z_{i1}} \cdots \pi_Q^{Z_{iQ}} \right),$$

where $D_{in}(i,j) = \left(p_{in} \left(\frac{\tau_{in}}{2\pi}\right)^{1/2} \exp\{-\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\} \right)^{1(Y_{ij}\neq 0)} (1-p_{in})^{1(Y_{ij}=0)}$ and

$$D_{out}(i,j) = \left(p_{out} \left(\frac{\tau_{out}}{2\pi}\right)^{1/2} \exp\{-\frac{\tau_{out}}{2}(Y_{ij}-\mu_{out})^2\} \right)^{1(Y_{ij}\neq 0)} (1-p_{out})^{1(Y_{ij}=0)}.$$

Finally, we have $Z_i|\ldots \sim \text{Multinomial}\left(\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iQ}\right)$, where $\tilde{\pi}_{iq}$ is given by

$$\tilde{\pi}_{iq} = \frac{\pi_q \prod\limits_{j\neq i} D_{in}(i,j)^{Z_{jq}} D_{out}(i,j)^{\sum\limits_{k\neq q} Z_{jk}}}{\sum\limits_{q=1}^{Q} \pi_q \prod\limits_{j\neq i} D_{in}(i,j)^{Z_{jq}} D_{out}(i,j)^{\sum\limits_{k\neq q} Z_{jk}}} \text{ for } q = 1,\ldots,Q.$$

The full conditional distribution of π is given by

$$\begin{aligned} p(\pi|Q,Z,Y) &\propto P(Z|\pi,Q)P(\pi|Q) \propto \left(\prod_{i=1}^{n} \pi_1^{Z_{i1}}\cdots\pi_Q^{Z_{iQ}}\right)\left(\prod_{q=1}^{Q}\pi_q^{a_q-1}\right) \\ &\propto \pi_1^{\sum\limits_{i=1}^{n} Z_{i1}+a_1-1}\cdots\pi_Q^{\sum\limits_{i=1}^{n} Z_{iQ}+a_Q-1}, \end{aligned}$$

which implies that $\pi|\ldots \sim \text{Dirichlet}\left(\sum\limits_{i=1}^{n} Z_{i1}+a_1,\ldots,\sum\limits_{i=1}^{n} Z_{iQ}+a_Q\right)$.

The full conditional distribution of p_{in} is given by

$$P(Y|\theta,p,\pi,Z,Q)P(p_{in}) \propto p_{in}^{n_A}(1-p_{in})^{n_C}\cdot 1(0<p_{in}<1) \sim \text{Beta}(n_A+1,n_C+1),$$

where $n_A = |A|$ and $n_B = |B|$. Similarly, we have

$$p(p_{out}|\cdots) \sim \text{Beta}(n_B+1,n_D+1),$$

where $n_C = |C|$ and $n_D = |D|$.

The full conditional distribution of μ_{in} is given by

$$\begin{aligned} &P(\mu_{in}|\tau,\pi,Z,Q,p,Y) \propto P(Y|\theta,p,\pi,Z,Q)P(\mu_{in}|\tau_{in}^{-1}) \\ &\propto \exp\{\sum_{A} -\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\}\exp\{-\frac{\tau_{in}}{2\sigma_{0,in}^2}(\mu_{in}-\mu_{0,in})^2\}, \end{aligned}$$

which implies that

$$\mu_{in}|\cdots \sim \mathrm{N}\left(\frac{\sigma_{0,in}^2 \sum_A Y_{ij} + \mu_{0,in}}{n_A\sigma_{0,in}^2 + 1}, \frac{\sigma_{0,in}^2\tau_{in}^{-1}}{n_A\sigma_{0,in}^2 + 1}\right).$$

Similarly, we have

$$\mu_{out}|\cdots \sim \mathrm{N}\left(\frac{\sigma_{0,out}^2 \sum_B Y_{ij} + \mu_{0,out}}{n_A\sigma_{0,out}^2 + 1}, \frac{\sigma_{0,out}^2\tau_{out}^{-1}}{n_B\sigma_{0,out}^2 + 1}\right).$$

The full conditional distribution of τ_{in} is given by

$$\begin{aligned}
& P(\tau_{in}|\mu,\pi,Z,Q,p,Y) \propto P(Y|\theta,p,\pi,Z,Q)P(\mu_{in}|\tau_{in})P(\tau_{in}) \\
\propto\ & \left[\prod_A \left(p_{in}(\frac{\tau_{in}}{2\pi})^{1/2}\exp\{-\frac{\tau_{in}}{2}(Y_{ij}-\mu_{in})^2\}\right)\right] \\
& \left(\frac{\tau_{in}}{2\pi\sigma_{0,in}^2}\right)^{1/2}\exp\{\frac{\tau_{in}}{2\sigma_{0,in}^2}(\mu_{in}-\mu_{0,in})^2\}\beta_{0,in}^{\alpha_{0,in}}\frac{1}{\Gamma(\alpha_{0,in})}\tau_{in}^{\alpha_{0,in}-1}\exp\{-\beta_{0,in}\tau_{in}\} \\
\propto\ & \tau_{in}^{\left(\frac{n_A+1}{2}+\alpha_{0,in}-1\right)}\exp\{\frac{-\tau_{in}}{2}\sum_A(Y_{ij}-\mu_{in})^2 - \frac{\tau_{in}}{2\sigma_{0,in}^2}(\mu_{in}-\mu_{0,in})^2 - \beta_{0,in}\tau_{in}\},
\end{aligned}$$

which implies that

$$\tau_{in}|\cdots \sim \mathrm{Gamma}\left(\frac{n_A+1}{2}+\alpha_{0,in}, \frac{1}{2}\sum_A(Y_{ij}-\mu_{in})^2 + \frac{1}{2\sigma_{0,in}^2}(\mu_{in}-\mu_{0,in})^2 + \beta_{0,in}\right).$$

Similarly, we have

$$\tau_{out}|\cdots \sim \mathrm{Gamma}\left(\frac{n_B+1}{2}+\alpha_{0,out}, \frac{1}{2}\sum_B(Y_{ij}-\mu_{out})^2 + \frac{1}{2\sigma_{0,out}^2}(\mu_{out}-\mu_{0,out})^2 + \beta_{0,out}\right).$$

APPENDIX B: CHAPTER 3

B.1 Figures

Figure B.1.1: Difference in overlap of communities between latent groups

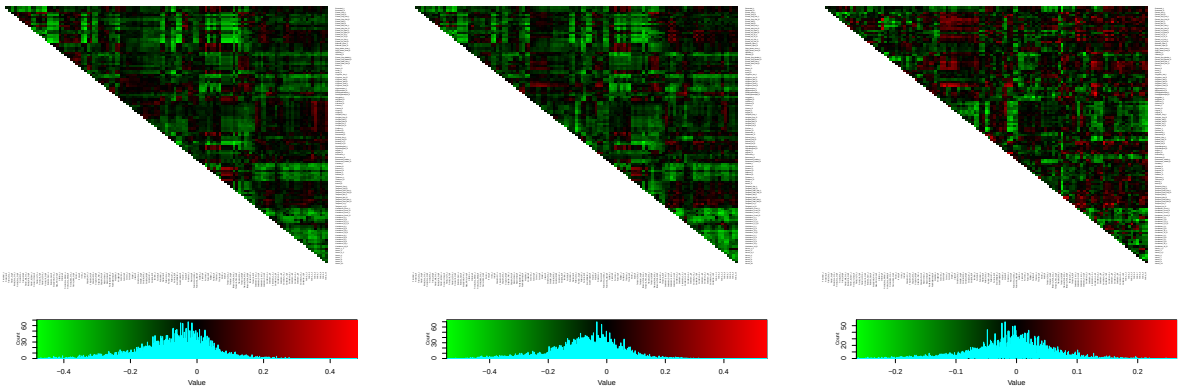


Figure B.1.1. Difference in the overlap of the latent class structure between the three estimated latent groups. Each element in these matrices gives the difference in proportions of the two groups' subjects in whom the corresponding two nodes fall in the same estimated latent class. The comparisons are groups 2 vs. 1 (L), 3 vs. 1 (C), and 3 vs. 2 (R).

B.2 Distributions

B.2.1 Prior and sampling distributions

The prior and sampling distributions are as follows: For each subject $m = 1, \ldots, M$,

$$P(V_m|\xi) = \xi_1^{V_{m1}} \cdots \xi_W^{V_{mW}},$$

where the probability vector ξ is distributed according to

$$P(\xi) = \prod_{w=1}^{W} \xi_w^{a_w - 1}.$$

Then, for each $i = 1, \ldots, n$ and $m = 1, \ldots, M$

$$P\left(Z_{i(m)}|V_m = w, \pi_{i(w)}\right) = \pi_{i1(w)}^{Z_{i1(m)}} \cdots \pi_{iQ(w)}^{Z_{iQ(m)}},$$

where, for $w = 1, \ldots, W$ and $i = 1, \ldots, n$,

$$P(\pi_{i(w)}) = \pi_{i1(w)}^{\alpha_{i1(w)}-1} \cdots \pi_{iQ(w)}^{\alpha_{iQ(w)}-1}.$$

Then $P(Y|V, Z, \mu, \tau, p_{in}, p_{out}, \mu_{in(w)}, \mu_{out(w)}, \tau_{in(w)}, \tau_{out(w)})$ is given by

$$\begin{aligned}
&\prod_{m=1}^{M} \prod_{1 \le i < j \le n} \prod_{w=1}^{W} \left[P(Y_{ij(m)} | V_m = w, Z_{i(m)}, Z_{j(m)}, \mu_{in(w)}, \mu_{out(w)}, \tau_{in(w)}, \tau_{out(w)}, p_{in(w)}, p_{out(w)}) \right]^{V_{mw}} \\
= &\prod_{m=1}^{M} \prod_{1 \le i < j \le n} \prod_{w=1}^{W} \left[\left[D_{in(m,w)}(i,j) \right]^{\sum_{q=1}^{Q} Z_{iq(m)} \cdot Z_{jq(m)}} \left[D_{out(m,w)}(i,j) \right]^{1 - \sum_{q=1}^{Q} Z_{iq(m)} \cdot Z_{jq(m)}} \right]^{V_{mw}},
\end{aligned}$$

where

$$D_{in(m,w)}(i,j) = \left(p_{in(w)} \left(\frac{\tau_{in(w)}}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\tau_{in(w)}}{2} (Y_{ij(m)} - \mu_{in(w)})^2 \right\} \right)^{1(Y_{ij(m)} \neq 0)} (1 - p_{in(w)})^{1(Y_{ij(m)} = 0)},$$

$$D_{out(m,w)}(i,j) = \left(p_{out(w)} \left(\frac{\tau_{out(w)}}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\tau_{out(w)}}{2} (Y_{ij(m)} - \mu_{out(w)})^2 \right\} \right)^{1(Y_{ij(m)} \neq 0)} (1 - p_{out(w)})^{1(Y_{ij(m)} = 0)}.$$

Then for the edge parameters, we set

$$\begin{aligned}
P(p_{in(w)}) &= \frac{1}{B(\alpha_{p_{in(w)}}, \beta_{p_{in(w)}})} \, p_{in(w)}^{\alpha_{p_{in(w)}}-1} (1 - p_{in(w)})^{\beta_{p_{in(w)}}-1}, \\
P(p_{out(w)}) &= \frac{1}{B(\alpha_{p_{out(w)}}, \beta_{p_{out(w)}})} \, p_{out(w)}^{\alpha_{p_{out(w)}}-1} (1 - p_{out(w)})^{\beta_{p_{out(w)}}-1}, \\
P(\mu_{in(w)} | \tau_{in(w)}) &= \left(\frac{\tau_{in(w)}}{2\pi\sigma_{0,in(w)}^2} \right)^{1/2} \exp\left\{ \frac{\tau_{in(w)}}{2\sigma_{0,in(w)}^2} (\mu_{in(w)} - \mu_{0,in(w)})^2 \right\}, \\
P(\tau_{in(w)}) &= \beta_{0,in(w)}^{\alpha_{0,in(w)}} \, \frac{1}{\Gamma(\alpha_{0,in(w)})} \, \tau_{in(w)}^{\alpha_{0,in(w)}-1} \exp\{-\beta_{0,in(w)} \tau_{in(w)}\}, \\
P(\mu_{out(w)} | \tau_{out(w)}) &= \left(\frac{\tau_{out(w)}}{2\pi\sigma_{0,out(w)}^2} \right)^{1/2} \exp\left\{ \frac{\tau_{out(w)}}{2\sigma_{0,out(w)}^2} (\mu_{out(w)} - \mu_{0,out(w)})^2 \right\}, \\
P(\tau_{out(w)}) &= \beta_{0,out(w)}^{\alpha_{0,out(w)}} \, \frac{1}{\Gamma(\alpha_{0,out(w)})} \, \tau_{out(w)}^{\alpha_{0,out(w)}-1} \exp\{-\beta_{0,out(w)} \tau_{out(w)}\}.
\end{aligned}$$

B.2.2 Full conditional posterior distributions

The full conditional distributions are then derived as follows. First, we have

$$P(V_m|\ldots) \propto P(V_m|\xi) \prod_{1\le i<j\le n} P(X_{ij(m)}|V_m, Z_{i(m)}, Z_{j(m)}, \mu, \tau, p_{in}, p_{out})$$

$$\propto \xi_1^{V_{m1}} \ldots \xi_W^{V_{mW}} \cdot \prod_{1\le i<j\le n} \prod_{w=1}^{W} \left[\left[D_{in(m,w)}(i,j)\right]^{\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}} \left[D_{out(m,w)}(i,j)\right]^{1-\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}} \right]^{V_{mw}}.$$

So, $V_m|\ldots \sim \text{Multinomial}(\tilde{\xi}_1, \ldots, \tilde{\xi}_W)$, where

$$\tilde{\xi}_w = \frac{\xi_w \prod_{1\le i<j\le n} \left[D_{in(m,w)}(i,j)\right]^{\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}} \left[D_{out(m,w)}(i,j)\right]^{1-\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}}}{\sum_{w=1}^{W} \xi_w \prod_{1\le i<j\le n} \left[D_{in(m,w)}(i,j)\right]^{\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}} \left[D_{out(m,w)}(i,j)\right]^{1-\sum_{q=1}^{Q} Z_{iq(m)}\cdot Z_{jq(m)}}}.$$

The full conditional distribution of ξ is proportional to

$$P(\xi|\ldots) \propto P(\xi) \prod_{m=1}^{M} P(V_m|\xi)$$

$$\propto \xi_1^{\sum_{m=1}^{M} V_{m1}+a_1-1} \cdots \xi_W^{\sum_{m=1}^{M} V_{mW}+a_W-1},$$

which implies that $\xi|\ldots \sim \text{Dirichlet}\left(\sum_{m=1}^{M} V_{m1} + a_1, \ldots, \sum_{m=1}^{M} V_{mW} + a_W\right)$.

For $m = 1, \ldots, M$, the full conditional distribution of $Z_{i(m)}$ given all other variables is proportional to

$$P(Z_{i(m)}|V_m = w, \pi_{i(w)}) \prod_{j\ne i} P(X_{ij(m)}|Z_{i(m)}, Z_{j(m)}, V_m, \mu, \tau, p_{in}, p_{out})$$

$$\propto \pi_{i1(w)}^{Z_{i1(m)}} \cdots \pi_{iQ(w)}^{Z_{iQ(m)}} \prod_{j\ne i} \left[D_{in(m,w)}(i,j)^{\sum_{q=1}^{Q} Z_{iq(m)}Z_{jq(m)}} \; D_{out(m,w)}(i,j)^{1-\sum_{q=1}^{Q} Z_{iq(m)}Z_{jq(m)}} \right]$$

$$\propto \left[\pi_{i1(w)} \prod_{j\ne i} \left(\frac{D_{in(m,w)}(i,j)}{D_{out(m,w)}(i,j)} \right)^{Z_{j1(m)}} \right]^{Z_{i1(m)}} \cdots \left[\pi_{iQ(w)} \prod_{j\ne i} \left(\frac{D_{in(m,w)}(i,j)}{D_{out(m,w)}(i,j)} \right)^{Z_{jQ(m)}} \right]^{Z_{iQ(m)}},$$

which implies that $Z_{i(m)}|\ldots \sim \text{Multinomial}(\tilde{\pi}_{i(m)})$, where

$$\tilde{\pi}_{iq(m)} = \frac{\prod_{w=1}^{W}\left[\pi_{iq(w)}\prod_{j\neq i}\left(\frac{D_{in(m,w)}(i,j)}{D_{out(m,w)}(i,j)}\right)^{Z_{jq(m)}}\right]^{V_{mw}}}{\sum_{q=1}^{Q}\prod_{w=1}^{W}\left[\pi_{iq(w)}\prod_{j\neq i}\left(\frac{D_{in(m,w)}(i,j)}{D_{out(m,w)}(i,j)}\right)^{Z_{jq(m)}}\right]^{V_{mw}}}.$$

For $w = 1,\ldots,W$, the full conditional distribution of $\pi_{i(w)}$ given all other variables is proportional to

$$\begin{aligned}
&P(\pi_{i(w)})\prod_{V_m=w}P(Z_{i(m)}|V_m=w,\pi_{i(w)})\\
&\propto \pi_{i1(w)}^{\alpha_{i1(w)}-1}\cdots\pi_{iQ(w)}^{\alpha_{iQ(w)}-1}\prod_{V_m=w}\pi_{i1(w)}^{Z_{i1(w)}}\cdots\pi_{iQ(w)}^{Z_{iQ(w)}}\\
&\propto \pi_{i(w)}^{\alpha_{i1(w)}+\sum_{V_m=w}Z_{i1(w)}-1}\cdots\pi_{iQ(w)}^{\alpha_{iQ(w)}+\sum_{V_m=w}Z_{iQ(w)}-1},
\end{aligned}$$

which implies that $\pi_{i(w)}|\ldots \sim \text{Dirichlet}(\alpha_{i1(w)}+\sum_{V_m=w}Z_{i1(w)},\ldots,\alpha_{iQ(w)}+\sum_{V_m=w}Z_{iQ(w)})$.

For each $w = 1,\ldots,W$, the full conditional distribution of $p_{in(w)}$ given all other variables is proportional to

$$\begin{aligned}
&P(p_{in(w)}) \propto P(p_{in(w)})P(X|V_m=w,Z,\mu,\tau,p_{in(w)},p_{out(w)})\\
&\propto p_{in(w)}^{\alpha_{p_{in(w)}}-1}\left(1-p_{in(w)}\right)^{\beta_{p_{in(w)}}-1}p_{in(w)}^{\sum_{V_m=w}\sum_{1\le i<j\le n}1(X_{ij(m)}\neq 0)\sum_{q=1}^{Q}Z_{iq(m)}Z_{jq(m)}}\\
&\times\left(1-p_{in(w)}\right)^{\sum_{V_m=w}\sum_{1\le i<j\le n}1(X_{ij(m)}=0)\sum_{q=1}^{Q}Z_{iq(m)}Z_{jq(m)}},
\end{aligned}$$

which implies that $p_{in(w)}|\ldots \sim \text{Beta}(n_{A(w)}+\alpha_{p_{in(w)}}, n_{C(w)}+\beta_{p_{in(w)}})$.

Similarly, $p_{out(w)}|\ldots \sim \text{Beta}(n_{B(w)}+\alpha_{p_{out(w)}}, n_{D(w)}+\beta_{p_{out(w)}})$, where

$$\begin{aligned}
&A(w) = \{m,i,j,q : V_m = w, 1\le i<j\le n, X_{ij(m)}\neq 0, Z_{iq(m)}\cdot Z_{jq(m)} = 1\}, n_{A(w)} = |A(w)|,\\
&B(w) = \{m,i,j,q\neq r : V_m = w, 1\le i<j\le n, X_{ij(m)}\neq 0, Z_{iq(m)}\cdot Z_{jr(m)} = 1\}, n_{B(w)} = |B(w)|,\\
&C(w) = \{m,i,j,q : V_m = w, 1\le i<j\le n, X_{ij(m)} = 0, Z_{iq(m)}\cdot Z_{jq(m)} = 1\}, n_{C(w)} = |C(w)|,\\
&\text{and } D(w) = \{m,i,j,q\neq r : V_m = w, 1\le i<j\le n, X_{ij(m)} = 0, Z_{iq(m)}\cdot Z_{jr(m)} = 1\}, n_{D(w)} = |D(w)|.
\end{aligned}$$

For $w = 1, \ldots, W$, the full conditional distribution of $\mu_{in(w)}$ given the other variables is proportional to

$$
\begin{aligned}
&P(\mu_{in(w)}|\tau_{in(w)}) \prod_{V_m=w} \prod_{1\le i<j\le n} P(X_{ij(m)}|V_m = w, Z_{ij(m)}, \mu_w, \tau_w, p_{in(w)}, p_{out(w)}) \\
&\propto \exp\{\frac{\tau_{in(w)}}{2\sigma^2_{0,in(w)}}(\mu_{in(w)} - \mu_{0,in(w)})^2\} \times \\
&\prod_{V_m=w} \prod_{1\le i<j\le n} \exp\{-\frac{\tau_{in(w)}}{2}(X_{ij(m)} - \mu_{in(w)})^2 \ 1(X_{ij(m)} \neq 0)\} \sum_{q=1}^{Q} Z_{iq(m)} Z_{jq(m)},
\end{aligned}
$$

which implies that

$$
\mu_{in(w)}|\ldots \sim \mathrm{N}\left(\frac{\sigma^2_{0,in(w)} \sum\limits_{A(w)} X_{ij(m)} + \mu_{0,in(w)}}{n_{A(w)}\sigma^2_{0,in(w)} + 1}, \frac{\sigma^2_{0,in(w)}\tau^{-1}_{in(w)}}{n_{A(w)}\sigma^2_{0,in(w)} + 1} \right).
$$

Similarly,

$$
\mu_{out(w)}|\ldots \sim N\left(\frac{\sigma^2_{0,out(w)} \sum\limits_{B(w)} X_{ij(m)} + \mu_{0,out(w)}}{n_{B(w)}\sigma^2_{0,out(w)} + 1}, \frac{\sigma^2_{0,out(w)}\tau^{-1}_{out(w)}}{n_{B(w)}\sigma^2_{0,out(w)} + 1} \right).
$$

Finally, for $w = 1, \ldots, W$ the full conditional distribution of $\tau_{in(w)}$ given the other variables is proportional to

$$
\begin{aligned}
&P(\tau_{in(w)})P(\mu_{in(w)}|\tau_{in(w)}) \prod_{V_m=w} \prod_{1\le i<j\le n} P(X_{ij(m)}|V_m = w, Z_{ij(m)}, \mu_w, \tau_w, p_{in(w)}, p_{out(w)}) \\
&\propto \tau^{\alpha_{0,in(w)}-1}_{in(w)} \exp\{-\beta_{0,in(w)}\tau_{in(w)}\} \left(\frac{\tau_{in(w)}}{2\pi\sigma^2_{0,in(w)}}\right)^{1/2} \exp\{\frac{\tau_{in(w)}}{2\sigma^2_{0,in(w)}}(\mu_{in(w)} - \mu_{0,in(w)})^2\} \times \\
&\prod_{V_m=w} \prod_{1\le i<j\le n} \left(p_{in(w)}(\frac{\tau_{in(w)}}{2\pi})^{1/2} \exp\{-\frac{\tau_{in(w)}}{2}(X_{ij(m)} - \mu_{in(w)})^2\}\right)^{1(X_{ij(m)}\neq 0) \sum\limits_{q=1}^{Q} Z_{iq(m)}Z_{jq(m)}} \\
&\propto \tau^{\left(\frac{n_{A(w)}+1}{2} + \alpha_{0,in(w)} - 1\right)}_{in(w)} \times \\
&\exp\{\frac{-\tau_{in(w)}}{2} \sum_{A(w)} \left(X_{ij(m)} - \mu_{in(w)}\right)^2 - \frac{\tau_{in(w)}}{2\sigma^2_{0,in(w)}} \left(\mu_{in(w)} - \mu_{0,in(w)}\right)^2 - \beta_{0,in(w)}\tau_{in(w)}\},
\end{aligned}
$$

which implies that

$$
\tau_{in(w)}|\ldots \sim \mathrm{Gamma}\left(\tilde{\alpha}_{in(w)}, \tilde{\beta}_{in(w)}\right),
$$

where

$$\tilde{\alpha}_{in(w)} = \frac{n_{A(w)} + 1}{2} + \alpha_{0,in(w)}$$
$$\text{and } \tilde{\beta}_{in(w)} = \frac{1}{2} \sum_{A(w)} \left(X_{ij(m)} - \mu_{in(w)} \right)^2 + \frac{1}{2\sigma^2_{0,in(w)}} \left(\mu_{in(w)} - \mu_{0,in(w)} \right)^2 + \beta_{0,in(w)}.$$

Similarly,

$$\tau_{out(w)} | \ldots \sim \text{Gamma} \left(\tilde{\alpha}_{out(w)}, \tilde{\beta}_{out(w)} \right),$$

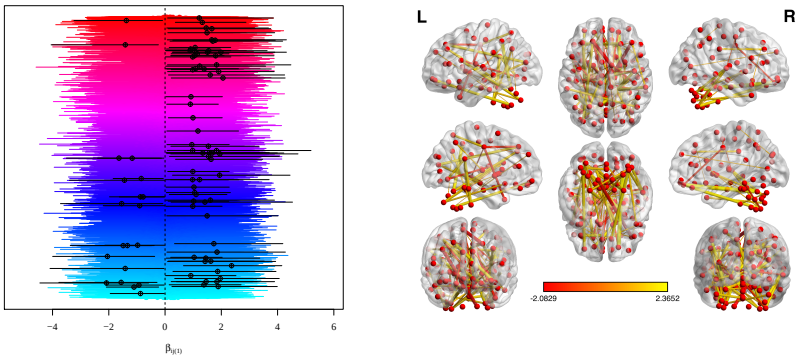
with

$$\tilde{\alpha}_{out(w)} = \frac{n_{B(w)} + 1}{2} + \alpha_{0,out(w)}$$
$$\text{and } \tilde{\beta}_{out(w)} = \frac{1}{2} \sum_{B(w)} \left(X_{ij(m)} - \mu_{out(w)} \right)^2 + \frac{1}{2\sigma^2_{0,out(w)}} \left(\mu_{out(w)} - \mu_{0,out(w)} \right)^2 + \beta_{0,out(w)}.$$

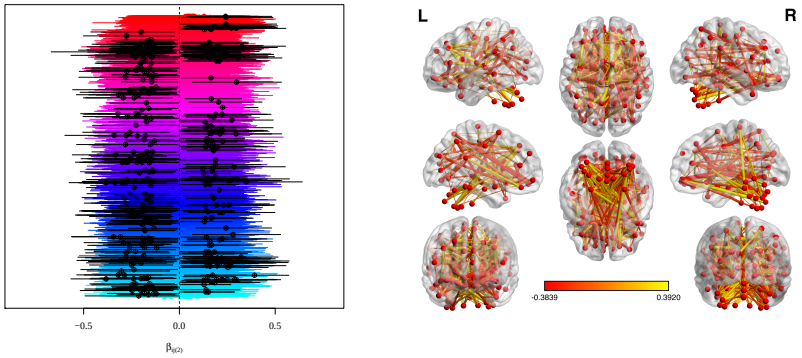
APPENDIX C: CHAPTER 4

C.1 Figures

Figure C.1.1: Parameter estimates: gender and age



(a) Gender



(b) Age

Figure C.1.1. Coefficient estimates comparing the probability of a latent connection between all $n \cdot (n-1)/2$ pairs of nodes. (L) β_{ij} posterior medians and 99% credible intervals and (R) the posterior medians just for the "significant" estimates where zero was not contained in the credible interval.

C.2 Distributions

C.2.1 Prior and sampling distributions

Prior and sampling distributions are listed below.

$$P(Z^*_{ij(m)}|x_m,\beta_{ij}) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}(Z^*_{ij(m)} - x_m^\top\beta_{ij})^2\},$$

where the distribution of the regression coefficients is given by

$$P(\beta_{ij}|b_{0,ij},\Lambda_{0,ij}) = \frac{1}{(2\pi)^{p+1}|\Lambda^{-1}_{0,ij}|}\exp\{-\frac{1}{2}(\beta_{ij} - b_{0,ij})^\top\Lambda_{0,ij}(\beta_{ij} - b_{0,ij})\}$$

with $b_{0,ij} = (b_{00,ij}, b_{01,ij}, \ldots, b_{0(p+1),ij})$ indicating the prior mean, $\Lambda_{0,ij}$ denoting the $(p+1)\times(p+1)$ prior precision matrix for the coefficients, and x_m denoting the covariate vector for subject m (including a one for the intercept). Then WLOG, we assume Gaussian-weighted edges for each subject $m = 1,\ldots,M$:

$$P(Y_{ij(m)}|Z^*_{ij(m)},\ldots) = \big[D_{in(m)}(i,j)\big]^{1(Z^*_{ij(m)}>0)}\big[D_{out(m)}(i,j)\big]^{1(Z^*_{ij(m)}\leq 0)},$$

where

$$D_{in(m)}(i,j) = \big[p_{in(m)}\big(\frac{\tau_{in(m)}}{2\pi}\big)^{1/2}\exp\{-\frac{\tau_{in(m)}}{2}(Y_{ij(m)} - \mu_{in(m)})^2\}\big]^{1(Y_{ij(m)}\neq 0)}(1-p_{in(m)})^{1(Y_{ij(m)}=0)}$$

and

$$D_{out(m)}(i,j) = \big[p_{out(m)}\big(\frac{\tau_{out(m)}}{2\pi}\big)^{1/2}\exp\{-\frac{\tau_{out(m)}}{2}(Y_{ij(m)}-\mu_{out(m)})^2\}\big]^{1(Y_{ij(m)}\neq 0)}(1-p_{out(m)})^{1(Y_{ij(m)}=0)}.$$

Then, for $m = 1,\ldots,M$ the edge distributions are as follows:

$$P(p_{in(m)}) = \frac{1}{B(\alpha_{p_{in}},\beta_{p_{in}})}\ p_{in(m)}^{\alpha_{p_{in}}-1}(1-p_{in(m)})^{\beta_{p_{in}}-1},$$

$$P(p_{out(m)}) = \frac{1}{B(\alpha_{p_{out}},\beta_{p_{out}})}\ p_{out(m)}^{\alpha_{p_{out}}-1}(1-p_{out(m)})^{\beta_{p_{out}}-1},$$

$$P(\mu_{in(m)}) = \left(\frac{1}{2\pi\sigma^2_{0,in}}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\sigma^2_{0,in}}(\mu_{in(m)} - \mu_{0,in})^2\},$$

$$P(\mu_{out(m)}|\mu_{in(m)}) \propto \left(\frac{1}{2\pi\sigma^2_{0,out(w)}}\right)^{\frac{1}{2}}\exp\{\frac{\tau_{out(m)}}{2\sigma^2_{0,out}}(\mu_{out(m)} - \mu_{0,out})^2\}\ 1(\mu_{in(m)} > \mu_{out(m)}),$$

$$P(\tau_{in(m)}) = \beta_{0,in}^{\alpha_{0,in}}\ \frac{1}{\Gamma(\alpha_{0,in})}\ \tau_{in(m)}^{\alpha_{0,in}-1}\exp\{-\beta_{0,in}\tau_{in(m)}\},$$

$$P(\tau_{out(m)}) = \beta_{0,out}^{\alpha_{0,out}}\ \frac{1}{\Gamma(\alpha_{0,out})}\ \tau_{out(m)}^{\alpha_{0,out}-1}\exp\{-\beta_{0,out}\tau_{out(m)}\}.$$

C.2.2 Full conditional posterior distributions

Here we show the derivations of all full conditional posterior distributions. The full conditional distribution of $Z^*_{ij(m)}$ given all other variables is proportional to

$$\begin{aligned}
&P(Z^*_{ij(m)}|x_m, \beta_{ij})P(Y_{ij(m)}|Z^*_{ij(m)}, \ldots) \\
&\propto \exp\{-\frac{1}{2}(Z^*_{ij(m)} - x_m^\top \beta_{ij})^2\}\big[D_{in(m)}(i,j)\big]^{1(Z^*_{ij(m)}>0)}\big[D_{out(m)}(i,j)\big]^{1(Z^*_{ij(m)}\leq 0)} \\
&= \exp\{-\frac{1}{2}(Z^*_{ij(m)} - x_m^\top \beta_{ij})^2\}\big[1(Z^*_{ij(m)} > 0)D_{in(m)}(i,j) + 1(Z^*_{ij(m)} \leq 0)D_{out(m)}(i,j)\big]
\end{aligned}$$

This implies that

$$P(Z^*_{ij(m)}|\ldots) = \frac{\phi(Z^*_{ij(m)} - x_m^\top \beta_{ij})\big[1(Z^*_{ij(m)} > 0)\cdot D_{in(m)}(i,j) + 1(Z^*_{ij(m)} \leq 0)\cdot D_{out(m)}(i,j)\big]}{D_{in(m)}(i,j)\big[1 - \Phi(-x_m^\top \beta_{ij})\big] + D_{out(m)}(i,j)\Phi(-x_m^\top \beta_{ij})},$$

and so

$$\begin{aligned}
Z^*_{ij(m)}|\ldots \sim\ & \frac{D_{in(m)}\big[1 - \Phi(-x_m^\top \beta_{ij})\big]}{D_{in(m)}\big[1 - \Phi(-x_m^\top \beta_{ij})\big] + D_{out(m)}\Phi(-x_m^\top \beta_{ij})}\mathrm{TN}(x_m^\top \beta_{ij}, 1, -\infty, 0) + \\
& \frac{D_{out(m)}\Phi(-x_m^\top \beta_{ij})}{D_{in(m)}\big[1 - \Phi(-x_m^\top \beta_{ij})\big] + D_{out(m)}\Phi(-x_m^\top \beta_{ij})}\mathrm{TN}(x_m^\top \beta_{ij}, 1, 0, \infty),
\end{aligned}$$

where $\mathrm{TN}(\mu, \sigma^2, a, b)$ refers to the truncated normal distribution with mean and variance (from the untruncated distribution) μ and σ^2, and truncation interval (a, b).
Then the full conditional distribution for each $\beta_{ij},\ 1 \leq i < j \leq n$ is proportional to

$$\begin{aligned}
&P(\beta_{ij}|b_{0,ij}, \Lambda_{0,ij}) \prod_{m=1}^{M} P(Z^*_{ij(m)}|x_m, \beta_{ij}) \\
&\propto \exp\{-\frac{1}{2}\sum_{m=1}^{M}(Z^*_{ij(m)} - x^\top \beta_{ij})^2 - \frac{1}{2}(\beta_{ij} - b_{0,ij})^\top \Lambda_{0,ij}(\beta_{ij} - b_{0,ij})\} \\
&\propto \exp\{-\frac{1}{2}(Z^*_{ij} - X\beta_{ij})^\top (Z^*_{ij} - X\beta_{ij}) - \frac{1}{2}(\beta_{ij} - b_{0,ij})^\top \Lambda_{0,ij}(\beta_{ij} - b_{0,ij})\} \\
&= \exp\{-\frac{1}{2}(\beta_{ij} - \beta^*_{ij})^\top (X^\top X + \Lambda_{0,ij})(\beta_{ij} - \beta^*_{ij})\},
\end{aligned}$$

where $\beta^*_{ij} = (X^\top X + \Lambda_{0,ij})^{-1}(X^\top Z^*_{ij} + \Lambda_{0,ij} b_{0,ij})$ and $Z^*_{ij} = (Z^*_{ij(1)}, \ldots, Z^*_{ij(M)})^\top$. This implies that

$$\beta_{ij}|\ldots \sim N_{(p+1)}\Big(\beta^*_{ij}, (X^\top X + \Lambda_{0,ij})^{-1}\Big).$$

The full conditional distribution for $p_{in(m)},\ m = 1, \ldots, M$, given the other variables, is proportional to

$$\begin{aligned} & P(p_{in(m)}) \prod_{1 \le i < j \le n} P(Y_{ij(m)} | p_{in(m)}, \ldots) \\ & \propto \left[p_{in(m)}^{\alpha_{p_{in}} - 1} (1 - p_{in(m)})^{\beta_{p_{in}} - 1} \right] \left[p_{in(m)}^{n_{A(m)}} (1 - p_{in(m)})^{n_{C(m)}} \right], \end{aligned}$$

which implies that

$$p_{in(m)} | \ldots \sim \text{Beta}\Big(\alpha_{p_{in}} + n_{A(m)}, \beta_{p_{in}} + n_{C(m)} \Big),$$

and similarly,

$$p_{out(m)} | \ldots \sim \text{Beta}\Big(\alpha_{p_{out}} + n_{B(m)}, \beta_{p_{out}} + n_{D(m)} \Big),$$

where

$$\begin{aligned} A(m) &= \{ i < j : Z^*_{ij(m)} > 0, Y_{ij(m)} \neq 0 \},\ n_{A(m)} = |A(m)|, \\ B(m) &= \{ i < j : Z^*_{ij(m)} \le 0, Y_{ij(m)} \neq 0 \},\ n_{B(m)} = |B(m)|, \\ C(m) &= \{ i < j : Z^*_{ij(m)} < 0, Y_{ij(m)} = 0 \},\ n_{C(m)} = |C(m)|, \\ D(m) &= \{ i < j : Z^*_{ij(m)} \le 0, Y_{ij(m)} = 0 \},\ n_{D(m)} = |D(m)|. \end{aligned}$$

The full conditional distribution for $\mu_{in(m)}$ given all other variables is proportional to

$$\begin{aligned} & P(\mu_{in(m)}) P(\mu_{out(m)} | \mu_{in(m)} \prod_{1 \le i < j \le n} P(Y_{ij(m)} | \mu_{in(m)}, \ldots) \\ & \propto \exp\{ -\frac{1}{2\sigma^2_{0,in}} (\mu_{in(m)} - \mu_{0,in})^2 \}\ 1(\mu_{in(m)} > \mu_{out(m)})\ \exp\{ \sum_{A(m)} -\frac{\tau_{in(m)}}{2} (Y_{ij(m)} - \mu_{in(m)})^2 \}, \end{aligned}$$

which implies that

$$\mu_{in(m)} | \ldots \sim \text{TN}\left(\frac{\sigma^2_{0,in} \sum\limits_{A(m)} Y_{ij(m)} + \mu_{0,in} \tau^{-1}_{in(m)}}{n_{A(m)} \sigma^2_{0,in} + \tau^{-1}_{in(m)}}, \frac{\sigma^2_{0,in} \tau^{-1}_{in(m)}}{n_{A(m)} \sigma^2_{0,in} + \tau^{-1}_{in(m)}}, \mu_{out(m)}, \infty \right),$$

and similarly,

$$\mu_{out(m)} | \ldots \sim \text{TN}\left(\frac{\sigma^2_{0,out} \sum\limits_{B(m)} Y_{ij(m)} + \mu_{0,out} \tau^{-1}_{out(m)}}{n_{B(m)} \sigma^2_{0,out} + \tau^{-1}_{out(m)}}, \frac{\sigma^2_{0,out} \tau^{-1}_{out(m)}}{n_{B(m)} \sigma^2_{0,out} + \tau^{-1}_{out(m)}}, -\infty, \mu_{in(m)} \right).$$

Finally, the full conditional distribution for $\tau_{in(m)}$ given the other variables is proportional

to

$$P(\tau_{in(m)}) \prod_{1\le i<j\le n} P(Y_{ij(m)}|\tau_{in(m)}\ldots)$$
$$\propto \tau_{in(m)}^{\alpha_{0,in}-1}\exp\{-\beta_{0,in}\tau_{in(m)}\}\ \tau_{in(m)}^{n_{A(m)}/2}\exp\{-\frac{\tau_{in(m)}}{2}\sum_{A(m)}(Y_{ij(m)}-\mu_{in(m)})^2\},$$

which implies that

$$\tau_{in(m)}|\ldots \sim \text{Gamma}\left(\frac{n_{A(m)}}{2}+\alpha_{0,in}, \frac{1}{2}\sum_{A(m)}(Y_{ij(m)}-\mu_{in(m)})^2+\beta_{0,in}\right),$$

and similarly,

$$\tau_{out(m)}|\ldots \sim \text{Gamma}\left(\frac{n_{B(m)}}{2}+\alpha_{0,out}, \frac{1}{2}\sum_{B(m)}(Y_{ij(m)}-\mu_{out(m)})^2+\beta_{0,out}\right).$$

BIBLIOGRAPHY

Abbe, E., Bandeira, A. S., and Hall, G. (2014), "Exact recovery in the stochastic block model," *arXiv preprint arXiv:1405.3267*, 1–31.

Ahmadlou, M., Adeli, H., and Adeli, A. (2012), "Graph theoretical analysis of organization of functional brain networks in ADHD," *Clinical EEG and neuroscience*, 43, 5–13.

Aicher, C., Jacobs, A. Z., and Clauset, A. (2015), "Learning latent block structure in weighted networks," *Journal of Complex Networks*, 3, 221–248.

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9, 33–40.

Ambroise, C. and Matias, C. (2012), "New consistent and asymptotically normal parameter estimates for random-graph mixture models," *Journal of the Royal Statistical Society. Series B*, 74, 3–35.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013), "Pseudo-likelihood methods for community detection in large sparse networks," *The Annals of Statistics*, 41, 2097–2122.

Attias, H. (2000), "A variational Bayesian framework for graphical models," *Advances in Neural Information Processing Systems (NIPS)*, 12, 209–215.

Bates, D. and Eddelbuettel, D. (2013), "Fast and Elegant Numerical Linear Algebra Using the RcppEigen Package," *Journal of Statistical Software*, 52, 1–24.

Biernacki, C., Celeux, G., and Govaert, G. (2000), "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719–725.

Bullmore, E. and Sporns, O. (2009), "Complex brain networks: graph theoretical analysis of structural and functional systems." *Nature Reviews. Neuroscience*, 10, 186–98.

— (2012), "The economy of brain network organization," *Nature Reviews Neuroscience*, 13, 336–349.

Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., Song, X.-W., Xia, M.-R., Shu, N., Dong, Q., et al. (2014), "Topological organization of the human brain functional connectome across the lifespan," *Developmental cognitive neuroscience*, 7, 76–93.

Castellanos, F. X. and Proal, E. (2012), "Large-scale brain systems in ADHD: beyond the prefrontal–striatal model," *Trends in cognitive sciences*, 16, 17–26.

Celisse, A., Daudin, J.-J., and Pierre, L. (2012), "Consistency of maximum-likelihood and variational estimators in the stochastic block model," *Electronic Journal of Statistics*, 6, 1847–1899.

Chen, P.-Y. and Hero III, A. O. (2015), "Universal phase transition in community detectability under a stochastic block model," *Physical Review E*, 91.

Chin, P., Rao, A., and Vu, V. (2015), "Stochastic Block Model and Community Detection in the Sparse Graphs : A spectral algorithm with optimal rate of recovery," *arXiv preprint arXiv:1501.05021*, 1–23.

Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2012), "Stochastic blockmodels with a growing number of classes," *Biometrika*, 99, 273–284.

Cocchi, L., Bramati, I. E., Zalesky, A., Furukawa, E., Fontenelle, L. F., Moll, J., Tripp, G., and Mattos, P. (2012), "Altered functional brain connectivity in a non-clinical sample of young adults with attention-deficit/hyperactivity disorder," *The Journal of Neuroscience*, 32, 17753–17761.

Côme, E. and Latouche, P. (2015), "Model selection and clustering in stochastic block models with the exact integrated complete data likelihood," *Statistical Modelling*.

Cron, A. J. and West, M. (2011), "Efficient Classification-Based Relabeling in Mixture Models," *Am. Stat.*, 65, 16–20.

Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., McGuire, P., and Bullmore, E. T. (2013), "Cognitive relevance of the community structure of the human brain functional coactivation network," *Proceedings of the National Academy of Sciences*, 110, 11583–11588.

Daudin, J. J., Picard, F., and Robin, S. (2008), "A mixture model for random graphs," *Statistics and Computing*, 18, 173–183.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011), "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, 84, 1–19.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B*, 39, 1–38.

Dickerson, B. C. and Sperling, R. A. (2009), "Large-scale functional brain network abnormalities in Alzheimer's disease: insights from functional neuroimaging," *Behavioural neurology*, 21, 63–75.

Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D., and Ushey, K. (2011), "Rcpp: Seamless R and C++ integration," *Journal of Statistical Software*, 40, 1–18.

Erdös, P. and Rényi, A. (1959), "On random graphs I." *Publ. Math. Debrecen*, 6, 290–297.

Fortunato, S. (2010), "Community detection in graphs," *Physics Reports*, 486, 75–174.

Fortunato, S. and Barthélemy, M. (2007), "Resolution limit in community detection." *Proceedings of the National Academy of Sciences of the United States of America*, 104, 36–41.

Fronczak, P., Fronczak, A., and Bujok, M. (2013), "Exponential random graph models for networks with community structure," *Physical Review E*, 88, 032810.

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2015), "Achieving Optimal Misclassification Proportion in Stochastic Block Model," *arXiv preprint arXiv:1505.03772*, 1–39.

Gilbert, E. N. (1959), *Random Graphs*, vol. 30.

Girvan, M. and Newman, M. E. J. (2002), "Community structure in social and biological networks." *Proceedings of the National Academy of Sciences of the United States of America*, 99, 7821–7826.

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*, 170, 301–354.

Hartigan, J. A. and Wong, M. A. (1979), "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society C*, 28, 100–108.

Herlau, T., Schmidt, M. N., and Mø rup, M. (2014), "Infinite-degree-corrected stochastic block model," *Physical Review E*, 90.

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.

Hofman, J. M. and Wiggins, C. H. (2008), "A Bayesian Approach to Network Modularity," *Physical review letters*, 100.

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

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.

Hubert, L. and Arabie, P. (1985), "Comparing partitions," *Journal of Classification*, 2, 193–218.

Karrer, B. and Newman, M. E. J. (2011), "Stochastic blockmodels and community structure in networks," *Physical Review E*, 83, 1–10.

Kolaczyk, E. D. (2009), *Statistical Analysis of Network Data: Methods and Models*, Springer Science+Business Media.

Konrad, K. and Eickhoff, S. B. (2010), "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder," *Human Brain Mapping*, 31, 904–916.

Lancichinetti, A., Kivelä, M., Saramäki, J., and Fortunato, S. (2010), "Characterizing the community structure of complex networks." *PloS one*, 5, e11976.

Latouche, P., Birmelé, E., and Ambroise, C. (2011), "Overlapping stochastic block models with application to the French political blogosphere," *Annals of Applied Statistics*, 5, 309–336.

Latouche, P., Birmele, E., and Ambroise, C. (2012), "Variational Bayesian Inference and Complexity Control for Stochastic Block Models," *Statistical Modelling*, 2, 93–115.

Lei, J. (2014), "A Goodness-of-fit Test for Stochastic Block Models," *arXiv preprint arXiv:1412.4857*, 1–20.

Lei, J. and Rinaldo, A. (2015), "Consistency of Spectral Clustering in Sparse Stochastic Block Models," *The Annals of Statistics*, 43, 215–237.

Lyzinski, V., Tang, M., Athreya, A., Park, Y., and Priebe, C. E. (2015), "Community Detection and Classification in Hierarchical Stochastic Blockmodels," *arXiv preprint arXiv:1503.02115*, 1–24.

Mariadassou, M. and Matias, C. (2015), "Convergence of the groups posterior distribution in latent or stochastic block models," *Bernoulli*, 21, 537–573.

Mariadassou, M., Robin, S., and Vacher, C. (2010), "Uncovering latent structure in valued graphs: A variational approach," *Annals of Applied Statistics*, 4, 715–742.

Matias, C. and Miele, V. (2015), "Statistical clustering of temporal networks through a dynamic stochastic block model," *arXiv preprint arXiv:1506.07464*.

McDaid, A. F., Murphy, T. B., Friel, N., and Hurley, N. J. (2013), "Improved Bayesian inference for the stochastic block model with application to large networks," *Computational Statistics and Data Analysis*, 60, 12–31.

Mena, R. H. and Walker, S. G. (2014), "On the Bayesian mixture model and identifiability," *Journal of Computational and Graphical Statistics*.

Mossel, E., Neeman, J., and Sly, A. (2014), "Reconstruction and estimation in the planted partition model," *Probability Theory and Related Fields*, 2, 1–36.

Newman, M. E. J. (2004), "Analysis of weighted networks," *Physical Review E*, 70, 9.

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. (2013), "Bayesian Degree-Corrected Stochastic Blockmodels for Community Detection," *arXiv preprint arXiv:1309.4796*, 1–23.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11.

Qiu, M.-g., Ye, Z., Li, Q.-y., Liu, G.-j., Xie, B., and Wang, J. (2011), "Changes of brain structure and function in ADHD children," *Brain topography*, 24, 243–252.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a), "An introduction to exponential random graph (p*) models for social networks," *Social Networks*, 29, 173–191.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007b), "Recent developments in exponential random graph (p*) models for social networks," *Social Networks*, 29, 192–215.

Rohe, K., Chatterjee, S., and Yu, B. (2011), "Spectral clustering and the high-dimensional stochastic blockmodel," *Annals of Statistics*, 39, 1878–1915.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., and Chun, M. M. (2015), "A neuromarker of sustained attention from whole-brain functional connectivity," *Nature neuroscience*.

Schweinberger, M. and Handcock, M. S. (2015), "Local Dependence in Random Graph Models: Characterisation, Properties, and Statistical Inference," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 647–676.

Schweinberger, M. and Luna, P. (2015), "HERGM: Hierarchical Exponential-Family Random Graph Models," *Journal of Statistical Software*.

Simpson, S. L., Bowman, F. D., and Laurienti, P. J. (2013), "Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain," *Statistics Surveys*, 7, 1–36.

Simpson, S. L., Hayasaka, S., and Laurienti, P. J. (2011), "Exponential random graph modeling for complex brain networks," *PLoS ONE*, 6, e20039.

Snijders, T. A. B. and Nowicki, K. (1997), "Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure," .

Sporns, O., Tononi, G., and Kötter, R. (2005), "The human connectome: a structural description of the human brain," *PLoS Comput Biol*, 1, e42.

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

Sussman, D. L., Tang, M., Fishkind, D. E., and Priebe, C. E. (2012), "A consistent adjacency embedding for stochastic blockmodel graphs," *Journal of the American Statistical Association*, 107, 17.

Suwan, S., Lee, D. S., Tang, R., Sussman, D. L., Tang, M., and Priebe, C. E. (2014), "Empirical Bayes estimation for the stochastic blockmodel," *arXiv preprint arXiv:1405.6070*, 1–19.

Tallberg, C. (2005), "A Bayesian Approach To Modeling Stochastic Blockstructures With Covariates," *The Journal of Mathematical Sociology*, 29, 1–23.

Traag, V. A. and Bruggeman, J. (2009), "Community detection in networks with positive and negative links," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 80, 7.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002), "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain." *NeuroImage*, 15, 273–289.

Vu, D. Q., Hunter, D. R., and Schweinberger, M. (2013), "Model-based clustering of large networks," *Annals of Applied Statistics*, 7, 1010–1039.

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., and He, Y. (2013), "Disrupted functional brain connectome in individuals at risk for Alzheimer's disease," *Biological psychiatry*, 73, 472–481.

Wang, Y. X. R. and Bickel, P. J. (2015), "Likelihood-based model selection for stochastic block models," *arXiv preprint arXiv:1502.02069*, 1–28.

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.

Wilson, J. D., Wang, S., Mucha, P. J., Bhamidi, S., and Nobel, A. B. (2014), "A Testing Based Extraction Algorithm for Identifying Significant Communities in Networks," *The Annals of Applied Statistics*, 8, 1853–1891.

Xia, M., Wang, J., and He, Y. (2013), "BrainNet Viewer: a network visualization tool for human brain connectomics," *PloS one*, 8, e68910.

Yan, X., Jensen, J., Krzakala, F., Moore, C., Shalizi, C. R., Zdeborova, L., Zhang, P., and Zhu, Y. (2014), "Model Selection for Degree-corrected Block Models," *Journal of Statistical Mechanics: Theory and Experiment*.

Yang, J. and Leskovec, J. (2012), "Community-affiliation graph model for overlapping network community detection," *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1, 1170–1175.

— (2014), "Structure and Overlaps of Ground-Truth Communities in Networks," *ACM Transactions on Intelligent Systems and Technology*, 5, 26.

Zhang, J., Wang, J., Wu, Q., Kuang, W., Huang, X., He, Y., and Gong, Q. (2011), "Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder," *Biological psychiatry*, 70, 334–342.

Zhang, Y., Friend, A. J., Traud, A. L., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2008), "Community structure in Congressional cosponsorship networks," *Physica A: Statistical Mechanics and its Applications*, 387, 1705–1712.

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