High-Dimensional Directed Network Analysis of Human Brains

by

Yaotian Wang

B.S. in Mathematics and Applied Mathematics, Peking University, 2017

M.S. in Statistics, University of Virginia, 2020

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yaotian Wang

It was defended on

March 27, 2023

and approved by

Tingting Zhang, Department of Statistics

Yu Cheng, Department of Statistics

Satish Iyengar, Department of Statistics

Guofen Yan, Department of Public Health Sciences, University of Virginia

Dissertation Advisor: Tingting Zhang, Department of Statistics

High-Dimensional Directed Network Analysis of Human Brains

Yaotian Wang, PhD

University of Pittsburgh, 2023

The brain is the most complex organ in the human body. Studying the functional organization of such a complex organ is fascinating. As statisticians, we study the human brain's functional organization by developing statistical modeling methods for brain data. Recent technological development brings opportunities and challenges: On one hand, enormous quantities of brain data in various modalities are produced in many fields, including biology, neurology, neuroscience, psychology, and psychiatry. On the other hand, human brain data bring new challenges to data analysts because their unique properties differ from conventional big data. This dissertation proposes novel statistical modeling methods to address these challenges and understand the brain's functional organization.

The human brain is a high-dimensional directed network system consisting of many regions as network nodes that exert influence on each other. The directed influence from one region to another is called directed connectivity and corresponds to one directed edge in the directed brain network. We understand the brain's functional organization by investigating how brain regions interact and form different network patterns when performing different brain functions. This dissertation illustrates two of our methods for revealing high-dimensional directed brain networks.

Chapter 2 explains a new Bayesian model for studying directed brain networks of patients with epilepsy using their intracranial electroencephalography (EEG) data. Epilepsy is a directed network disorder, as epileptic activity spreads from a seizure onset zone (SOZ) to many other regions after seizure onset. Intracranial EEG data are multivariate time series recordings of many brain regions. Using our model, we revealed the evolution of brain networks during seizure development and uncovered unique directed connectivity properties of the SOZ.

Chapter 3 presents a new Bayesian model for characterizing whole-brain directed networks of the healthy human population based on functional magnetic resonance imaging

(fMRI) data. We also propose a computationally efficient algorithm to address the challenge of analyzing thousands of subjects' fMRI data. Using our new model and algorithm, we analyzed the resting-state fMRI data of around one thousand subjects from the Human Connectome Project (HCP) and revealed both population-mean and subject-specific whole-brain directed networks of them.

Keywords: Brain network, directed connectivity, epileptic network, multi-subject fMRI, variational Bayes.

Table of Contents

List of Tables

List of Figures

Acknowledgement

From the bottom of my heart, I would like to express my sincere gratitude to my dissertation advisor Professor Tingting Zhang. Her training substantially developed my research skills and improved my thinking and communications skills. This led to my remarkable progress over the past six years. I am also deeply grateful to her for spending a lot of time preparing me for the academic job market. She makes it possible for me to pursue a career in scientific research, which has been my goal since childhood.

I would also like to thank the rest of my dissertation committee members, Professor Yu Cheng, Professor Satish Iyengar, and Professor Guofen Yan, for their questions, suggestions, patience, and trust during my Ph.D. research.

I also wish to thank all the members of Professor Zhang's lab for their collaboration and support in work and life.

Additionally, I am thankful to all my collaborators, who open-up my eyes outside the Statistics. Learning scientific knowledge from them helps me identify good and interesting scientific questions.

Finally, I would like to express my thank to my parents, my aunt, and my grandparents. I genuinely appreciate their love, encouragement, and support over nearly thirty years. Especially their support helps me to concentrate on my research during the pandemic. I would like to dedicate this dissertation to them.

1.0 Introduction

The brain is the most complex organ in the human body. It is truly fascinating to study the functional organization of such a complex organ. We want to understand how the brain's functional organization brings about different brain functions and how it differs in psychiatric and neurological illnesses. As statisticians, we study the human brain by developing statistical modeling methods for brain data. Recent technological development brings opportunities and challenges: On one hand, enormous quantities of human brain data in various modalities have been produced in many fields, including biology, neurology, neuroscience, psychology, and psychiatry. On the other hand, as one of the most commonly cited forms of big data, human brain data bring new challenges to data analysts because of brain data's unique properties that are different from conventional data. We develop new statistical modeling methods for brain data in various modalities to address these challenges and understand the human brain.

The human brain is a high-dimensional directed network system (Van Den Heuvel and Pol, 2010) consisting of many brain regions as network nodes that exert influence on each other. The directed influence from one region to another is called directed connectivity (also called directed functional connectivity or effective connectivity (Friston, 1994; Mill et al., 2017)) and corresponds to one directed edge in the directed brain network. To understand how brain regions interact with each other and form different brain network patterns when performing different brain functions, we develop statistical modeling methods to reveal high-dimensional directed brain networks using brain data.

At the beginning of the dissertation research, we wanted to start by analyzing high-quality brain data and studying a clear network phenomenon. The intracranial electroencephalography (EEG) data from patients with epilepsy satisfy these two requirements, as explained below. Epilepsy is the fourth most common neurological disorder, affecting approximately 3 million people in the United States. In the epilepsy setting, the propagation of seizures from a seizure onset zone (SOZ) to other healthy brain regions is a clinically important example of pathological directed brain networks (Bernhardt et al., 2015; Englot

et al., 2016). Patients with epilepsy typically take medications to control their symptoms. Unfortunately, about 40% of patients develop drug-resistant epilepsy (Burns et al., 2014). The best chance for these patients to achieve seizure control is surgically resecting the SOZ. Therefore, localization of the SOZ is critical to the success of the surgery. To localize the SOZ, intracranial EEG data are recorded using cortical or depth electrodes (typically, $>$ 30 regions). Given that intracranial EEG is an invasive method, its recordings have high temporal and spatial resolution and a strong signal-to-noise ratio (SNR) (Cervenka et al., 2013). Overall, seizures are a clear pathological network phenomenon, and intracranial EEG data from patients with epilepsy are of high quality. As a result, our initial study focused on investigating the epileptic high-dimensional directed brain network using intracranial EEG data.

Existing directed brain network results of epilepsy were mostly about low-dimensional directed networks between the SOZ and a few contiguous regions (Korzeniewska et al., 2014). Results of high-dimensional directed networks between the SOZ, its adjacent regions, and many other non-SOZ regions are lacking. To address this knowledge gap, We develop a new statistical model to characterize the SOZ, its adjacent regions, and many other non-SOZ regions as one integrated high-dimensional directed network system. We develop a Bayesian method to produce efficient estimates of the model parameters by incorporating the prior information that high-dimensional brain networks tend to have a modular organization (Meunier et al., 2009; Newman, 2006; Park and Friston, 2013; Sporns and Betzel, 2016). The modular organization (also called modularity) is a network pattern in which network nodes within the same module interact more strongly and densely with each other than with nodes in different modules. Overall, one crucial novelty of the proposed Bayesian method is to develop a unified Bayesian framework to simultaneously identify modules and directed connections among a large number of brain regions.

We applied the proposed Bayesian method to intracranial EEG recordings from six patients with medically intractable focal epilepsy. We revealed the evolution of brain networks from normal to abnormal epileptic states and uncovered unique directed connectivity properties of the SOZ during seizure development. To demonstrate the effectiveness of our method in uncovering different directed connectivity properties of different brain regions, we de-

veloped a new SOZ localization method based on the network results and compared our identified SOZ candidate regions against patients' clinically localized SOZ (used as the given truth). Our method achieved high accuracies in localizing the SOZ (100% true positive rates and less than 3% false positive rates) for all six patients.

Despite the success in revealing the brain network evolution and uncovering directed connectivity properties, our network results based on intracranial EEG data have several limitations. First, since intracranial EEG is an invasive method, intracranial EEG data are generally collected from patients with epilepsy or brain tumors and cannot be collected from healthy subjects. Second, intracranial EEG records the activity of only a small portion of the brain. Therefore, we cannot investigate whole-brain directed networks using intracranial EEG data. Third, intracranial EEG electrode placement varies across different patients. As a result, our analysis of intracranial EEG data can only be performed subject by subject and cannot be generalized to the population of subjects. As our understanding of the human brain deepened, we wanted to scale up our study for whole-brain directed networks of the healthy human population. Therefore, we started to study many subjects' functional magnetic resonance imaging (fMRI) data, as they provide non-invasive measurements of the activity of the entire human brain with a high spatial resolution (Lindquist, 2008).

Given that whole-brain directed networks also have modular organization because of functional specialization and integration (Friston, 1994), we extend the previous Bayesian model to analyze resting-state functional magnetic resonance imaging (fMRI) data from thousands of subjects and characterize their whole-brain directed networks. To accommodate the variation of brain networks across subjects, we build a Bayesian framework allowing each region to be in different modules and to have different directed connections in different subjects' brain networks. We also develop a novel variational Bayesian method for fMRI data analysis to address the computational challenges that arise. As far as we know, this is the first method that can simultaneously identify modules and directed connections in the whole-brain directed networks of many subjects.

Using our new method, we analyzed the resting-state fMRI data of around one thousand subjects from the Human Connectome Project (Van Essen et al., 2013, HCP). We revealed both population-mean and subject-specific whole-brain directed networks. We identified

modules in accord with functional brain systems specialized for various functions (Power et al., 2011), revealed directed connections between modules with different specialized functions, and uncovered several regions that could be involved in more than one brain function. We also assessed our method's reproducibility by comparing independent analysis results of different fMRI runs of the same subjects.

The rest of the dissertation is organized as follows. Chapter 2 illustrates the study on directed brain networks of patients with epilepsy using intracranial EEG Data. The methodology has been previously published in (Li et al., 2021). The real intracranial EEG data analysis has been previously published in (Wang et al., 2022). Chapter 3 explains the study on whole-brain directed networks of many subjects using fMRI data. Wang et al. (2023) is the published version. Chapter 4 describes the future research.

2.0 Study on Directed Brain Networks of Patients with Epilepsy Using Intracranial EEG Data

The majority of this chapter has been previously published in (Li et al., 2021; Wang et al., 2022).

2.1 Introduction

Brain activities form a directed network, where network nodes are brain regions and each network edge represents a directed influence exerted by one region on another. Such directed information flow from one region to another is referred to as directed connectivity also called directed functional connectivity or effective connectivity (Friston, 1994; Mill et al., 2017). The purposes of this chapter are to present a new statistical approach for the analysis of intracranial electroencephalographic (intracranial EEG) data and to use our approach to uncover the normal and abnormal directed brain networks of patients with epilepsy over the course of seizure development.

Seizures are a directed network phenomenon (Rosenow and Lüders, 2001), as abnormal, excessive, and synchronous neuronal activities start from the seizure onset zone (SOZ) and propagate to otherwise healthy brain regions. Brain surgery to remove the SOZ is a common treatment consideration for patients with drug-resistant epilepsy. Pre-surgical evaluation includes localization of the SOZ using intracranial EEG, which is absolutely critical to the success of the surgery. Clinicians place intracranial EEG electrodes on the exposed brain (inside the skull) of patients with epilepsy to record their neuronal activities in many regions. The recorded data are high-dimensional multivariate time series of voltage waveforms, which often exceed 50 channels (with each channel corresponding to one region). Figure 2.4b shows the electrode placement of a patient who underwent intracranial EEG recordings in epilepsy evaluation. Figure 2.4a illustrates a segment of intracranial EEG recordings of this patient.

To localize the SOZ, trained EEG experts visually examine intracranial EEG waveforms

and designate the region that first shows abnormal epileptic activity to be the SOZ (Jacobs et al., 2012). However, despite careful planning, sometimes visual analysis of intracranial EEG fails to localize the SOZ clearly (Harroud et al., 2012). One crucial reason is that sometimes seizure onsets consist of low amplitude, very fast activity. This activity may not generate appropriate power that can be visually detected until the seizure is well underway. Activity with greater power that can be identified may occur later, by which time seizure activity has spread beyond the actual SOZ and involves brain regions that are involved in seizure occurrence but do not serve as the electrical source. Given that seizures are a directed network phenomenon, our method for mapping directed brain networks (i.e., identifying directed connections) using intracranial EEG data is expected to improve understanding of the brain system and localization of the SOZ.

Intracranial EEG data are high-dimensional multivariate time series recordings of many small regions' neuronal activities at a high temporal resolution (millisecond scale) and spatial resolution (about 10 mm in diameter) and with a strong signal-to-noise ratio (SNR) (Cervenka et al., 2013), in contrast to popular functional magnetic resonance imaging (fMRI) with a low temporal resolution and scalp EEG with a low spatial resolution. As such, intracranial EEG data provide valuable information about directed brain networks.

Mapping directed brain networks based on high-dimensional multivariate time series, however, faces multiple challenges. First, it is difficult to construct a model that can accurately characterize the complex mechanism of a high-dimensional brain system, i.e., how each region's activity depends on many others' activities. Second, the estimation of a high-dimensional model has a large variance. With many regions being studied and enormous possibilities in directed connections among the regions, it is challenging to identify only a few strong connections among enormous candidate ones. Though incorporating anatomic connectivity (AC) information into the directed connectivity model can improve the estimation of directed connections (Hahn et al., 2019), AC information is not always available. Here, we consider mapping directed brain networks without relying on AC information. Simple sparsity regularization does not address the challenge because high-dimensional sparse networks have many different forms, most of which do not accurately reflect the brain's functional organization. For example, standard L_1-regularized estimates (Basu and Michailidis, 2015;

Nicholson et al., 2017) lead to a sparse network in which every region has only a few connections with other regions. However, this sparse network is inconsistent with known brain networks in which regions with similar functions tend to be closely connected (Petersen and Sporns, 2015). Third, the computation for analyzing high-dimensional multivariate time series data can be intensive. Existing approaches to mapping directed networks usually address only a part of these challenges, as explained below.

Network mapping approaches fall into two major categories: information-theoretic-measure-based methods and model-based methods. The former includes correlations, cross-correlations (Kramer et al., 2008; Schiff et al., 2005), cross-coherence (Schröder and Ombao, 2019), transfer entropy (Vicente et al., 2011), directed transinformation (Hinrichs et al., 2006), directed information (Liu and Aviyente, 2012), and many others (Van Mierlo et al., 2013; Wilke et al., 2011). Although these measures are fast to compute, they are mainly for quantifying the pairwise relationship between regions and ignore system features of the brain in which each region's activity depends on many other regions' activities. Thus, information-theoretic-measure-based approaches lack the ability to delineate the entire signal pathway of directed connections from regions to regions.

Model-based methods have been developed to describe simultaneous directed connectivity among all the recorded regions. The most popular models include dynamic causal modeling (DCM, Friston et al., 2003) and neural mass models (NMM, David and Friston, 2003), which use ordinary differential equations (ODE) to characterize directed connectivity. Because of their complex mathematical formulation, the DCM and NNM are typically used for low-dimensional brain networks (consisting of only a few brain regions being studied). To address this limitation, Zhang et al. (2020, 2015, 2017) proposed to use linear ODEs to approximate high-dimensional brain systems (consisting of many regions). However, parameter estimation of deterministic ODE models is sensitive to model specification, data noise, and data-sampling frequency.

We propose to use a multivariate autoregressive state-space (MARSS) model for intracranial EEG data to address the limitation of existing methods. First, the state-space framework allows for separating the model error due to the inherent model inadequacy for a complex system and the data measurement error. The MARSS with the two errors is flex-

ible to approximate different systems and is robust to various deviations from the assumed model. Equally importantly, the formulation of MARSS is much simpler than ODE models, which thus, enables fast computation for high-dimensional data.

Different from standard MAR (Goebel et al., 2003; Harrison et al., 2003; Korzeniewska et al., 2008) and MARSS (Cheung et al., 2010; Riera et al., 2004), our MARSS is uniquely constructed for analyzing intracranial EEG data to map directed brain networks. It has been widely documented (Newman, 2006; Sporns, 2010) that brain networks have the modular organization (or modularity), in which regions are more densely connected with regions in the same module than with regions otherwise. Our approach incorporates the modular organization to greatly improve the model estimation. Specifically, we propose a stochastic blockmodel (SBM)-motivated prior for the MARSS parameters, restricting the estimated network to have the modular organization. The SBM (Airoldi et al., 2008; Durante and Dunson, 2014; Geng et al., 2019; Nowicki and Snijders, 2001) is a generative model for networks that have the modular organization. However, existing applications of the SBM (Arroyo-Relión et al., 2017; Paul and Chen, 2018) and most module identification methods (also called community detection, a terminology often used in social network literature) (Goldenberg et al., 2010; Zhao et al., 2012) are for observed networks with known edges. The proposed method addresses a more challenging problem of inferring *unobserved networks* based on multivariate time series measurements of network nodes' activities.

Using the SBM-motivated prior for MARSS parameters, we develop a Bayesian framework to make inferences about the underlying network. The proposed Bayesian approach has three major advantages. First, our method improves the efficiency in identifying connected brain regions (i.e., a high true positive) and produces scientifically interpretable network results by incorporating the modular organization into the model. Second, the proposed Bayesian framework accounts for the model error due to the model inadequacy for the complex system as well as the statistical uncertainty in identifying connected regions. Third, the simple MARSS formulation brings the flexibility to approximate various brain systems and enables fast computation for high-dimensional multivariate time series data. As such, our approach effectively addresses the three major challenges in mapping high-dimensional brain networks.

We apply the new MARSS model to multi-channel intracranial EEG recordings from six patients with epilepsy. We show the evolution of these patients' high-dimensional directed networks from interictal to ictal phases. Changes in directed connections and modules are uncovered not only for the SOZ and adjacent regions but also for many more distant non-SOZ regions. We also use these network results to localize the SOZ independently from traditional visual analysis in clinical practice to demonstrate the effectiveness of our method in uncovering different directed connectivity properties of different regions. We compare identified candidate SOZ regions against patients' clinically localized SOZ (used as the given truth). We show that our method achieved high accuracies in localizing the SOZ (100% true positive rates and less than 3% false positive rates) for all six patients.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the new MMSAR model for directed brain networks with the modular organization. We build a Bayesian hierarchical model with an SBM-motivated prior to make inferences of MARSS parameters and develop an efficient Markov chain Monte Carlo (MCMC) simulation algorithm for the ensuing posterior inference. In Section 2.3, we apply the developed Bayesian model to data simulated under two different model settings and network patterns and compare the ensuing results with those of existing network mapping methods. We show that the proposed method is robust to various deviations from the assumed model and outperforms existing methods by achieving much higher accuracy in identifying connected brain regions. Section 2.4 illustrates the analysis of intracranial EEG recordings of six patients with epilepsy using our new MMSAR model. Section 2.5 concludes with a discussion.

2.2 Dynamic System Models and Bayesian Inference

2.2.1 The Multivariate Autoregressive State-Space Model

Let $\boldsymbol{y}(t) = (y_1(t), \ldots, y_d(t))'$ be observed intracranial EEG measurements of d brain regions (equivalently d network nodes of the brain network under study) at time t and $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))'$ be the underlying neuronal state functions of the d brain regions

at time t for $t = 1, \ldots, T$. Since each intracranial EEG electrode directly records one brain region's neuronal activity with high spatial and temporal resolutions, we propose a simple space model that links $y_i(t)$ to $x_i(t)$:

$$y_i(t) = c_i \cdot x_i(t) + \epsilon_i(t), \quad i = 1, \ldots, d, \tag{2.1}$$

where c_i is a unknown constant, and $\epsilon_i(t)$ is a data measurement error with mean zero.

For the state model that describes directed connectivity among the d regions at the neuronal level, we propose to use the simplest dynamic system model, i.e., the first-order multivariate-autoregression (MAR), for $\boldsymbol{x}(t)$:

$$x_i(t) = \sum_{j=1}^{d} A_{ij} \cdot x_j(t-1) + \eta_i(t), \quad i = 1, \ldots, d, \ t = 1, \ldots, T,$$

where $\eta_i(t)$ is the model error due to the model inadequacy in characterizing the dynamics of region i.

Our goal is to develop a parsimonious model to detect the existence of temporal dependence among neuronal activities of regions rather than building a comprehensive model that can explain all the neuronal activities. Due to the high-dimensionality and the current limited understanding of the brain system, it is extremely difficult to build such a comprehensive dynamic system model. Even though more complex models, such as high-order MARs, may fit the observed data better, they still suffer from the model inadequacy. More seriously, high-order MARs have large estimation errors because they have at least d^2 more parameters than first-order MARs. Consequently, the first-order MAR is more efficient for detecting connected regions and addresses our needs.

Under the MARSS, identifying connected regions and mapping the brain network are equivalent to selecting statistically significant nonzero A_{ij}s. To distinguish nonzero directed connections from zero ones, we introduce indicators for A_{ij}s:

$$x_i(t) = \sum_{j=1}^{d} \gamma_{ij} \cdot A_{ij} \cdot x_j(t-1) + \eta_i(t), \quad i = 1, \ldots, d, \ t = 1, \ldots, T, \tag{2.2}$$

where γ_{ij} is an indicator, taking values of either 0 or 1. We use γ_{ij}s to stand for the set of indicators $\{\gamma_{ij}, \ i, j = 1, \ldots, d\}$. The use of indicators is similar to the "spike and slab"

prior (Ishwaran and Rao, 2005; Miller, 2002; Theo and Mike, 2004) in the Bayesian variable selection framework (Brown et al., 1998; George and McCulloch, 1993, 1997; Yi et al., 2003). Under (2.2), identifying connected brain regions, i.e., selecting directed network edges, is equivalent to selecting nonzero γ_{ij}s, which is the focus of our model estimation.

The observation model (2.1) and the state model (2.2) together are the proposed MARSS for the brain's directed connectivity. Note that the first-order MARSS is different from the first-order MAR: The former is robust to violations of model assumptions, but the latter is not. This is because the MARSS uses two error terms, $\eta_i(t)$ and $\epsilon_i(t)$, to accommodate the model inadequacy and measurement error separately.

We let $\eta_i(t) \overset{\text{i.i.d}}{\sim} \text{N}(0,1)$ for several reasons. First, c_i in (1) and the variance of $\eta_i(t)$ are not uniquely defined. Since we treat the former as unknown, we fix the latter at 1 to avoid the identifiability issue. Second, letting $\eta_i(t)$ be independent between regions enables γ_{ij} and A_{ij} to capture the dependence between regions more efficiently than otherwise. Third, letting $\eta_i(t)$ be independent over time brings parsimony to the model. Again, our purpose is to detect the existence of temporal dependence between regions' intracranial EEG rather than capturing all possible temporal dependence. Similarly, for the latter two reasons, we let $\epsilon_i(t) \overset{\text{i.i.d}}{\sim} \text{N}(0,\tau_i)$. We show through simulation studies (Section 2.3) that our approach is robust to violations of model assumptions.

2.2.2 Bayesian Hierarchical Model for MARSS

Since nonzero γ_{ij}s define the brain's directed network, we impose the modular organization on the estimated brain network by using a stochastic blockmodel (SBM)-motivated (Airoldi et al., 2008; Durante and Dunson, 2014; Fienberg et al., 1985; Nowicki and Snijders, 2001) prior for γ_{ij}s. The modular organization means that regions within the same module connect more closely with each other than with regions in a different module. The modular organization fits the brain's functional organization reported in the literature (Newman, 2006; Sporns, 2010) and is also useful in epilepsy diagnosis. For example, regions in the SOZ's module are those affected by the SOZ's activities most. Information about the SOZ's module and its changes during seizure development can help neurologists assess the effect of

seizures on brain functions. In summary, developing the SBM-motivated prior for MARSS parameters to impose the modular organization on estimated networks is another important novelty of our approach.

Let K be the pre-specified number of modules. Let $\boldsymbol{m}_i = (m_{i1}, \ldots, m_{iK})'$ be a K-dimensional vector with only one element being 1 and the rest being 0; $\boldsymbol{m}_i$ labels the module of region i, i.e., $m_{ik} = 1$ indicates region i in the kth module. Let $B_{k_1k_2}$, $k_1, k_2 = 1, \ldots, K$, denote the prior probability of a nonzero directed connection from a region in module k_2 to another region in module k_1. Let $\mathbf{B}$ be a $K \times K$ matrix with entries $B_{k_1k_2}$ for $k_1, k_2 = 1, \ldots, K$.

2.2.2.1 Prior Specification for the Modular Organization

The prior for the brain network with the modular organization is a joint distribution for indicators γ_{ij}s, the module labels $\boldsymbol{m}_i$s, and the probability matrix $\mathbf{B}$ as follows:

$$\gamma_{ij} | \boldsymbol{m}_i, \boldsymbol{m}_j, \mathbf{B} \overset{\text{ind}}{\sim} \text{Bernoulli}(\boldsymbol{m}_i' \, \mathbf{B} \, \boldsymbol{m}_j); \tag{2.3}$$

$$\boldsymbol{m}_i \overset{\text{i.i.d}}{\sim} \text{Multinomial}(1; p_1, \ldots, p_K) \text{ for } i = 1, \ldots, d, \text{ and } (p_1, \ldots, p_K) \sim \text{Dirichlet}(\boldsymbol{\alpha}) \tag{2.4}$$

$$B_{kk} \overset{\text{i.i.d}}{\sim} \text{Uniform}(l_0, 1) \text{ and } B_{k_1k_2} \overset{\text{i.i.d}}{\sim} \text{Uniform}(0, u_0), k, k_1, k_2 = 1, \ldots, K, \; k_1 \neq k_2; \tag{2.5}$$

where l_0 and u_0 are given constants between 0 and 1, and $\boldsymbol{\alpha} = (1, \ldots, 1)$, assigning uniform weights to different modules. The distribution (2.3) specifies the probabilities of both within-module and between-module connections. For example, if $m_{ik_1} = 1$ and $m_{jk_2} = 1$, then $\boldsymbol{m}_i' \, \mathbf{B} \, \boldsymbol{m}_j = B_{k_1k_2}$, which is the probability of existing a directed connection from module k_2 to module k_1; if $m_{ik} = 1$ and $m_{jk} = 1$, $\boldsymbol{m}_i' \, \mathbf{B} \, \boldsymbol{m}_j = B_{kk}$, which is the prior probability of existing a directed connection between two regions in the same module k. Since within-module connections are dense and strong, while between-module connections are sparse (Park and Friston, 2013), we let $u_0 = 0.1$ and $l_0 = 0.9$. The large difference between u_0 and l_0 facilitates differentiating within-module connections from between-module ones and identifying modules.

The distributions (2.3), (2.4), and (2.5) together define the SBM-motivated prior for γ_{ij}s. Our goal is to identify modules and select significant edges by estimating the module labels for regions, $\boldsymbol{m}_i$s, and the indicators for edges, γ_{ij}s.

2.2.2.2 Prior Specification for A_{ij}s

We assign a normal prior to A_{ij}:

$$A_{ij} \overset{\text{i.i.d}}{\sim} \text{N}(0, \xi_0^2), \tag{2.6}$$

where ξ_0 is a positive constant so that the density of A_{ij} is almost flat within its domain.

2.2.2.3 Priors for Other Parameters

Let $\boldsymbol{x}(0) = (x_1(0), \ldots, x_d(0)$, $\boldsymbol{c} = (c_1, \ldots, c_d)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_d)$. We assign the following priors to the rest parameters:

$$x_i(0) \overset{\text{ind}}{\sim} \text{N}(\mu_i, 1),\ \mu_i \overset{\text{i.i.d}}{\sim} \text{N}(0, \xi_1^2),\ c_i \overset{\text{i.i.d}}{\sim} \text{N}(0, \xi_1^2),\ p(\tau_i) \propto \frac{1}{\tau_i^{1+\rho_0}} \exp\{-\frac{\rho_0}{\tau_i}\},\ i = 1, \ldots, d, \tag{2.7}$$

where ρ_0 is a pre-specified small positive constant to give an almost flat prior for $\boldsymbol{\tau}$ and ξ_1 is a large positive constant to give almost flat priors for c_i and μ_i.

2.2.2.4 Joint Posterior Distribution

All the parameters to be estimated in the proposed Bayesian framework are $\boldsymbol{\Theta} = \{\boldsymbol{\Gamma}, \mathbf{B}, \mathbf{M}, \mathbf{A}, \boldsymbol{c}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{p}\}$, where $\boldsymbol{\Gamma}$ is a $d \times d$ matrix with entries γ_{ij} for $i, j = 1, \ldots, d$, $\mathbf{M}$ is a $K \times d$ matrix with the ith column being $\boldsymbol{m}_i$, $\mathbf{A}$ is a $d \times d$ matrix with entries A_{ij} for $i, j = 1, \ldots, d$, and $\boldsymbol{p} = \{p_1, \ldots, p_K\}$.

Let $\mathbf{X} = \{\boldsymbol{x}(0), \ldots, \boldsymbol{x}(T)\}$ and $\mathbf{Y} = \{\boldsymbol{y}(1), \ldots, \boldsymbol{x}(T)\}$. The MARSS model (2.1) and (2.2) with prior distributions (2.3), (2.4), (2.5), (2.6), and (2.7) lead to the posterior distribution: $p(\mathbf{X}, \boldsymbol{\Theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Theta}) \cdot p(\mathbf{X}|\boldsymbol{\Theta}) \cdot p(\boldsymbol{\Theta})$. The detailed formulation of the joint posterior distribution is provided in the Appendix A.1.

2.2.3 EM Algorithm for Setting Initial Values and Hyperparameter

We simulate from $p(\mathbf{X}, \boldsymbol{\Theta}|\mathbf{Y})$ with a partially collapsed Gibbs Sampler (Van Dyk and Park, 2008), whose Markov Chain Monte Carlo (MCMC) simulation steps are provided in the Appendix A.2.

The MCMC simulation can take many iterations to converge especially for large d. To address this issue, following the practice suggested in (Chapter 13.1, Gelman et al., 2013), we use an expectation-maximization (EM) algorithm to find the starting values for the MCMC simulation. Specifically, we optimize $p(\mathbf{Y}|\hat{\mathbf{\Theta}}) = \int p(\mathbf{Y}|\hat{\mathbf{\Theta}}, \mathbf{X}) \cdot p(\mathbf{X}|\mathbf{\Theta}) d\mathbf{X}$ by the EM algorithm, in which the state functions $\mathbf{X}$ are treated as missing values. The output of the EM algorithm, $\hat{\mathbf{\Theta}}$ in the final step, is used as the initial value for the following MCMC iterations. For all our simulation and real data analysis, we verified that the MCMC algorithm converged upon evaluating the Gelman-Rubin statistic (Gelman and Rubin, 1992).

We need to determine the value of K, the number of modules, for the proposed Bayesian model. Standard approaches to selecting hyperparameters for Bayesian methods include information criteria and cross-validation. However, these methods are time-consuming for large d, because they all require running the posterior simulation for each candidate K. We propose to select the value for K by the EM algorithm. Specifically, we let $K = d$ in our EM algorithm. We set the initial values of m_{ii} to 1 for $i = 1, \ldots, d$, that is, we let each region form one independent module at the start of the EM algorithm. As the algorithm iterates, several regions fall into the same module, and the number of distinct modules of the d regions becomes stable. Since the EM algorithm can find the number of modules that leads to a locally optimal posterior, we let the K in the Bayesian model be the number of distinct modules in the final step of the algorithm.

2.2.4 Posterior Inference

We use two posterior probabilities to map the brain network: $\hat{P}^m_{ij} = \frac{1}{S}\sum_{s=1}^{S} \delta(\boldsymbol{m}_i^{(s)}, \boldsymbol{m}_j^{(s)})$ and $\hat{P}^\gamma_{ij} = \frac{1}{S}\sum_{s=1}^{S} \gamma_{ij}^{(s)}$, where S is the total number of MCMC samples after burn-in. The former is the posterior probability of two regions i and j in the same module, and the latter is the posterior probability of nonzero directed connectivity from region j to i. We use $\hat{P}^m_{ij}$, $i, j = 1, \ldots, d$, to identify modules. Given a threshold $\hbar^m$, if $\hat{P}^m_{ij} > \hbar^m$, regions i and j are put in the same module; if additionally, $\hat{P}^m_{jk} > \hbar^m$, then the three regions i, j, and k are put in the same module regardless of the value of $\hat{P}^m_{ik}$. We use $\hat{P}^\gamma_{ij}$ to select directed network edges. Given a threshold $\hbar^\gamma$, if $\hat{P}^\gamma_{ij} > \hbar^\gamma$, we deem the directed connection from region j to

i nonzero and select the directed network edge from j to i.

2.2.4.1 Choice of Thresholds.

The total numbers of potential network edges and possible network patterns are enormous for high-dimensional networks. Because of the uncertainty resulted from the high dimensionality, posterior probabilities $\hat{P}_{ij}^m$ and $\hat{P}_{ij}^\gamma$ are all small. To address this issue, many Bayesian methods select variables based on the ranks of their posterior probabilities (Li et al., 2015; Zhang et al., 2017). We here propose to determine the thresholds for $\hat{P}_{ij}^m$ and $\hat{P}_{ij}^\gamma$ based on their significance/p-values under the null hypothesis that all the regions are *independent* from each other. For simulated data, we generate a null data set, as explained in detail below.

We first generate a null data set $\mathbf{Y}^0$ that satisfies the null hypothesis. Specifically, given the long simulated time series before the time of our interest, we randomly sample a short segment $Y_i^0 = \{y_i(t), t = t_i + 1, \ldots, t_i + T\}$ of each region i and let the pairwise distance between any two regions' segments, $|t_i - t_j|$, no smaller than $2T$. All the regions' segments Y_i^0, $i = 1, \ldots, d$, form $\mathbf{Y}^0$, in which the temporal dependence of each region's time-series data points remains while the dependence between regions' time series is almost none. Applying our Bayesian method to $\mathbf{Y}^0$, we obtain the ensuing posterior probabilities, which form the empirical null distributions for $\hat{P}_{ij}^m$'s and $\hat{P}_{ij}^\gamma$'s, respectively. We evaluate the p-values of $\hat{P}_{ij}^m$'s and $\hat{P}_{ij}^\gamma$'s based on the null distributions and determine the thresholds for $\hat{P}_{ij}^m$'s and $\hat{P}_{ij}^\gamma$'s corresponding to the chosen p-value. We here use the p-value of 1% to ensure a low false positive rate.

For real intracranial EEG data, we propose to use the baseline period, a period way before seizure onset time, as the null data set. The details are described in Section 2.4.2.

2.3 Simulation Studies

2.3.1 Example 1: Simulation from A Third-Order MARSS

We simulated multivariate time-series data from the following third-order MARSS.

$$\begin{aligned} x_i(t) &= \sum_{j=1}^{d} A_{1,ij}\ x_j(t-1) + \sum_{j=1}^{d} A_{2,ij}\ x_j(t-2) + \sum_{j=1}^{d} A_{3,ij}\ x_j(t-3) + \eta_i(t) \text{ and} \\ y_i(t) &= c_i \cdot x_i(t) + \epsilon_i(t). \end{aligned}$$

The above system has three modules of sizes 15, 15, and 20. We consider region j has a directed influence over i, if at least one of $A_{1,ij}$, $A_{2,ij}$, and $A_{3,ij}$ is nonzero. Figure 2.1a shows the simulated network pattern, where the presence of a directed connection is indicated by an edge (grey edges for within-module connections and purple edges for between-module connections).

We simulated $\eta_i(t)$ from the model

$$\boldsymbol{\eta}(t) = 0.5\boldsymbol{\eta}(t-1) + \boldsymbol{\delta}(t) \text{ and } \boldsymbol{\delta}(t) \overset{i.i.d}{\sim} \text{MNV}(0, \boldsymbol{\Sigma}_1), \tag{2.8}$$

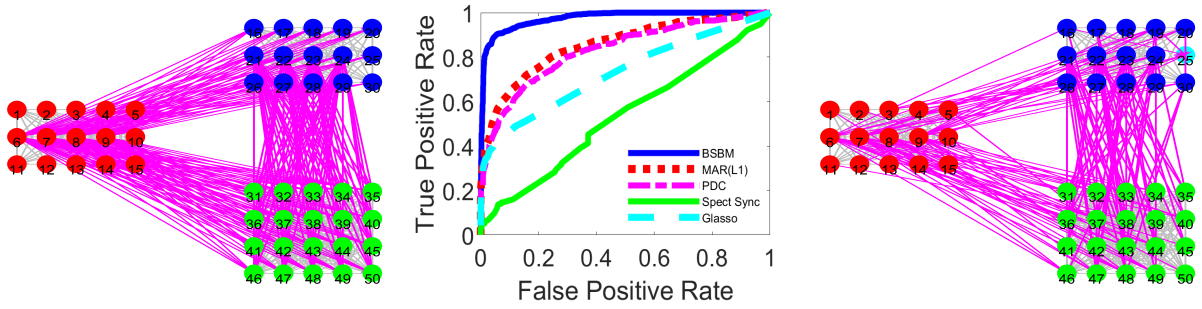
where $\boldsymbol{\Sigma}_1$ is a block diagonal matrix with each block corresponding to one module. The diagonal entries of $\boldsymbol{\Sigma}_1$ all equal 1 and off-diagonal entries in diagonal submatrices follow Uniform(0,0.5). The upper bound of off-diagonal entries is chosen such that $\boldsymbol{\Sigma}_1$ is strictly positive definite.

We generated the observation errors $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_d(t))'$ from the model

$$\boldsymbol{\epsilon}(t) = 0.5\boldsymbol{\epsilon}(t-1) + \boldsymbol{\zeta}(t) \text{ and } \boldsymbol{\zeta}(t) \overset{i.i.d}{\sim} \text{MVN}(0, \mathbf{D}^{\frac{1}{2}}\boldsymbol{\Sigma}_2\mathbf{D}^{\frac{1}{2}}), \tag{2.9}$$

where $\boldsymbol{\Sigma}_2$ is created in the same way as $\boldsymbol{\Sigma}_1$, and $\mathbf{D}$ is a d-by-d diagonal matrix with the diagonal entries chosen such that the SNRs of all the time series equal 10. The median SNR of real intracranial EEG data is much higher than 10 (Zhang et al., 2015). As such, the simulated model errors and data errors are all spatially and temporally correlated, which violates the model assumptions of the proposed MARSS.

Using the simulated edges as the true values, we calculated false positive rates (FPR) and true positive rates (TPR) of network edge selection based on different thresholds for $\hat{P}_{ij}^{\gamma}$s.



(a) Simulated Brain Network (b) ROC Curves (c) Estimated Brain Network

Figure 2.1: Simulation Study of Data Generated from a Third-Order MARSS. (a) The true simulated network structure. (b) The ROC curves of the proposed Bayesian method with the SBM-motivated prior (BSBM) and competing methods including MAR(L_1), PDC, the spectrum synchronicity, and Glasso. (c) The estimated network corresponds to 1% p-value.

For comparison, we examined the FPRs and TPRs of popular competing methods, including the third-order MAR with L_1 regularization (implemented by using the R package BigVAR (Nicholson et al., 2017)), denoted by MAR(L_1), partial directed coherence (PDC) (Baccalá and Sameshima, 2001), the spectrum synchronicity (Euán et al., 2018), and graphical lasso (Glasso) (Friedman et al., 2014; Witten et al., 2011). Figure 2.1b shows the ROC curves of TPRs vs. FPRs for these methods. The proposed Bayesian method with the SBM-motivated prior (BSBM) outperformed the other methods as evidenced by its much greater TPRs given the same FPRs.

Figure 2.1c shows the estimated network pattern using the thresholds corresponding to 1% p-value for $\hat{P}^m_{ij}$ and $\hat{P}^\gamma_{ij}$. The proposed method was able to identify three modules. For detecting the directed connections among the 50 regions, the overall TPR and FPR are 0.84 and 0.02. More specifically, the TPR and FPR are 0.95 and 0 for within-module connections and 0.45 and 0.02 for between-module connections. The comparably low TPR for selecting between-module connections is due to several reasons. First, since the module identification (similar to clustering) is subjective, our selection of directed network edges based on P^γ_{ij} does not account for the identified modules. As within-module connections (accounting for 32.6% of all candidate connections) are much denser than between-module connections (9.0% of all candidate connections), network edge selection is more towards selecting within-module

connections, so the overall network edge selection accuracy is high. Second, the number of candidate between-module connections is enormous and even more than the total number of true network edges. As such, the true between-module connections are highly sparse and more difficult to identify than within-module connections. Third, since the number of null connections is large, we used a high threshold for P_{ij}^{γ} to avoid many false selections, which also leads to a low TPR for selecting between-module connections. Overall, the proposed method outperformed existing methods by achieving a higher TPR and almost zero FPR.

In summary, this simulation demonstrates the robustness of our MARSS to violations of model assumptions and its efficiency in identifying connected regions and modules.

2.3.2 Example 2: Simulations from DCMs

We simulated time series from the dynamic causal modeling (DCM) (Friston et al. (2003); Kiebel et al. (2008)), the most popular ODE-based model for the brain's directed connectivity. We used the simulation studies in this section to assess the robustness and efficiency of our proposed Bayesian method in identifying directed connections for data generated from models highly distinct from our MARSS.

We first used the neuronal state equations (which consist of many ordinary differential equations) in the DCM for EEG data (DCM-EEG) (Kiebel et al. (2008)) to generate the state functions $\boldsymbol{x}(t)$, because both EEG and intracranial EEG signals are mainly from pyramidal neurons. The state equations characterized a directed network among 50 regions with 3 modules. Figure 2.2a shows the simulated network, where an edge indicates the presence of a directed connection between a pair of regions. Based on the generated $\boldsymbol{x}(t)$, we used the observation model (2.1) to simulate $\boldsymbol{y}(t)$, and (2.9) to generate observational errors $\epsilon_i(t)$'s. As such, the data-generating model was distinct from the assumed MARSS. We applied our Bayesian method to the simulated data and obtained posterior probabilities for all the directed connections.

Given the simulated directed connections (directed edges) as the true values, we calculated FPR and TPR of network edge selection based on different thresholds for $\hat{P}_{ij}^{\gamma}$s. For comparison, we also examined FPRs and TPRs of other competing methods. In addition to

the ones mentioned in Section 2.3.1, we apply transfer entropy (TE) (Vicente et al. (2011)) to our simulated data. Note that the DCM (Kiebel et al. (2008)) could not produce the network estimate because its estimation for 50 regions is computationally infeasible. The ROC curves of TPRs vs. FPRs for these methods are shown in Figure 2.2b. BSBM outperformed other methods by achieving a much larger area under the curve.

Secondly, we compared our BSBM with competing methods using fMRI data simulated from a whole-brain directed network model (DCM-fMRI) (Frässle et al. (2018)). Figure 2.2c shows the simulated network, which had the small world architecture of the human brain. Though the proposed MARSS is distinct from the DCM-fMRI, our BSBM achieved the largest area under the ROC curve, as shown in Figure 2.2d. Note that our method also outperformed the sparse regression DCM (srDCM) (Frässle et al. (2018)), which is a recent extension of the DCM-fMRI to high-dimensional directed networks.

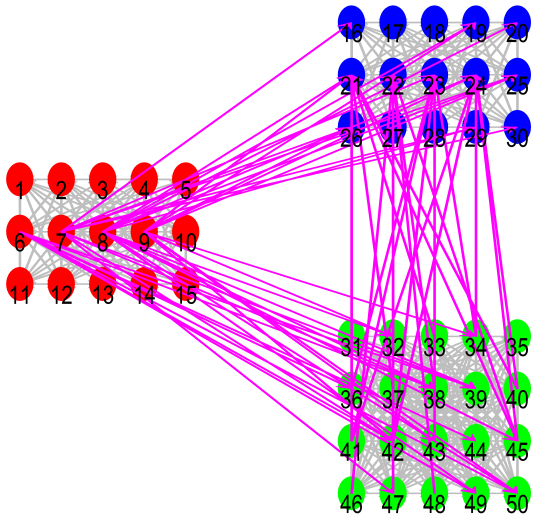
In summary, these simulations demonstrate that the proposed BSBM is still efficient even though the underlying models for the data generation are highly distinct from the proposed MARSS.

2.4 Real Intracranial EEG Data Analysis

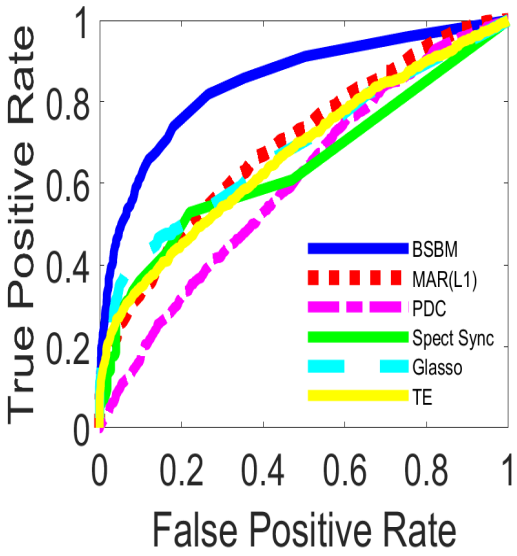
2.4.1 Subjects and Intracranial EEG Recordings

We analyzed intracranial EEG recordings of six patients with medically intractable focal seizures who underwent evaluation for epilepsy surgery. Intracranial EEG implantation was customized for each patient. Each patient had at least three clinical and electrographic seizures with clear SOZs. A board-certified EEG expert examined intracranial EEG data and determined SOZs and seizure onset times. Table 2.1 presents information about six patients with drug-resistant epilepsy in this study. The information includes the number of seizures analyzed, the age of the patient at the recording, epilepsy types, etiology, and surgical outcome for each patient.

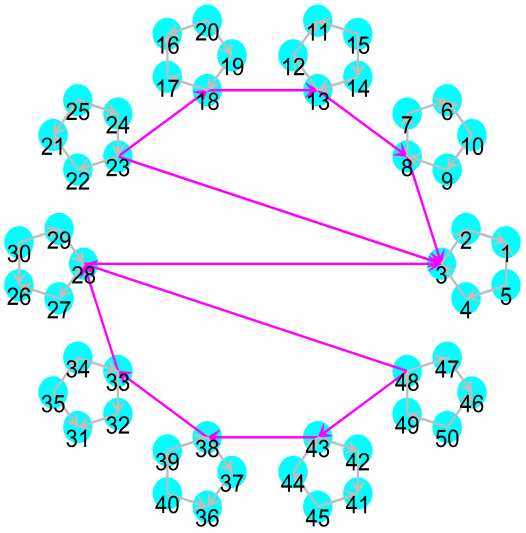
For each patient, we analyzed his/her intracranial EEG recordings from 300 seconds



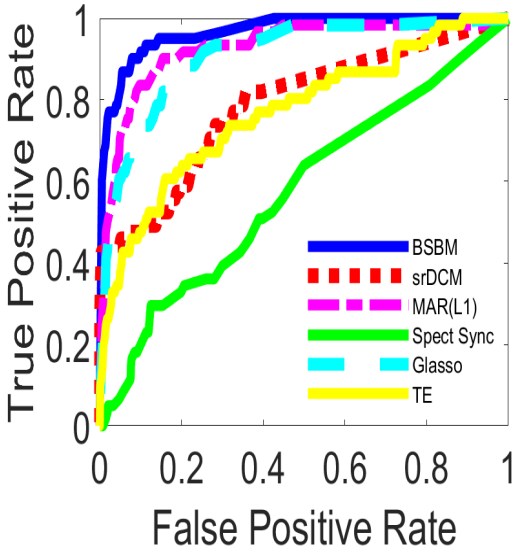
(a) Simulated network DCM-EEG



(b) ROC curves for DCM-EEG



(c) Simulated network DCM-fMRI



(d) ROC curves for DCM-fMRI

Figure 2.2: Simulation Studies of Data Generated from DCMs. (a) The simulated network from the DCM for EEG data. (b) ROC curves for directed connections identified by six methods for the data generated in (a). (c) The simulated network from the DCM for fMRI data. (d) ROC curves for directed connections identified by six methods for the data generated in (c).

Table 2.1: Patient Information.

Patient	Number of Seizures Analyzed	Age	Gender	Epilepsy Type	Etiology	Surgical Outcome
1	3	23	M	left lateral frontal	dysplasia	EC2 (No surgery, but RNS installed)
2	7	57	M	right temporal	gliosis	EC1B
3	4	28	F	left inferior frontal	dysplasia	EC1B
4	4	30	M	left temporal parietal	dysplasia	EC1A
5	5	44	M	left lateral frontal	dysplasia	EC1A
6	3	45	M	right subtemporal	unk	EC3

Age indicates the patient's age at recording.

before to 300 seconds after electrographic seizure onset. The recordings were down-sampled to 1,024 Hz (collected at a 4,096 Hz [Natus, Middleton, WI USA]), filtered with a 60 Hz notch filter, and underwent removal of the first principal component to minimize artifacts. The analysis was blinded to the location of clinically determined SOZ but not the seizure onset time so as to provide a common reference time point with respect to the starting time of electrographic ictal activity. Identified directed brain networks (averaged across seizures) were presented for every 25-second window.

2.4.2 Methods: Identification of Directed Brain Networks

For each patient, we identified his/her directed brain network in any 25-second window by simultaneously identifying directed connections and modules in the network, using $\hat{P}_{ij}^{m}$s and $\hat{P}_{ij}^{\gamma}$s defined in Section 2.2.4, as explained in detail below.

We first used the proposed BSBM to analyze the patient's 1-second intracranial EEG segments independently and obtained the above two posterior probabilities for every 1-second segment. For every pair of regions i and j, we evaluated the evidence for the directed connection from region j to i being present in the window from t to $t+25$ seconds (the $[t, t+25]$ window) by averaging $\hat{P}_{ij}^{\gamma}$ (the posterior probabilities of $\gamma_{ij}=1$) from all non-overlapping 1-second segments in the window across all seizures, where $t=0$ is the seizure onset time, and $t=-300,-299,\ldots,275$. We denoted the ensuing average probability by $\hat{P}_{ij}^{t,\gamma}$ and referred to it as the network edge probability. The network edge probability, $\hat{P}_{ij}^{t,\gamma}$, quantifies how likely, on average, the directed connection from region j to i was present in the 25-second window. Similarly, we obtained the average posterior probability of two regions, i and j, in the same module in the window. This average probability is denoted by $\hat{P}_{ij}^{t,m}$ and referred to as the module probability. The module probability, $\hat{P}_{ij}^{t,m}$, quantifies how likely, on average, regions j and i were in the same module in the $[t, t+25]$ window.

For each 25-second window $[t, t+25]$, we identified its directed brain network by using network edge probabilities, $\hat{P}_{ij}^{t,\gamma}$s, to identify directed connections and by using module probabilities, $\hat{P}_{ij}^{t,m}$s, to identify modules. For every pair of regions, i and j, if $\hat{P}_{ij}^{t,\gamma}$ was greater than a threshold, the directed connection from region j to i was deemed to be present in

that window. Similarly, if $\hat{P}_{ij}^{t,m}$ was greater than a threshold, regions j and i were deemed to be in the same module in that window. To determine the thresholds, we used the earliest 25-second window (275-300 seconds before seizure onset time in our study) as the baseline window. The top 1% percentile of all the $\hat{P}_{ij}^{-300,\gamma}$s in the baseline window was used as the threshold for network edge probabilities (in all windows), and the top 1% percentile of all the $\hat{P}_{ij}^{-300,m}$s in the baseline window was used as the threshold for module probabilities. We used the top 1% percentiles as the thresholds to select the most significant connections and to keep a low false discovery rate. In addition, top 1% percentiles have been commonly used as thresholds in Bayesian variable selection problems (Li et al., 2015; McCann and Welsch, 2007).

2.4.3 Methods: Quantification of the Extent of Directed Connectivity

For each region j, we quantified its extent of directed connectivity in the window from t to $t + 25$ seconds by averaging region j's all the network edge probabilities in the window:

$$D_j^t = \sum_{i=1}^{d} (\hat{P}_{ij}^{t,d} + \hat{P}_{ji}^{t,d})/2d,$$

where d is the total number of regions.

We then examined the change in the extent of directed connectivity for region j at time t by

$$DC_j^t = D_j^t - D_j^{t-25}.$$

That is, DC_j^t is the difference between the average directed connectivity in the two windows that overlap at the time point t only. We evaluated DC_j^t at every one second t. We refer to DC_j^t as the directed connectivity change of region j at time t.

2.4.4 Methods: SOZ Localization

To assess the effectiveness of our method in revealing different regions' directed connectivity properties, we proposed to localize the SOZ based on DC_j^t. Let $t = 0$ denote the seizure onset time. Let $M_j = \max\{DC_j^t, t \leq 0\}$, the maximum directed connectivity change

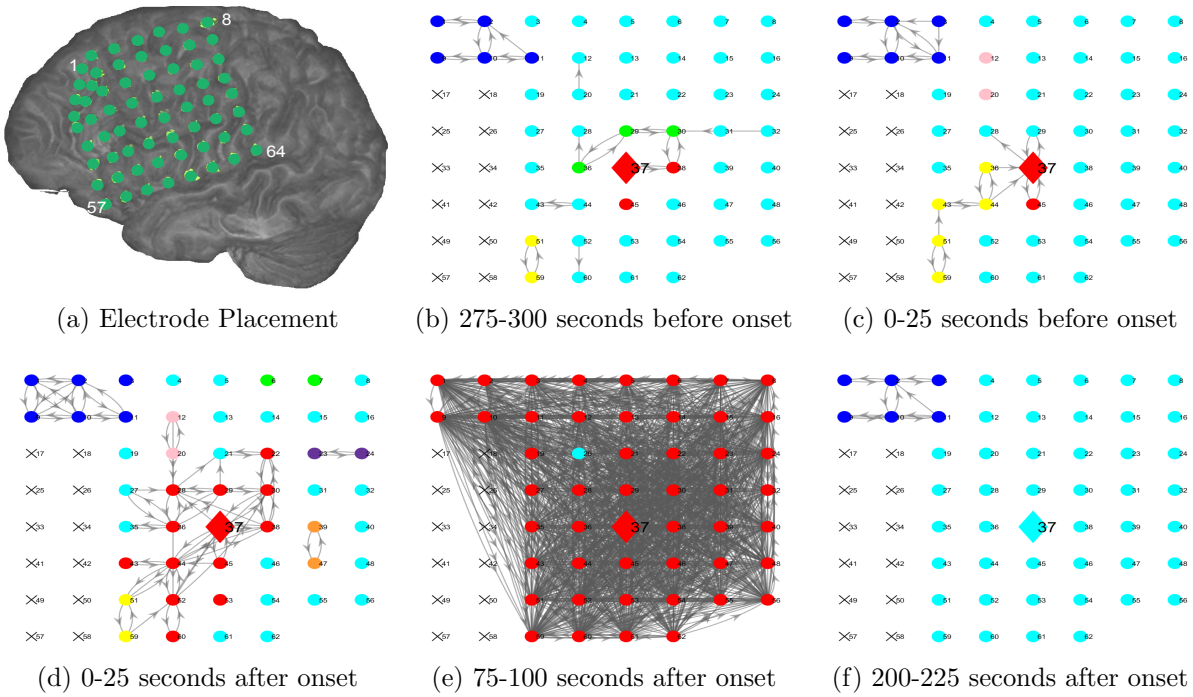
of region j over time no later than the seizure onset time. Let C_0 be the top 10% percentile of $\{DC_j^0, j = 1, \ldots, d\}$, the directed connectivity changes of all the regions at $t = 0$. A region l was selected to be a candidate for the SOZ if DC_l^0, its directed connectivity change at $t = 0$, was greater than both $0.9M_l$ and C_0.

2.4.5 Results: Increase in the Number of Directed Connections of SOZ after Seizure Onset

For all patients analyzed, we found that their interictal directed brain networks were stable up to the seizure onset time. Moreover, directed connections were sparse in these interictal brain networks. Specifically, before seizure onset (Figures 2.3b and 2.3c), the (clinically determined) SOZ was connected to just a few regions, most of which were immediately adjacent to the SOZ. However, immediately after seizure onset (in the window of 0-25 seconds after seizure onset), the number of directed connections significantly increased (Figure 2.3d) with P-values $\leq$0.03 for all patients (see details of hypothesis testing procedures in Appendix A.3). The majority of these new directed connections occurred in the area of the SOZ and its adjacent regions. The number of directed connections continued to grow as seizure progressed (Figure 2.3e). The cessation of seizure activity was marked by a return to the number of directed connections in interictal phases (Figure 2.3f). Overall, the proposed BSBM revealed the evolution of the directed brain network from a normal to abnormal epileptic state and then back to the normal state.

2.4.5.1 Comparison with Other Connectivity Approaches

We compared directed connections identified by BSBM with those identified by three common connectivity methods, including: i) short-time direct directed transfer function (SdDTF) (Korzeniewska et al., 2014), ii) partial directed coherence (PDC) (Baccalá and Sameshima, 2001), and iii) cross-coherence (CC) (Varela et al., 2001). SdDTF and PDC measure directed connectivity. CC measures function connectivity without directionality information. To assess the performance of BSBM and these three methods in detecting brain network changes at the seizure onset time, we compared the number of identified



(a) Electrode Placement (b) 275-300 seconds before onset (c) 0-25 seconds before onset

(d) 0-25 seconds after onset (e) 75-100 seconds after onset (f) 200-225 seconds after onset

Figure 2.3: Patient 1's Directed Brain Networks Identified by the Proposed BSBM. (a) The intracranial EEG electrode placement on the left hemisphere of Patient 1. (b)-(f) The identified directed brain networks from 300 seconds before to 225 seconds after seizure onset. The diamond at electrode G37 is the SOZ identified by expert interpretation of EEG data. Nodes in light blue are the regions that did not belong to any modules. Nodes in the same other colors (either dark blue, green, pink, red, purple, brown, or yellow) denote different identified modules of regions. All nodes in red color belong to the SOZ module. Grey arrows indicate the identified directed connections between regions. Anteroinferior electrodes preceded by an "X" were resected in a previous epilepsy surgery. The evolution of the brain network from 300 seconds before to 300 seconds after seizure onset is shown as a video, which can be downloaded via the link in Appendix A.4.

connections (with or without direction) of the SOZ across three 25-second windows: 275-300 seconds before seizure onset (*baseline*), 0-25 seconds before seizure onset (*immediate preictal*), and 0-25 seconds after seizure onset (*immediate postictal*).

As shown in Table 2.2, BSBM consistently detected increases in the number of directed connections of the SOZ in the immediate postictal window, compared to the preictal windows, for all six patients analyzed. In contrast, none of the three competing methods (SdDTF, PDC, or CC) consistently detected such increases. SdDTF detected increases in the number of directed connections in the immediate postictal window, compared to the baseline and immediate preictal windows, for Patients 1, 3, 5, and 6, but failed to show similar increases in the directed network for Patients 2 and 4. PDC failed to detect more directed connections in the immediate postictal window versus the other two interictal windows for Patients 4 and 6. CC detected zero connections in all the three windows for Patients 1, 3, and 4. Overall, BSBM outperformed the other three methods by consistently detecting changes in the directed brain networks around the SOZs at seizure onset.

2.4.6 Results: Increase in the Number of Regions in the SOZ Module

Our BSBM enables identifying modules of a large number of regions. In this study, the SOZ module is of particular interest because it is a group of regions that are most affected by the activity in the SOZ. With BSBM, we identified the SOZ module and revealed its changes from interictal to ictal phases to demonstrate the effect of seizure propagation on changing the brain network structure.

For all patients analyzed, their SOZ module sizes (the number of regions contained in the identified SOZ module) remained stable over time in interictal phases. As shown in Figures 2.3b and 2.3c, the SOZ module often included only the SOZ and a few immediately adjacent regions, indicating that very few regions were affected by activity in the SOZ before seizure onset. However, immediately after seizure onset, the SOZ module expanded significantly (P-values $\leq$0.02 for all patients) to include more nearby regions (Figures 2.3d). The SOZ module size continued to increase as the seizure progressed (Figure 2.3e). With seizure cessation, similar to the changes in the number of directed connections, the size of the SOZ

Table 2.2: Comparison of Different Connectivity Methods

Methods	Windows	Patients					
		1	2	3	4	5	6
BSBM	Baseline	2	10	3	0	8	16
	Immediate Preictal	7	9	0	0	14	45
	Immediate Postictal	11	15	7	6	63	220
PDC	Baseline	6	0	3	1	15	29
	Immediate Preictal	6	0	4	0	10	18
	Immediate Postictal	8	1	7	1	29	23
SdDTF	Baseline	15	0	0	2	20	75
	Immediate Preictal	7	2	1	1	35	12
	Immediate Postictal	19	2	6	1	53	87
CC	Baseline	0	4	0	0	14	11
	Immediate Preictal	0	6	0	0	30	23
	Immediate Postictal	0	9	0	0	152	49

The numbers of directed connections of the SOZ identified by BSBM versus three other methods, including: i) short-time direct directed transfer function (SdDTF), ii) partial directed coherence (PDC), and iii) cross-coherence (CC) for six patients in the baseline (275-300 seconds before seizure onset), immediate preictal (0-25 seconds before seizure onset), and immediate postictal (0-25 seconds after seizure onset) windows.

module returned to the level in interictal phases.

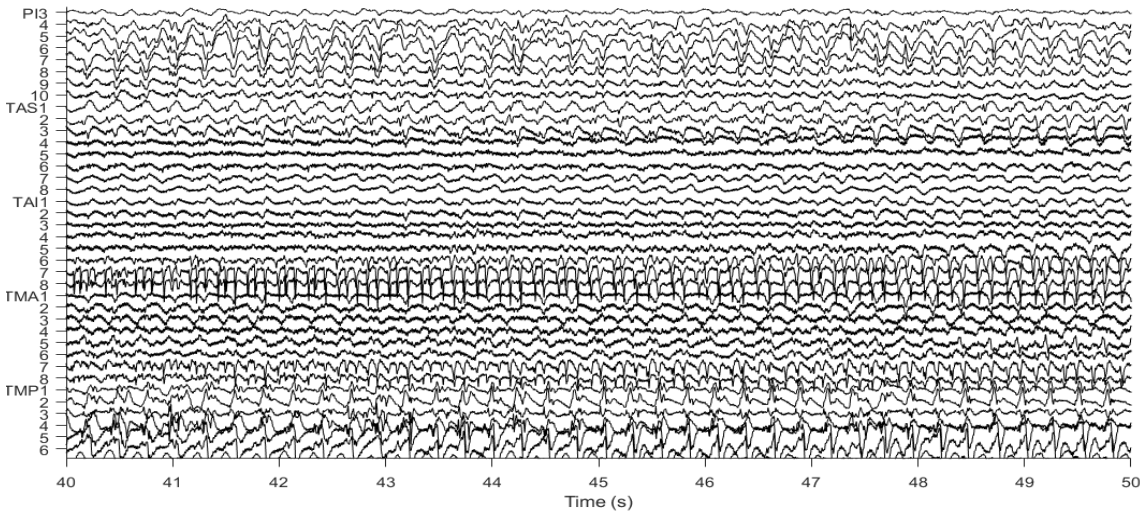
2.4.7 Results: Changes in Directed Connections of Non-SOZ Regions

Most previous studies focused on the changes in the directed brain network around the SOZ only (Jirsa et al., 2014). With BSBM, we distinguished among the SOZ, non-SOZ regions in the SOZ module, and other non-SOZ regions outside the SOZ module. We found that the latter two types of non-SOZ regions experienced different changes in directed connections during the transition from interictal to ictal phases. Regions in the SOZ module (regardless of time window) all had substantial increases in the number of directed connections in ictal phases compared to interictal phases. For example, for Patient 1 (electrode placement is shown in Figure 2.3a), all the regions in the SOZ module (electrodes in red) had significant increases in the number of directed connections during 75-100 seconds after seizure onset (Figure 2.3e), compared to their numbers of directed connections in interictal windows.

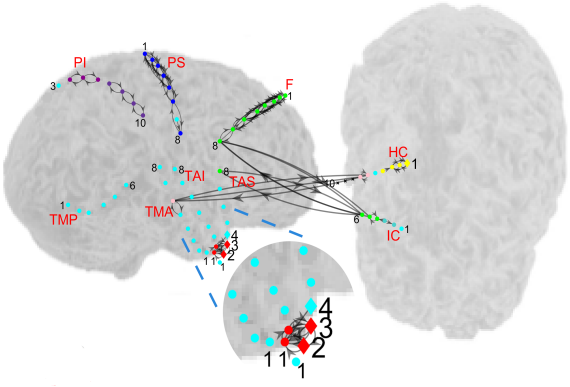
On the other hand, many regions outside the SOZ module did not exhibit increases in directed connections though these regions displayed ictal activity based on visual interpretation of EEG. For example, for Patient 2 (electrode placement is shown in Figure 2.4b), regions outside the SOZ module, such as TMA3-TMA5 and TMP3-TMP6, had no direct connections with the SOZ in both interictal and ictal phases. These regions had no changes in directed connections from interictal to ictal phases (Figures 2.4b-2.4e), despite that their intracranial EEG data signaled ictal activity during seizure propagation (Figure 2.4a). This result indicates that not all regions showing epileptic activity are actually involved in seizure propagation or directly influenced by the activity in the SOZ during seizure propagation.

2.4.8 Results: Directed Connectivity Analysis for SOZ Localization

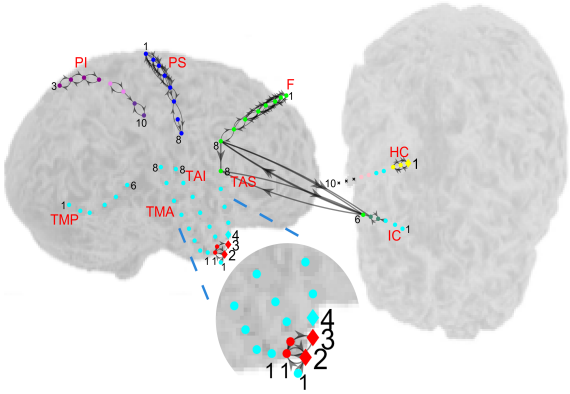
The above results show the identified changes with respect to the overall directed network structure from interictal to ictal phases. We next show the change in directed connectivity for individual regions. Figure 2.5b presents directed connectivity changes (DC_j^t defined in Section 2.4.4) of all individual regions for Patient 3 from 275 seconds before to 275 seconds



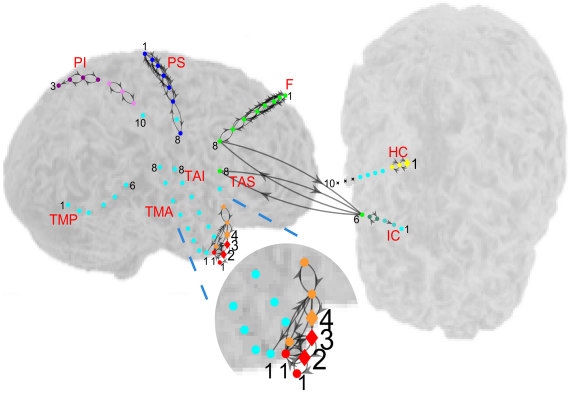
(a) Intracranial EEG of Patient 2



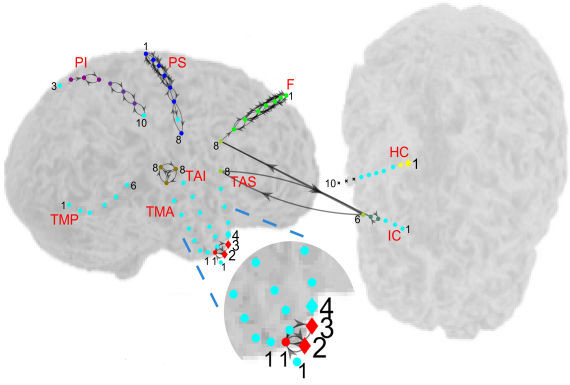
(b) 25-50 seconds before onset



(c) 0-25 seconds before onset

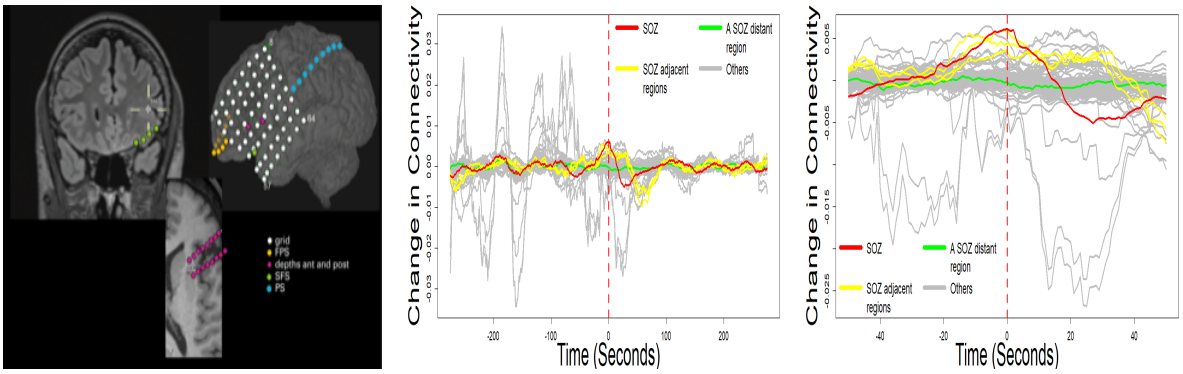


(d) 0-25 seconds after onset



(e) 25-50 seconds after onset

Figure 2.4: Patient 2's Directed Brain Networks Identified by the Proposed BSBM. (a) A segment of intracranial EEG recordings of Patient 2. (b)-(e) The identified directed brain networks from 50 seconds before to 50 seconds after seizure onset. The diamonds are the SOZ identified by expert interpretation of EEG data. Nodes in light blue are the regions that did not belong to any modules. Nodes in the same other colors denote different identified modules of regions. Arrows indicate the identified directed connections between regions. The evolution of the brain network is shown as a video, which can be downloaded via the link in Appendix A.4.



(a) Electrode Placement (b) Connectivity Changes $t \in (-275, 275)$ (c) Connectivity Changes $t \in (-50, 50)$

Figure 2.5: The Directed Connectivity Changes over Time of All Regions for Patient 3. (a) Electrode placement of Patient 3. (b) The time series of directed connectivity changes, DC_j^t, of all regions from 275 seconds before to 275 seconds after seizure onset. $t = 0$ is the seizure onset time. (c) The time series of directed connectivity changes of all regions from 50 seconds before to 50 seconds after seizure onset. The red curve is the directed connectivity changes of the SOZ.

after seizure onset, with Figure 2.5c focused on the shorter time interval around the seizure onset time (± 50 seconds around seizure onset). We found that at $t = 0$ when seizures started, the SOZ and its a few adjacent regions experienced the greatest directed connectivity changes. Based on this unique directed connectivity property of the SOZ, we developed a SOZ localization method (see Section 2.4.4). Using this method, we selected G30 and its three adjacent regions (G22-G24) to be candidate regions for the SOZ. Since G30 is the true SOZ, we regard this selection as accurately identifying the same region as verified by expert visual interpretation of EEG for Patient 3.

We assessed the effectiveness of our method in detecting the unique directed connectivity property of the SOZ by evaluating the accuracy of our method in localizing SOZs for all six patients. We used clinically localized SOZs as the given truth. We deemed our method to successfully detect one SOZ location (i.e., a true positive) if at least one of the regions selected by the method was within one electrode of the SOZ location. A false positive was scored when a region selected by the method was beyond one electrode of any SOZ location. Our localization method achieved 100% true positive rates (TPR) for all six patients (Table 2.3) despite their different SOZ locations and different numbers of regions inside the SOZ

areas. The false positive rates (FPR) were zero for three patients and 2-3% for the other three patients. Note that Patient 6 had multiple SOZ locations, which are difficult to accurately localize by using existing quantitative methods. In contrast, our method successfully detected all the SOZ locations for this patient.

2.5 Discussion

This chapter develops a new high-dimensional dynamic system model for mapping directed brain networks using intracranial EEG data. The proposed approach has three novelties. First, the proposed first-order MARSS for the brain network is effective for approximating various high-dimensional brain systems. The model is robust to violations of model assumptions. Second, in contrast to standard MARSS and MAR models, the proposed Bayesian framework incorporates the prior knowledge of the modular organization into the model estimation, which addresses the challenge of detecting connected brain regions among many possible ones. Our method produces scientifically meaningful network results. Third, we develop a stochastic blockmodel (SBM)-motivated prior to impose the modular organization on the MARSS parameters that denote directed edges. This is novel from standard SBMs for observed networks where network edges are directly known.

The proposed method can robustly detect directed connections with high accuracy, even if the underlying model for the brain network is nonlinear and highly distinct from the proposed MARSS for three reasons. First, we apply the MARSS to short intracranial EEG time segments so that the linear model can effectively approximate the underlying network system. Second, we use the proposed model to identify the directed connections by detecting the existence of temporal dependence among neuronal activities of regions rather than estimating the nonlinear interactions among regions. The first-order MARSS focuses only on the primary temporal dependence (rather than the exact order or nonlinearity of the dependence) among multivariate time series. Thus, the model is parsimonious in terms of the number of model parameters and enables efficient detection of directed connections among many regions. Third, the SBM-motivated prior can effectively capture potential brain net-

Table 2.3: SOZ Localization

Patient Number	Clinically Diagnosed SOZ	Other Resected Regions	Number of Analyzed Regions	Selected Regions	TPR	FPR
1	G37	N/A	50	G28, G38, G30, **G37**, G52	100%	2%
2	TAS2-4, HC1*	TAS1, TAS5	67	**TAS2, TAS5**	100%	0%
3	G30	N/A	95	**G30**, G23, G22, G24	100%	2%
4	G17	G16, G19	20	**G16, G17**	100%	0%
5	AS3-6, MS4-6	MM3-6	38	**MM3, MM4, MS5, AS5**	100%	0%
6	RTM2-4, ROD1-2, RTA1-3	N/A	128	RTP1, **RTM4**, RTM5, RTP2 RTM6, **ROD2**, RTM8, **RTA1**	100%	3%

The true positive rates (TPR) and false positive rates (FPR) of the proposed SOZ localization method in comparison to the SOZ locations determined by the clinical practice. The correctly identified SOZ regions are marked in bold. * For Patient 2, HC1 and TAS2 were spatially close to each other and deemed to be in one SOZ location.

work patterns. Using the SBM-motivated prior increases the efficiency in detecting directed connections. In summary, the proposed integration of a conventional MARSS and the modular organization yields robustness, flexibility, efficiency, and computational feasibility in modeling and estimating brain network systems.

Epilepsy is a directed network disorder (Englot et al., 2016; Kramer and Cash, 2012). Existing studies of epileptic networks have mainly focused on functional connectivity without directionality information (Khambhati et al., 2015; Stacey et al., 2019) or on low-dimensional directed networks of only a few brain regions around the SOZ (Korzeniewska et al., 2014). In our study, we used BSBM to characterize the SOZ, its adjacent regions, and many more distant non-SOZ regions as one integrated high-dimensional directed network system. With BSBM, we identified not only local changes of the directed brain network in the SOZ's neighboring area from interictal to ictal phases but also changes in the directed brain network at many non-SOZ regions that were spatially distant from the SOZ. Specifically, BSBM identified increases in the number of directed connections of the SOZ after seizure onset, along with the expansion of the SOZ module. These network results with respect to the SOZ were consistent with known properties of seizure propagation (Alarcón and Valentín, 2012; Englot et al., 2016). More importantly, our approach uncovered two types of regions that differed in their changes in directed connections during the transition from interictal to ictal phases: 1) Regions within SOZ modules demonstrated substantial increases in directed connections after seizure onset, while 2) many regions outside SOZ modules demonstrated no changes in directed connections during seizure propagation despite visual evidence of these regions' ictal activity. Earlier studies also suggested that connectivity properties of non-SOZ regions vary by their distances to the SOZ (Englot et al., 2015; Zaveri et al., 2009). However, these studies focused on functional connectivity. The high-dimensional directed network results by BSBM better displayed the difference between regions that showed visual ictal activity of "bystander" regions versus those involved in ictal propagation and thus, shed new light on the brain's normal and abnormal network mechanisms.

An important and unique feature of BSBM is to simultaneously identify modules and directed connections. Both identifying connections (Friston, 2011) and identifying modules (Sporns and Betzel, 2016; Sporns et al., 2007) in brain networks are of great interest

in computational neuroscience, as they are critical to understanding the brain mechanism. However, identifying modules and identifying connections are usually performed separately with different approaches, resulting in two errors in the ensuing estimated networks. The proposed BSBM fills this gap by providing a unified tool for simultaneously identifying modules and directed connections. As demonstrated in the results, BSBM, as an integration of the MARSS and the feature of the modular organization, is able to more accurately and robustly identify high-dimensional directed brain networks, compared to those approaches that identify connections only without the feature of the modular organization. More importantly, despite patients' heterogeneous electrode arrays and underlying anatomic seizure foci, BSBM identified consistent network properties, including increases in the number of directed connections and the expansion of SOZ modules after seizure onset. Future work will attempt to confirm our method's utility across larger samples.

We posit that the identification of modules as well as directed connections of regions better reveals the interactions between regions that underlay seizure propagation. For example, BSBM accurately reflected the clinical history of Patient 1 (electrode placement shown in Figure 2.3a), who tended to experience secondarily generalized seizures arising from a SOZ in the frontal lobe, as his SOZ module encompassed his entire EEG array (Figure 2.3e). In contrast, the other 5 patients, whose electrographic seizures tended to remain focal, had complex partial seizures without secondary generalization. Accordingly, these 5 patients' SOZ modules remained confined to focal electrodes (Figures 2.4d and 2.4e). Such results suggest an association between the clinical symptoms of patients and the size of their SOZ modules during seizure propagation. This association promises to be a key pursuit in future research, which will improve understanding of the relationship between patients' epileptic brain networks and their medical impairments.

The SOZ and adjacent regions had the highest directed connectivity changes at seizure onset. However, at other times (before seizure onset and several seconds after seizure onset), the directed connectivity changes of the SOZ did not differ from other regions (Figure 2.5b). Though existing network studies of interictal intracranial EEG data suggested that the SOZ had different connectivity properties as compared to other regions (Wilke et al., 2008) during interictal phases, these studies were generally based on special interictal periods containing

spikes or high-frequency oscillations (Korzeniewska et al., 2014). In contrast, our analysis was not based on these short interictal events that required selection using the appearance of epileptiform phenomena. Our result suggested that the directed connectivity properties of the SOZ on average were no different from other regions most of the time. Future work will evaluate directed connectivity during those short high-frequency interictal periods.

Although we believe BSBM provides unique insights into the pathophysiology of epileptic networks, the clinical importance of the new method lies in its ability–albeit preliminarily—to localize SOZs independently from the traditional interpretation of EEG for patients with focal epilepsy. Many methods were developed to localize the SOZ, usually by comparing regions' connectivity strengths (Korzeniewska et al., 2014; Van Mierlo et al., 2013). In contrast, the proposed SOZ localization method is a novel method of using the high-dimensional directed network results generated from BSBM, yielding great efficiency in detecting the unique directed connectivity property of the SOZ among many regions. More importantly, our SOZ localization method focuses on comparing regions' changes in directed connectivity at the seizure onset time. Thus, it is more sensitive to detect the network change due to seizure propagation (Englot et al., 2016; Fisher et al., 2005).

Although the directed connectivity change of the SOZ at seizure onset was consistently among the highest across all recorded regions for all six patients, selecting the region with the highest directed connectivity change at seizure onset alone does not always lead to accurate localization of the SOZ. This is because such SOZ localization depends on the accuracy of estimating the timing of seizure onset. Another potential problem is that a patient can have more than one region in the SOZ area. Our method addressed these limitations by examining directed connectivity changes both spatially and temporally for each region. In addition to selecting regions whose directed connectivity changes at the seizure onset time were among the top 10% (spatially) across all the regions, we also excluded the regions whose maximum directed connectivity change (temporally) in interictal phases was 10% larger than the region's directed connectivity change at the seizure onset time. The second step reduced false selections for SOZ localization. These two criteria for directed connectivity changes in the proposed SOZ-localization method ensured a high TPR and a low FPR for localizing the SOZ. Overall, our initial results show high concordance between the proposed method with

the clinical "gold standard" for patients with different epilepsy types and SOZ locations.

Despite the flexibility and robustness of BSBM in identifying directed connections, the effectiveness of the proposed SOZ-localization method relies on clinical limitations of electrode coverage of the SOZ. In addition, since the proposed MARSS contains many parameters for a high-dimensional directed network, its high accuracies in SOZ localization also rely on sufficient data information, including recordings of at least three seizures per patient. Subsequent work will focus on generalizing our findings with a larger sample of patients with a range of surgical outcomes to determine the accuracy and prognostic values.

While epilepsy surgery for the right patient can be transformative, epilepsy surgery remains underutilized because identification of the SOZ remains difficult and expensive, especially for patients without a clear clinical target (Jacobs et al., 2012; Rosenow and Lüders, 2001; Surges and Elger, 2013). Our statistical network analysis can be transformative in medical practice by providing an independent tool to localize the SOZ for patients with drug-resistant epilepsy, which can be used with clinical information to improve patient outcomes.

3.0 Study on Whole-Brain Directed Networks of Many Subjects Using FMRI Data

The majority of this chapter has been previously published in (Wang et al., 2023).

3.1 Introduction

The brain is a high-dimensional directed network system as it consists of many regions as network nodes that exert influence on each other. We refer to the directed influence exerted by one region on another as directed connectivity (also called directed functional connectivity or effective connectivity (Friston, 2011; Mill et al., 2017)). Identifying directed connections between all the regions and revealing the whole-brain directed network are essential to understanding the functional organization of the brain. However, it is both statistically and computationally challenging to produce brain network estimates that are scientifically meaningful because of the enormous numbers of potential directed connections and possible patterns of the directed network between many network nodes. To address this challenge, we propose a new directed network model that incorporates the principles of the functional organization of the brain.

The functional organization of the brain is governed by two principles: functional specialization and functional integration (Friston, 1994). The former indicates that different brain areas are specialized for different brain functions, while the latter suggests different brain areas interact with each other to process information and perform various functions. Enormous brain networks studies (Meunier et al., 2009; Park and Friston, 2013; Sporns and Betzel, 2016) have suggested that the modular organization (also called modularity) of networks gives rise to functional specialization and integration. Specifically, the brain network comprises modules of brain regions, whose connections with regions in the same module are stronger and denser than connections with regions in different modules. Brain regions in the same module tend to be specialized for the same or similar functions. Directed connections

within and between modules ensure integration among different functionally specialized brain areas. Because modular networks have been widely reported in the literature to reflect the brain's functional organization (Fodor, 1983; Sporns, 2013), we assume whole-brain directed networks to have a modular organization. The goal is to identify modules as well as directed connections in whole-brain directed networks using resting-state functional magnetic resonance imaging (fMRI) data of a large number of subjects. We use fMRI data because they provide non-invasive measurements of the activity of the entire human brain with a high spatial resolution (Lindquist, 2008).

We recognize multiple challenges in simultaneously identifying directed connections and modules in whole-brain directed networks based on fMRI data of a large number of subjects. First, it is difficult to find a "perfect" model that can accurately characterize the complex interactive relationship between many regions for many subjects due to the limited understanding of the brain's functional organization. Therefore, a model for the whole-brain directed network inevitably has a model error, that is, the deviation of the assumed model from the true network. Second, brain network structures vary across subjects (Mennes et al., 2010; Moussa et al., 2012). Third, fMRI data have a high degree of noise (Lindquist, 2008), bringing an additional difficulty to the network analysis. Fourth, the analysis of massive fMRI data and simultaneous identification of brain modules and directed connections for many subjects can be computationally intensive. Existing approaches address part of these challenges, as explained in detail below.

Most information theoretic measures, such as cross-correlations (Kramer et al., 2008; Schiff et al., 2005), cross-coherence (Schröder and Ombao, 2019), transfer entropy (Vicente et al., 2011), directed transinformation (Hinrichs et al., 2006), and directed information (Liu and Aviyente, 2012), and many others (Van Mierlo et al., 2013; Wilke et al., 2011), quantify pairwise connectivity between regions and cannot be directly used to identify modules. Popular models such as dynamic causal modeling (DCM, Frässle et al., 2018; Friston et al., 2003) and neural mass models (David and Friston, 2003) characterize directed connectivity but not modules. Methods such as independent component analysis (Calhoun and Adali, 2012; Mejia et al., 2020; van de Ven et al., 2004) and spectral clustering (Craddock et al., 2012) are effective in identifying modules or functional systems in the brain. However, because these

methods are based on functional connectivity (i.e., statistical associations between activity in different regions (Friston, 2011)), they cannot provide information about the direction of connectivity between regions or the existence of directed connectivity between modules. Overall, existing brain network studies identify modules (Sporns and Betzel, 2016; Sporns et al., 2007) and directed connections (Chiang et al., 2017; Friston, 2011; Kook et al., 2021) separately with different approaches, resulting in two different and hard-to-track errors in the estimated directed network. Despite the recent development of models (Li et al., 2021; Zhang et al., 2020, 2015, 2017) to characterize both directed connectivity and modules in the human brain, these models are for single-subject analysis, and the estimation of these models based on fMRI data of many subjects is computationally infeasible.

To address limitations in existing directed network analysis, we develop a new Bayesian model for whole-brain directed networks of many subjects. At the subject level, we use a multivariate autoregressive state-space (MARSS) model for fMRI data of each subject, because the MARSS has the properties of robustness and flexibility in approximating various network systems (see Chapter 2 or Li et al. (2021)). At the population level, we assign a mixed-membership stochastic blockmodel (MMSB) as a prior to all the subjects' MARSS parameters that denote directed connections. The use of the MMSB prior enables brain network estimates to have the modular organization. That is, connections between regions in the same modules are much denser than connections between regions in different modules. The use of the MMSB prior also allows for each region to be in different modules and to have different directed connections in different subjects' brain networks, and accommodates the variation of directed brain networks across subjects. Overall, the proposed Bayesian model provides a flexible and robust framework for combining fMRI data of many subjects to characterize brain networks in modular organizations. Thus, the Bayesian model enables us to address the first three challenges in directed network analysis of many subjects' fMRI data.

We address the computational challenge in analyzing fMRI data of many subjects by developing a variational Bayesian method to estimate the proposed Bayesian model. Through both simulation and real data analysis, we show that our new variational method is able to identify the whole-brain directed network with both computational efficiency and estimation

accuracy. As far as we know, this is the first method that can identify brain modules and directed connections simultaneously and reveal whole-brain directed networks for many subjects.

We applied our method to all four resting-state fMRI runs of all subjects (995 subjects) from the Human Connectome Project (HCP Van Essen et al., 2013). Specifically, we divided the entire resting-state fMRI data into two sets, each consisting of two fMRI runs collected on two separate days for each of the 995 subjects. We analyzed the two fMRI data sets independently. Modules identified by our method are consistent with known brain functional systems with different specialized functions, such as visual, default mode, auditory, cingulo-opercular task-control systems, and many others. Our method also identified directed connections between the somatosensory-motor and auditory modules and between the cingulo-opercular task control and salience modules. Moreover, we evaluated the reproducibility of our method by taking advantage of multiple fMRI runs for each subject. We showed that brain network results from independent analysis of two fMRI data sets are highly similar with overlap coefficients above 80%.

The rest of the chapter is organized as follows. In Section 3.2, we introduce the MARSS model for multiple resting-state fMRI runs of multiple subjects. We then propose a new Bayesian hierarchical model that uses the MMSB as a prior for MARSS parameters. In Section 3.3, we develop a variational Bayesian approach to estimate the new Bayesian model. In Section 3.4, we examine the robustness and effectiveness of the proposed method compared to existing network methods through simulation studies. Section 3.5 presents the analysis results of resting-state fMRI data of many subjects. Section 3.7 concludes with a discussion.

3.2 The Directed Brain Network Model

We propose a directed network model for fMRI data from L runs in d regions of S subjects. In the real data analysis, we used the functional atlas in the literature (Power et al., 2011) to divide the entire brain into $d = 264$ non-overlapping functional regions. These regions span the cerebral cortex, the cerebellum, and subcortical structures.

3.2.1 The Multivariate Autoregressive State-Space Model

Let $\boldsymbol{y}^{s,l}(t) = (y_1^{s,l}(t), \ldots, y_d^{s,l}(t))'$ be fMRI measurements in d brain regions (i.e., d network nodes of the whole-brain directed network) at time t from the lth fMRI run of subject s for $s = 1, \ldots, S$, $t = 1, \ldots, T$, and $l = 1, \ldots, L$. Each data point, $y_j^{s,l}(t)$, is an average of fMRI data of all voxels in region j at time t in the lth run for subject s. Each time series, $\{y_j^{s,l}(1), \ldots, y_j^{s,l}(T)\}$, is standardized to have mean zero and variance one. Let $\boldsymbol{x}^{s,l}(t) = (x_1^{s,l}(t), \ldots, x_d^{s,l}(t))'$ be the state functions of the d brain regions at time t in the lth fMRI run of subject s. The state function, $\boldsymbol{x}^{s,l}(t)$, represents the brain activity in d regions at time t in the lth fMRI run for subject s. We model directed connections between the d regions of each subject s using a multivariate autoregressive state-space model (MARSS):

$$y_i^{s,l}(t) = c_i^{s,l} \cdot x_i^{s,l}(t) + \epsilon_i^{s,l}(t), \;\; i = 1, \ldots, d, \;\; s = 1, \ldots, S, \;\; l = 1, \ldots, L, \tag{3.1}$$

$$x_i^{s,l}(t) = \sum_{j=1}^{d} \gamma_{ij}^{s} \cdot A_{ij}^{s,l} \cdot x_j^{s,l}(t-1) + \eta_i^{s,l}(t), \;\; t = 1, \ldots, T_l, \tag{3.2}$$

where c_i^s is an unknown parameter for standardizing activity of different regions; γ_{ij}^s is an indicator with 1 indicating the presence of the directed connection from region j to region i in the directed brain network of subject s and 0 for the absence; $A_{ij}^{s,l}$s are coefficients; and $\eta_i^s(t)$ and $\epsilon_i^s(t)$ are error terms with mean zero.

We use the first-order MARSS to model directed connectivity among many brain regions because it is robust to the model error and data error and also is parsimonious in terms of the number of free parameters for characterizing directed connectivity between many regions (see Chapter 2 or Li et al. (2021)).

We use indicators, γ_{ij}^ss, to distinguish nonzero directed connections from zero ones. Models (3.1) and (3.2) distinguish two connections in different directions between every pair of regions i and j by using two different indicators, γ_{ij}^s and γ_{ji}^s, to represent the two connections in two different directions between the two regions. For example, suppose only γ_{ij}^s is identified to be nonzero, and γ_{ji}^s is identified to be zero. We deem that a directed connection exists only from region j to region i in subject s's brain network and not otherwise.

Following standard practice in connectivity studies (Hayden et al., 2016; Sato et al., 2010), we fix $\gamma_{ii}^s = 0$ for $i = 1, \ldots, d, \;\; s = 1, \ldots, S$. We let indicators for directed connections,

γ_{ij}^s, be shared in common across different fMRI runs for each subject. This is because fMRI data in separate runs for each subject were collected under the same condition, and it is intuitive to assume that the subject's brain networks are identical in these runs. Moreover, this assumption enables combining data information across multiple fMRI runs to estimate directed networks more efficiently than otherwise.

Under the MARSS, (3.1) and (3.2), we focus on identifying nonzero γ_{ij}^ss for all pairs of regions i and j and for every subject s. That is, we identify directed connections by using the MARSS as a working model to detect the existence of temporal dependencies between activity of different regions. Detecting the existence of temporal dependencies is robust to the model error and data noise, as demonstrated in Chapter 2 and the simulation study (see Section 3.4). For mathematical simplicity and computational efficiency, we let $\eta_i^{s,l}(t) \overset{\text{i.i.d}}{\sim} \text{N}(0,1)$ and $\epsilon_i^{s,l}(t) \overset{\text{i.i.d}}{\sim} \text{N}(0,\tau_i^2)$.

3.2.2 Bayesian Hierarchical Model for Modular Networks

Given that the modular brain network is tied with functional specialization and integration of the brain (Newman, 2006; Sporns, 2010), we impose modularity on γ_{ij}^ss by using a mixed membership stochastic blockmodel (MMSB) (Airoldi et al., 2008; Durante and Dunson, 2014; Fienberg et al., 1985; Nowicki and Snijders, 2001) prior for γ_{ij}^ss. The details of the prior specification are given below.

Let K be the pre-specified number of modules. Let $\boldsymbol{m}_i^s = (m_{i1}^s, \ldots, m_{iK}^s)'$ label the module of region i in the directed brain network of subject s. Only one element of $\boldsymbol{m}_i^s$ equals 1 and the rest elements equal 0. For example, $m_{ik}^s = 1$ indicates that region i is in module k in the brain network of subject s. Let $B_{k_1k_2}$, $k_1, k_2 = 1, \ldots, K$, denote the prior probability of a nonzero directed connection from a region in module k_2 to another region in module k_1. Let $\mathbf{B}$ be a $K \times K$ matrix with entries $B_{k_1k_2}$ for $k_1, k_2 = 1, \ldots, K$.

3.2.2.1 Prior Specification for Modularity

The prior for whole-brain directed networks with modularity is a joint distribution for γ_{ij}^ss (indicators), $\boldsymbol{m}_i^s$s (module labels), and $\mathbf{B}$ (the probability matrix) as follows:

$$\gamma_{ij}^s | \boldsymbol{m}_i^s, \boldsymbol{m}_j^s, \mathbf{B} \overset{\text{ind}}{\sim} \text{Bernoulli}((\boldsymbol{m}_i^s)' \, \mathbf{B} \, \boldsymbol{m}_j^s), \; i, j = 1, \ldots, d; \tag{3.3}$$

$$\boldsymbol{m}_i^s \overset{\text{i.i.d}}{\sim} \text{Multinomial}(1; p_{i1}, \ldots, p_{iK}) \text{ and } (p_{i1}, \ldots, p_{iK}) \sim \text{Dirichlet}(\frac{1}{K}\mathbf{1}_K); \tag{3.4}$$

$$B_{kk} \overset{\text{i.i.d}}{\sim} \text{Uniform}(l_0, 1) \text{ and } B_{k_1 k_2} \overset{\text{i.i.d}}{\sim} \text{Uniform}(0, u_0), \; k_1, k_2 = 1, \ldots, K, \; k_1 \neq k_2, \tag{3.5}$$

where l_0 and u_0 are pre-specified constants between 0 and 1, and $\mathbf{1}_K$ is a K-dimensional vector with all entries equal to 1.

The distribution (3.3) specifies prior probabilities for nonzero directed connections between regions either in the same module (referred to as within-module directed connections) or in different modules (referred to as between-module directed connections) in the directed brain network of subject s. For example, if $m_{ik_1}^s = 1$ and $m_{jk_2}^s = 1$, the prior probability of the nonzero directed connection from region j to regions i equals $(\boldsymbol{m}_i^s)' \, \mathbf{B} \, \boldsymbol{m}_j^s = B_{k_1 k_2}$.

We let $l_0 = 0.9$ and $u_0 = 0.1$ to reflect the prior belief that within-module connections are dense while between-module connections are much sparser (Park and Friston, 2013). We make the difference between the lower bound, l_0, and the upper bound, u_0, large to facilitate module identification. The practice of module identification rests on the difference between the densities of within-module and between-module connections. The closer are the densities of within-module and between-module connections, the more difficult it is to identify modules correctly. We choose a high lower bound (i.e., $l_0 = 0.9$) for prior distributions of within-module connections to identify the most closely connected regions. More importantly, we found that if we lower the upper bound l_0 from 0.9 to 0.8, many modules would be merged together because a lower l_0 allows for regions with fewer connections to form one module. On the other hand, the upper bound $u_0 = 0.1$ is chosen because it is the upper bound threshold used by Power et al. (2011) to detect connections. Through both simulation and real data analysis, we found that the combination of $l_0 = 0.9$ and $u_0 = 0.1$ leads to the most accurate module identification: the regions identified to be in the same module have the same brain functions according to the functional atlas provided by Power et al. (2011).

The MMSB prior, (3.3)-(3.5), allows for each region to be in different modules and have different directed connections in different subjects' brain networks and thus accommodates the variation of brain networks across subjects. Under the MARSS, (3.1) and (3.2), with the MMSB prior (3.3)-(3.5) (BMMSB), our goal is to identify modules and directed connections by estimating the population-mean probabilities of region i in different modules, $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iK})$, posterior probabilities of $\boldsymbol{m}_i^s$s, and posterior probabilities of γ_{ij}^ss, for all regions $i, j = 1, \ldots d$ and subjects $s = 1, \ldots, S$.

3.3 Variational Bayesian Inference

The standard Bayesian approach that uses Markov chain Monte Carlo simulations is computationally infeasible to estimate the above Bayesian model for the massive fMRI data under study (the number of regions, d, is in hundreds, the number of subjects, S, is almost one hundred, and the number of time series points, T_l, is in thousands). We develop a variational Bayesian approach to estimate the above Bayesian model and address the computational challenge, as explained below.

We first estimate $\boldsymbol{x}^{s,l}(t)$ through a standard MARSS (Holmes et al., 2012) (where γ_{ij}^ss in (3.2) are all fixed at 1) instead of using a fully Bayesian Approach. State functions, $\boldsymbol{x}^{s,l}(t)$, are not of interest in our study, but their estimation through a fully Bayesian approach is computationally time-consuming. In addition, we found that estimated $\boldsymbol{x}^{s,l}(t)$ under the standard MARSS (Holmes et al., 2012) are similar to those under the fully Bayesian approach.

Let $\mathbf{A}^{s,l}$ be a $d \times d$ matrix whose (i,j)th entry is $A_{ij}^{s,l}$, $i, j = 1, \ldots, d$ and $l = 1, \ldots, L$, $X^{s,l} = \{\boldsymbol{x}^{s,l}(0), \ldots, \boldsymbol{x}^{s,l}(T_l)\}$, and $\mathbf{X} = \{X^{s,l},\ l = 1, \ldots, L\ s = 1, \ldots, S\}$. Let $\boldsymbol{\Theta}$ denote all unknown parameters,

$$\boldsymbol{\Theta} = \Big\{\gamma_{ij}^s,\ \mathbf{A}^{s,l},\ \boldsymbol{m}_i^s,\ \boldsymbol{p}_i,\ \mathbf{B},\ i, j = 1, \ldots, d,\ l = 1, \ldots, L\ s = 1, \ldots, S\Big\}.$$

. We treat $\mathbf{X}$ as given data, and the posterior of $\mathbf{\Theta}$, given $\mathbf{X}$, is

$$p(\mathbf{\Theta}|\mathbf{X}) \propto \prod_{s=1}^{S}\prod_{l=1}^{L}\left\{\prod_{t=1}^{T_l} p\Big(\boldsymbol{x}^{s,l}(t)\Big|\boldsymbol{x}^{s,l}(t-1),\mathbf{\Theta}\Big)\right\}\cdot p(\mathbf{\Theta}). \tag{3.6}$$

where $p\Big(\boldsymbol{x}^{s,l}(t)\Big|\boldsymbol{x}^{s,l}(t-1),\mathbf{\Theta}\Big)$ is derived using the state model (3.2). The prior distribution for the parameters γ_{ij}^s, $\boldsymbol{m}_i^s$, and $\mathbf{B}$ is the MMSB prior, (3.3), (3.4), and (3.5). We assign normal priors to $A_{ij}^{s,l}$,

$$A_{ij}^{s,l} \overset{\text{i.i.d}}{\sim} \mathrm{N}(0,\xi_0^2), \tag{3.7}$$

where ξ_0 is a pre-specified positive constant. Explicit formulas of the posterior distribution, $p(\mathbf{\Theta}|\mathbf{X})$, are provided in Appendix B.1.

We use a variational method to approximate the posterior distribution $p(\mathbf{\Theta}|\mathbf{X})$ in (3.6). Variational methods (Blei et al., 2017) have received enormous popularity in estimating graphical models and network models (Airoldi et al., 2008; Durante and Dunson, 2014; Fienberg et al., 1985; Nowicki and Snijders, 2001; Wainwright and Jordan, 2008). However, existing variational methods are mainly for observed networks whose network edges are known. We here address a more complicated problem: simultaneously identifying directed network edges (i.e., directed connections) and modules based on multivariate time series measurements of activity of many networks nodes. Our new variational method is based on a new factorized approximation to $p(\mathbf{\Theta}|\mathbf{X})$. The factorized distribution is given as follows:

$$\begin{aligned} q(\mathbf{\Theta}|\mathbb{V}) = &\prod_{s=1}^{S}\prod_{i,j=1,i\neq j}^{d} q_1(\mathrm{A}_{ij}^{s,1},\ldots,\mathrm{A}_{ij}^{s,L},\gamma_{ij}^s|\mathbf{\Phi}_{ij}^s)\cdot\prod_{s=1}^{S}\prod_{i=1}^{d} q_2(\boldsymbol{m}_i^s|\mathbf{\Phi}^{\boldsymbol{m}_i^s})\cdot\prod_{i=1}^{d} q_3(\boldsymbol{p}_i|\mathbf{\Phi}^{\boldsymbol{p}_i}) \\ &\cdot\prod_{k_1,k_2=1}^{K} q_4(\mathrm{B}_{k_1k_2}|\mathbf{\Phi}^{\mathrm{B}_{k_1k_2}}), \end{aligned} \tag{3.8}$$

where $\mathbb{V}=\{\mathbf{\Phi}_{ij}^s,\mathbf{\Phi}^{\boldsymbol{m}_i^s},\mathbf{\Phi}^{\boldsymbol{p}_i},\mathbf{\Phi}^{\mathrm{B}_{k_1k_2}},s=1,\ldots,S,\ i,j=1,\ldots,d,\ i\neq j,\ k_1,k_2=1,\ldots,K\}$ is the set of free variational parameters.

The variational distribution factors in the factorized distribution (3.8) are given below,

$$\begin{aligned} &q_1(\gamma_{ij}^s|\mathbf{\Phi}_{ij}^s) = \mathrm{Bernoulli}(\gamma_{ij}^s|\alpha_{ij}^s); \\ &q_1(\mathrm{A}_{ij}^{s,1},\ldots,\mathrm{A}_{ij}^{s,L}|\gamma_{ij}^s,\mathbf{\Phi}_{ij}^s) = \prod_{l=1}^{L} q_1(\mathrm{A}_{ij}^{s,l}|\gamma_{ij}^s,u_{ij}^{s,l},w_{ij}^{s,l}), \end{aligned}$$

$$\text{where } q_1(\mathrm{A}_{ij}^{s,l}|\gamma_{ij}^s, u_{ij}^{s,l}, w_{ij}^{s,l}) = \begin{cases} \text{Normal}(\mathrm{A}_{ij}^{s,l}|u_{ij}^{s,l}, w_{ij}^{s,l}) & \text{if } \gamma_{ij}^s = 1, \\ \text{Normal}(\mathrm{A}_{ij}^{s,l}|0, \xi_0^2) & \text{if } \gamma_{ij}^s = 0; \end{cases}$$

$$q_2(\boldsymbol{m}_i^s|\boldsymbol{\Phi}^{\boldsymbol{m}_i^s}) = \text{Multinomial}(\boldsymbol{m}_i^s|1, \boldsymbol{\Phi}^{\boldsymbol{m}_i^s});$$

$$q_3(\boldsymbol{p}_i|\boldsymbol{\Phi}^{\boldsymbol{p}_i}) = \text{Dirichlet}(\boldsymbol{p}_i|\boldsymbol{\Phi}^{\boldsymbol{p}_i});$$

$$q_4(\mathrm{B}_{k_1k_2}|\boldsymbol{\Phi}^{\mathrm{B}_{k_1k_2}}) = \begin{cases} \text{Beta}(\mathrm{B}_{k_1k_2}|\beta_{1,k_1}, \beta_{2,k_1}) \cdot 1_{\{l_0<\mathrm{B}_{k_1k_2}<1\}}(\mathrm{B}_{k_1k_2}) & \text{if } k_1 = k_2, \\ \text{Beta}(\mathrm{B}_{k_1k_2}|\beta_{1,k_1k_2}, \beta_{2,k_1k_2}) \cdot 1_{\{0<\mathrm{B}_{k_1k_2}<u_0\}}(\mathrm{B}_{k_1k_2}) & \text{if } k_1 \neq k_2, \end{cases}$$

where $\boldsymbol{\Phi}_{ij}^s = \{\alpha_{ij}^s, u_{ij}^{s,l}, w_{ij}^{s,l},\ l = 1, \ldots, L\}$, $\boldsymbol{\Phi}^{\boldsymbol{m}_i^s} = \{\Phi_1^{\boldsymbol{m}_i^s}, \ldots, \Phi_K^{\boldsymbol{m}_i^s}\}$, $\boldsymbol{\Phi}^{\boldsymbol{p}_i} = \{\Phi_1^{\boldsymbol{p}_i}, \ldots, \Phi_K^{\boldsymbol{p}_i}\}$, $\boldsymbol{\Phi}^{\mathrm{B}_{k_1k_2}} = \{\beta_{1,k_1}, \beta_{2,k_1}\}$ for $k_1 = k_2$, $\boldsymbol{\Phi}^{\mathrm{B}_{k_1k_2}} = \{\beta_{1,k_1k_2}, \beta_{2,k_1k_2}\}$ for $k_1 \neq k_2$, , and $1_{\aleph}(x)$ is an indicator function that equals 1 if x falls into the set $\aleph$ and 0 otherwise.

A crucial novelty of our variational Bayesian method is to let γ_{ij}^s and $A_{ij}^{s,l}$ be dependent on each other in our approximating distribution (3.8). Although using a fully factorized approximating distribution is more common in practice, it is not effective in approximating our target distribution, $p(\boldsymbol{\Theta}|\mathbf{X})$. A fully factorized approximating distribution is based on the mean field theory (Chaikin et al., 1995). The theory suggests that a joint distribution of many random variables that are dependent on each other can be effectively approximated by a product of independent distributions of these variables. However, the mean field approximation is usually effective when each random variable depends on many other variables and pairwise dependencies between variables are weak. In the posterior distribution (3.6), each $A_{ij}^{s,l}$ mostly depends on γ_{ij}^s, and a full factorization of the posterior distributions of $A_{ij}^{s,l}$ and γ_{ij}^s leads to a large bias. Therefore, we keep the dependence structure between $A_{ij}^{s,l}$ and γ_{ij}^s in the approximating distribution (3.8). A similar idea is implemented in structured variational inference (Hoffman and Blei, 2015).

We determine the values of $\mathbb{V}$ through iteratively minimizing the KL-divergence between the approximating distribution $q(\boldsymbol{\Theta}|\mathbb{V})$ and the posterior distribution $p(\boldsymbol{\Theta}|\mathbf{X})$,

$$\text{KL}\left(q(\boldsymbol{\Theta}|\mathbb{V})||p(\boldsymbol{\Theta}|\mathbf{X})\right) = -\,\text{E}_q\left(\log\frac{p(\boldsymbol{\Theta}|\mathbf{X})}{q(\boldsymbol{\Theta}|\mathbb{V})}\right).$$

To provide a flexible Bayesian model, we let $K = d$ and the initial values of the variational parameters for module labels, $\Phi_i^{\boldsymbol{m}_i^s} = 1$ and $\Phi_k^{\boldsymbol{m}_i^s} = 0$ for $k \neq i$, $i = 1, \ldots, d$, and $s = 1, \ldots, S$.

Algorithm 1 Pseudocode for the Variational Bayesian Method

Let $t = 0$ and set initial values $\mathbb{V}^0$

Let $\mathbb{V} = \mathbb{V}^0$.

while $t = 0$ or $KL^t - KL^{t-1} > 0.01 \times M_{KL}$ **do**

Let $t = t + 1$.

1. For $s = 1, \ldots, S$ and $i = 1, \ldots, d$:

   Update $\mathbf{\Phi}^{\boldsymbol{m}_i^s}$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.

2. For $s = 1, \ldots, S$ and $i, j = 1, \ldots, d$: :

   Update $\mathbf{\Phi}_{ij}^s$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.

3. For $i = 1, \ldots, d$:

   Update $\mathbf{\Phi}^{\boldsymbol{p}_i}$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.

4. For $k_1, k_2 = 1, \ldots, K$:

   Update $\mathbf{\Phi}^{B_{k_1 k_2}}$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.

5. Let $\mathbb{V}^t = \mathbb{V}$.

6. If $t = 1$:

   Let $M_{KL} = KL^t - KL^{t-1}$.

7. Else if $t > 1$ and $M_{KL} < KL^t - KL^{t-1}$:

   Let $M_{KL} = KL^t - KL^{t-1}$.

end while

The initial values of the other variational parameters and detailed steps in the iterative optimization algorithm for evaluating variational parameters are provided in Appendix B.2.

Algorithm 1 provides the pseudocode of the iterative optimization algorithm. Let KL^t denote the KL-divergence value calculated (up to an arbitrary additive constant) at the t iteration and $M_{KL} = \max\{KL^t - KL^{t-1},\ t = 1, ...\}$, where M_{KL} can be estimated based on the algorithm outputs in the first a few iterations.

We employ parallel computing (Kontoghiorghes, 2005; Rosenthal, 2000) to implement the above iterative algorithm. The use of parallel computing with a 16-core node can reduce the computation time by 90%. The analysis of two runs of fMRI data of 1000 subjects by our method takes no more than 20 hours.

3.3.1 Posterior Inference

Posterior inference of directed brain networks is equivalent to identifying directed connections and modules in these networks. In the following we elaborate the procedures to identifying modules and directed connections using the variational parameters output from the above variational Bayesian method.

3.3.1.1 Identification of Modules in Subject-Specific Brain Networks

Intuitively, given an appropriate number of modules K, one can use the variational parameters $\mathbf{\Phi}^{m_i^s}$ output from the variational Bayesian method to determine the module for region i in the directed brain network of subject s. However, we let $K = d$ instead of using a carefully chosen K. This is because, even though we can identify the correct number of modules, it is difficult to correctly specify initial module assignments for many regions under study with K much smaller than d. As pointed out by Blei et al. (2017), the KL-divergence, $\mathrm{KL}\left(q(\mathbf{\Theta}|\mathbb{V})||p(\mathbf{\Theta}|\mathbf{X})\right)$, is a nonconvex optimization function, and its optimization is sensitive to initial values. If K is assigned a value much smaller than d, many regions would be incorrectly assigned to the same module in the initial step, resulting in the algorithm being stuck at a local mode that can be far from the truth. In contrast, in our initialization with $K = d$, we let each region be in one unique module and separate from each other. This initialization lets the algorithm automatically group regions and find the right module for every region. We found that this approach is more reliable than using the initial values where many regions could be incorrectly grouped together. Moreover, this initialization avoids the issues of identifying the correct number of modules and rerunning the algorithm.

On the other hand, because $K = d$ is much larger than the true number of modules, bringing uncertainty in determining the module of each region i, the probabilities, $\mathbf{\Phi}^{m_i^s}$, of each region i in different modules are small. More importantly, allowing for each region to be in different modules in different subjects' networks in the Bayesian model can lead to an identifiability issue because the same module can be given different labels in different subjects' networks.

We propose the following computationally fast steps to determine an appropriate num-

ber of modules and reevaluate posterior probabilities of each region i in different modules. We first identify the regions that are in the same module inmost subjects' directed brain networks. We use these regions to determine modules and the number of modules, based on which, we reevaluate the probabilities of module assignments for the other regions. In the following, $\mathbf{\Phi}$ denotes the variational parameter output of the variational Bayesian method, and a notation $\hat{\theta}$ denotes a quantity evaluated based on the output:

1. Evaluate the probability of two regions, i and j, in the same module in the directed brain network of each subject s by $\hat{\Omega}_{ij}^s = \sum_{k=1}^d \Phi_k^{\boldsymbol{m}_i^s} \cdot \Phi_k^{\boldsymbol{m}_j^s}$.
2. Two regions i and j are deemed to be in the same module in the directed brain network of subject s if $\hat{\Omega}_{ij}^s > \frac{1}{d}$.
3. Identify sets of regions, $C_k,\ k = 1, \ldots, \hat{K}$, that satisfy three conditions: (1) Each C_k contains at least two regions; (2) for any two regions $i_{k_1}, i_{k_2} \in C_k$, either i_{k_1} and i_{k_2} are in the same module in more than 50% of subjects' directed brain networks or there exists a third region $j_k \in C_k$ such that i_{k_1} with j_k and j_k with i_{k_2} are in the same module in more than 50% of subjects' directed brain networks; (3) for any two regions in two different sets, $i \in C_k,\ j \in C_{\tilde{k}}$, and $k \neq \tilde{k}$, i and j are different regions, and i and j are in the same module in fewer than 50% of subjects' brain networks.
4. For all regions $i_k \in C_k$, let $\hat{m}_{i_k,k}^s = 1$ and $\hat{p}_{i_k,k} = 1$. That is, we deem all the regions in C_k to be in the same module k in directed brain networks of all subjects.

In Step 1, we calculate $\hat{\Omega}_{ij}^s$, based on the factorized distribution (3.8), in which the distributions of module labels for regions i and j are independent. In Step 2, the value $1/d$ is calculated based on the worst scenario where the probabilities of module labels of either region i or region j are identical for $K = d$ modules (i.e., $\Phi_k^{\boldsymbol{m}_i^s}$ or $\Phi_k^{\boldsymbol{m}_j^s} = 1/d$ for all $k = 1, \ldots, d$). Step 3 identifies groups of regions that are in the same module in most subjects' brain networks. Step 4 lets the $\hat{K}$ sets of regions, identified in Step 3, define $\hat{K}$ modules.

Given the $\hat{K}$ region sets, C_k, $k = 1, \ldots, \hat{K}$, we reevaluate the variational parameters of

module labels for each region $i \notin \{C_k, k = 1, \ldots, \hat{K}\}$ and subject s. Specifically, we let

$$\hat{\Phi}_k^{\boldsymbol{m}_i^s} = \sum_{h=1}^{d} \Phi_h^{\boldsymbol{m}_i^s} \cdot \max\{\Phi_h^{\boldsymbol{m}_{i_k}^s}, i_k \in C_k\} \quad \text{for} \quad k = 1, \ldots, \hat{K},$$

and $\hat{\Phi}_k^{\boldsymbol{m}_i^s} = 0$ for $k = \hat{K}+1, \ldots, d$. The above calculates the probability of region i in the same module as any one of the regions in C_k. Then we standardize $\hat{\Phi}_k^{\boldsymbol{m}_i^s}$, $k = 1, \ldots, \hat{K}$ such that their sum equals 1 for every region i and subject s.

We use $\hat{\boldsymbol{\Phi}}^{\boldsymbol{m}_i^s} = \{\hat{\Phi}_1^{\boldsymbol{m}_i^s}, \ldots, \hat{\Phi}_{\hat{K}}^{\boldsymbol{m}_i^s}\}$ to identify the module of region i in the directed brain network of subject s. If region i's largest module probability, $\hat{\Phi}_{k_{(1)}}^{\boldsymbol{m}_i^s}$, is larger than 50%, we deem that region i falls into module $k_{(1)}$ in the directed brain network of subject s; otherwise, region i does not fall into any module.

3.3.1.2 Identification of Modules in the Population-Mean Brain Network

Given modules identified in S subjects' directed brain networks, we reevaluate the population-mean probability of region i in module k, $\hat{p}_{ik}$, by the percentage of the S subjects' networks in which region i is in module k:

$$\hat{p}_{ik} = \frac{1}{S} \sum_{s=1}^{S} 1_{\hat{\Phi}_k^{\boldsymbol{m}_i^s} > 50\%}.$$

After normalizing $\hat{\boldsymbol{p}}_i = \{\hat{p}_{i1}, \ldots, \hat{p}_{i\hat{K}}\}$ to have a sum one, we use it to determine the module(s) of each region i in the population-mean directed brain network. The module assignment of each region i falls into 4 scenarios. (1) If the largest module probability of region i, $\hat{p}_{ik_{(1)}}$, is larger than 50%, we deem that region i falls into module $k_{(1)}$ only; (2) if $\hat{p}_{ik_{(1)}} \leq 50\%$ and $\hat{p}_{ik_{(1)}} + \hat{p}_{ik_{(2)}} > 50\%$, we deem that region i falls into modules $k_{(1)}$ and $k_{(2)}$; (3) if $\hat{p}_{ik_{(1)}} + \hat{p}_{ik_{(2)}} \leq 50\%$ and $\hat{p}_{ik_{(1)}} + \hat{p}_{ik_{(2)}} + \hat{p}_{ik_{(3)}} > 50\%$, we deem that region i falls into three modules, $k_{(1)}$, $k_{(2)}$, and $k_{(3)}$; (4) if $\hat{p}_{ik_{(1)}} + \hat{p}_{ik_{(2)}} + \hat{p}_{ik_{(3)}} \leq 50\%$, we deem that the modules of region i are unidentifiable in the population-mean brain network. We consider each region to be in no more than three different modules (corresponding to three different specialized functions) for easy scientific interpretation and to detect the most significant modules for each region. We also found that very few regions can fall into more than three different

modules.

3.3.2 The Choice of Hyperparameter

The hyperparameter ξ_0^2 can affect modules identified in each subject's network. Specifically, if ξ_0^2 is too small, the values of $A_{ij}^{s,l}$s would be tiny which will result in small differences between the posterior probabilities of including ($\gamma_{ij}^s = 1$) and excluding ($\gamma_{ij}^s = 0$) directed connections as well as small differences between the posterior probabilities of each region being in different modules. On the other hand, if ξ_0^2 is too large, $A_{ij}^{s,l}$s tend to be large, and indicators, γ_{ij}^ss, tend to be 0 regardless of regions' module assignments. The probabilities of each region being in different modules are also similar. Overall, either too large or too small ξ_0^2 makes it difficult to identify correct modules for each region.

Considering that modules identified affect the number of free parameters in the state model (3.2), we propose a Bayesian information criterion (BIC) to choose ξ_0^2.

For easy calculation of BIC, we treat all regions in the same module to be pairwisely connected and regions in different modules are disconnected. Given ξ_0^2, let $\mathcal{C}_{i,\xi_0^2}^s$ be the set of regions (excluding region i) in the same module as region i in the directed brain network of subject s. If region i does not fall into any modules in the directed brain network of subject s (i.e., $\hat{\Phi}_{k_{(1)}}^{\boldsymbol{m}_i^s} \leq 50\%$), $\mathcal{C}_{i,\xi_0^2}^s = \emptyset$. Given $\mathbf{X}$, let $\hat{L}_{i,\xi_0^2}^{s,l}$ denote the maximized value of the likelihood function of the state model (also a linear regression model), $x_i^{s,l}(t) = \sum_{j \in \mathcal{C}_{i,\xi_0^2}^s} A_{ij}^{s,l} \cdot x_j^{s,l}(t-1) + \eta_i^{s,l}(t)$ for $t = 1, \ldots, T_l$. Let $\kappa_{\xi_0^2}$ be the total number of free parameters in these $S \cdot d \cdot L$ regression models. Our BIC is

$$\text{BIC}(\xi_0^2) = \kappa_{\xi_0^2} \cdot \log(\sum_{l=1}^{L} S \cdot d \cdot T_l) - 2 \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{i=1}^{d} \log(\hat{L}_{i,\xi_0^2}^{s,l}).$$

We select ξ_0^2 that leads to the smallest $\text{BIC}(\xi_0^2)$ and and more than 90% of regions having identifiable modules.

Note that the above procedure allows us to analyze the massive fMRI data just once for each candidate hyperparameter ξ_0^2 and thus, requires much less computational time to determine the appropriate number of modules.

3.3.3 Directed Connection Identification

We use α_{ij}^s to identify directed connections in the subject-specific directed network for each subject s and use average posterior probabilities $\bar{\alpha}_{ij} = \sum_{s=1}^{S} \alpha_{ij}^s/S,\ i, j = 1, \ldots, d$ to identify directed connections in the population-mean directed network.

Because it is hard to know the density of true between-module connections versus within-module connections, we followed the approach by Power et al. (2011) and selected directed connections with top posterior probabilities ranging from top 1% to top 10%. We present directed connections with the highest possible posterior probabilities for easy visualization and minimal false selections while ensuring the number of selected between-module directed connections is no smaller than 1% of the number of selected within-module connections. The connections selected by this approach are easy to visualize and scientifically interpretable.

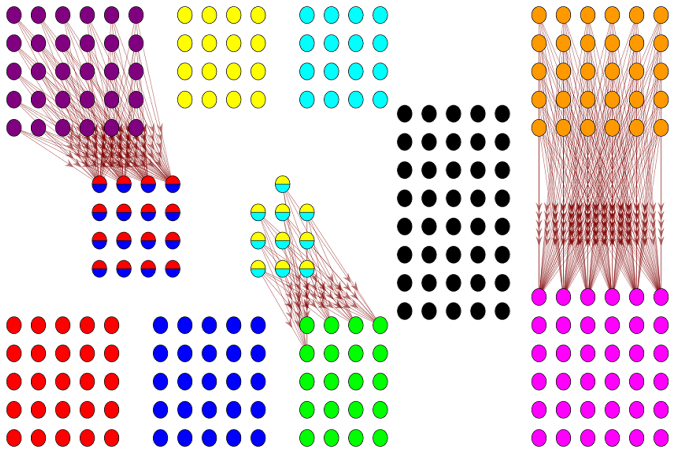
3.4 Simulation Studies

We used SPM software (Penny et al., 2011) to simulate fMRI data from the DCM (Friston et al., 2003) because it is the most popular model for directed connectivity. The DCM uses many complex ordinary differential equations (ODEs) in the state model to characterize interactions between neuronal activity in different regions and uses ODEs in the observation model to link regions' neuronal activity to their blood oxygen level dependent signals. We first used the ODEs in the state model of the DCM to generate state functions, $\boldsymbol{x}^{s,l}(t)$, of $d = 264$ egions in each of two ($l = 1, 2$) 15-minute runs for each subject s. The state functions $\boldsymbol{x}^{s,1}(t)$ and $\boldsymbol{x}^{s,2}(t)$ in two different runs were generated using the same ODEs but different initial values so that $\boldsymbol{x}^{s,1}(t) \neq \boldsymbol{x}^{s,2}(t)$ which is consistent with real data from different fMRI runs of each subject. Then, we used the ODEs in the observation model of the DCM to generate fMRI data $\boldsymbol{y}^{s,1}(t)$ in which the observation noise $\epsilon_j^{s,l}(t)$ of each region j is chosen such that the signal-to-noise ratio $\text{var}(x_j^{s,l}(t))/\text{var}(\epsilon_j^{s,l}(t)) = 1$ for $j = 1, \ldots, d = 264,\ , s = 1, \ldots, S = 1000$, and $l = 1, 2$. The chosen signal-to-noise ratio is considered low in the literature (Frässle et al., 2018). Note that simulation from the ODE model, DCM, generates

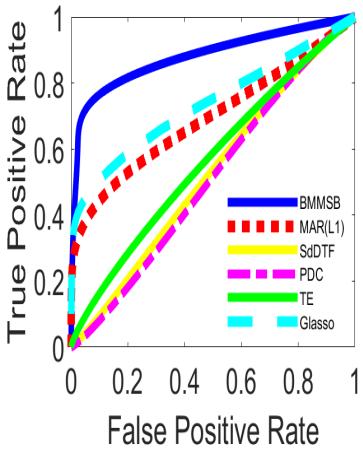
continuous data. We kept $T = 1200$ equally distanced data points with repetition time (TR) of 0.72 s as our simulated data, the same as the TR of real fMRI data under study.

Figure 3.1a shows simulated network patterns. We used the BrainNet Viewer (Xia et al., 2013) to visualize networks. The number of modules and the sizes of modules were chosen to be close to those of functional systems determined by Power et al. (2011). Network nodes in the same color are in the same module in all subjects' networks. Network nodes witt two colors are in one module (in one color) in 50% of subjects' networks and in the other module (in the other color) in the other 50% of subjects' networks. All network nodes in the same module are pairwise connected. We show only between-module connections in figures for easy visualization. Edges in dark red indicate between-module directed connections from an upper module to a lower module. Edges in green indicate between-module connections from a lower module to an upper module. The between-module connections are chosen to make easy visualization of the network. The number of between-module connections is around 5% of that of within-module connections.

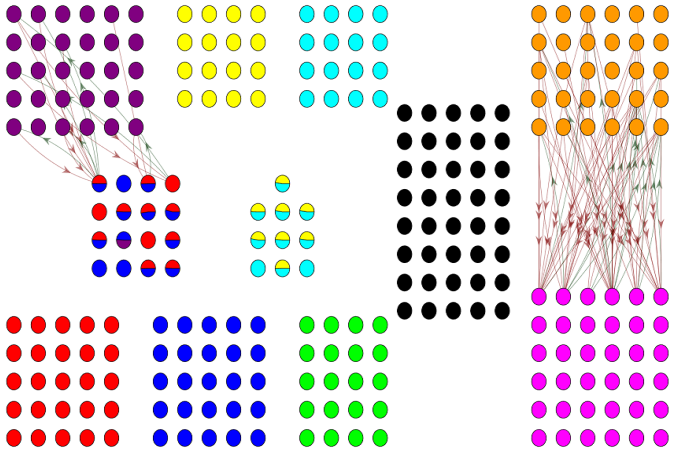
Using simulated directed connections (i.e., directed network edges) of all the subjects as the truth, we calculated the false positive rate (FPRs) and true positive rate (TPRs) of selecting directed network edges for all the subjects based on different thresholds for α_{ij}^ss. For comparison, we examined the FPRs and TPRs of popular competing methods, including the third-order MAR with L_1 regularization (implemented by the R package BigVAR (Nicholson et al., 2017)), denoted by MAR(L_1), transfer entropy (TE) (Sabesan et al., 2009; Schreiber, 2000; Vicente et al., 2011), partial directed coherence (PDC) (Baccalá and Sameshima, 2001), short-time direct transfer function (SdDTF) (Korzeniewska et al., 2014), and graphical lasso (Glasso) (Friedman et al., 2014; Witten et al., 2011). Figure 3.1b shows the ROC curves of TPRs vs. FPRs for these methods. We also tried the sparse regression DCM (Frässle et al., 2018), but it is computationally infeasible for identifying 1000 subjects' whole-brain directed networks. We also performed the simulation study 100 times independently and found that the accuracy of directed connection selection is stable across different simulations. The lowest value of the area under the curve (AUC) is 0.82, and the highest one is 0.89. In summary, the proposed variational Bayesian method with the MMSB prior (BMMSB) outperformed the other methods by achieving the largest area under the ROC curve.



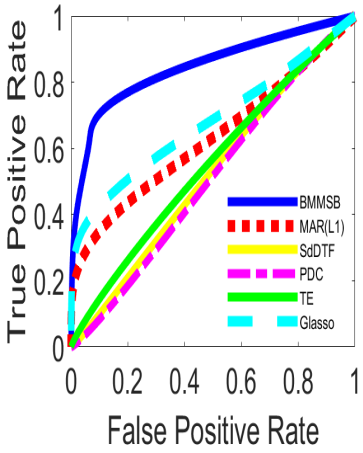
(a) Simulated Network



(b) ROC Curves $T = 1200$



(c) Estimated Mean Network for $T = 1200$



(d) ROC Curves $T = 600$

Figure 3.1: The Simulation Study of Data Generated from the DCM. (a) The simulated network patterns. Nodes in the same color are in the same module in all subjects' brain networks. Nodes with two colors are in different modules in different subjects' brain networks. Edges in dark red indicate between-module directed connections from an upper module to a lower module. Edges in green indicate between-module connections from a lower module to an upper module. (b) ROC curves for directed connections identified by six competing methods. (c) The estimated population-mean directed network. (d) ROC curves for directed connections identified by six competing methods based on data with $T = 600$ time points.

Figure 3.1c shows the estimated population-mean directed network. Our method successfully identified nine modules and the existence of two groups of regions with mixed module memberships. The TPR and FPR of selecting within-module directed connections are 66.3% and 0, respectively. The TPR and FPR of selecting between-module connections are 40.3% and 2.6%, respectively.

The TPR of selecting within-module connections is much higher than that of between-module connections for several reasons. First, module identification, similar to clustering, is subjective, so our selection of directed connections does not take into account identified modules and is purely based on posterior probabilities of directed connections (i.e., α_{ij}^ss). Since the number of true within-module connections is much larger than that of true between-module connections, and the number of candidate between-module connections is much greater than the total number of true directed connections, within-module connections are much easier to detect and their posterior probabilities tend to be much higher than those of between-module connections. Second, since the number of within-module connections is much larger than between-module connections, connection selection is more toward selecting within-module connections so that the overall accuracy of connection selection is high. Third, since the number of void connections is large, a slightly lower threshold for directed connections can lead to many selections. These selections not only could contain many false selections but also lead to a network result that is difficult to interpret scientifically. Consequently, we used a high threshold for α_{ij}^ss to avoid many false selections which also rendered only a few between-module connections selected. Overall, the proposed method outperformed existing methods by achieving a higher TPR and a low FPR.

We also analyzed the first half of the simulated fMRI data with $T = 600$ to assess the effect of the data length on the accuracy of connection selection. Figure 3.1d shows ROC curves of six competing methods. The proposed variational method has a slightly smaller AUC (0.85 compared to the AUC of 0.88 with $T = 1200$) in identifying directed connections with fewer data points and still outperformed other methods.

We performed another simulation study to compare the proposed variational Bayesian method and a fully Bayesian approach based on simulated fMRI data in $d = 62$ regions of a single subject. The ROC curve of the variational method is only slightly lower than that

of the fully Bayesian approach: The AUC of the former method is 0.82, and the AUC of the latter method is 0.87. This result suggests that the variational method can effectively approximate the target posterior distribution. More details of this simulation study can be found in Appendix B.3.

3.5 Real FMRI Data Analysis

We analyzed resting-state fMRI data of $S = 995$ healthy subjects in total from the Human Connectome Project (HCP) (Van Essen et al., 2013). All subjects went through one-hour (in total) resting-state fMRI scanning at 3T (Smith et al., 2013) in two pairs of 15-minute runs on each of two separate days. The data of each subject per run consist of functional images at 1200 time points with a repetition time (TR) of 0.72s and a 2-mm isotropic spatial resolution. The resting-state fMRI data downloaded from the HCP had been preprocessed according to the HCP minimal preprocessing pipeline. More detailed descriptions of the preprocessing steps, including optimized spatial preprocessing and temporal preprocessing, can be found in the papers by Glasser et al. (2013); Smith et al. (2013). Following the practice by Power et al. (2011), we extracted fMRI time series from the 10-mm diameter sphere of each of 264 regions of interest using the DPABI toolbox (Yan et al., 2016). We averaged fMRI time series of all voxels in each region j from each run l for each subject s and standardized the average time series to have mean zero and variance one. The ensuing time series was $\{y_j^{s,l}(1), \ldots, y_j^{s,l}(T_l)\}$ in our analysis.

We applied the proposed variational method to analyze subjects' fMRI data in $L = 2$ runs collected on separate days. Therefore, we analyzed two sets of fMRI data independently. The first set contains $S = 995$ subjects' resting-state fMRI data in the two runs with phase encoding in the left-to-right direction, and the second set contains the same subjects' resting-state fMRI data in the two runs with phase encoding in the right-to-left direction.

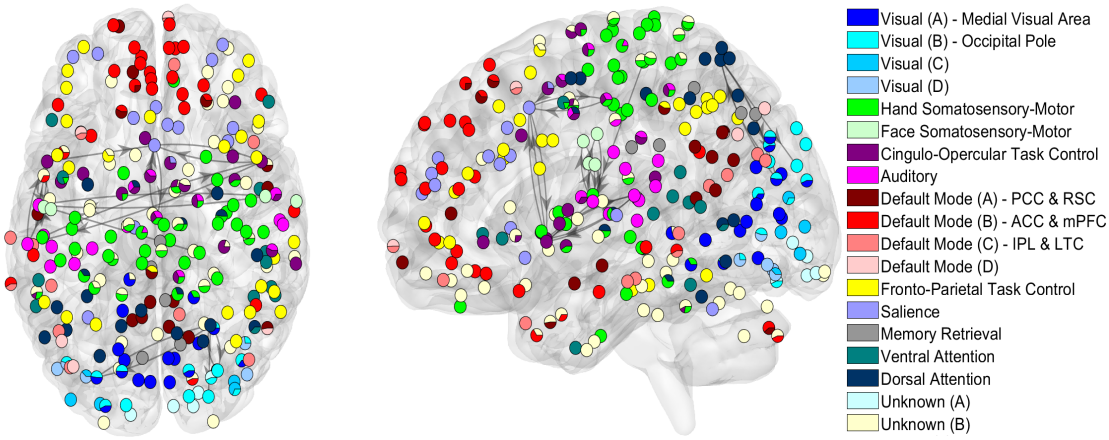
We present four major results of our directed network analysis of the fMRI data. First, modules identified by our method are accordant with functional brain systems specialized for various functions. The accordance between the identified modules and functional brain sys-

tems provides validation of module identification by our directed network method. Second, we revealed directed connections between brain modules with different specialized functions. These identified between-module directed connections are consistent with those discovered in low-dimensional directed network analysis of task-based fMRI data in just a few regions of interest. Third, we uncovered several regions that can be in different modules in different subjects' networks. This result suggests that these regions can be involved in more than one brain function. Fourth, we evaluated reproducibility by comparing the results of the independent analysis of the two fMRI data sets. We found both modules and directed connections identified are similar across different data sets. We elaborate on these results below.

3.5.1 Identification of Modules

Our method identified modules specialized for different functions, though the method did not use spatial information of regions. Figure 3.2 shows the identified population-mean whole-brain directed network in axial and sagittal views using the first fMRI data set. The identified modules are specialized for functions including visual (several blue colors), hand somatosensory-motor (green), face somatosensory-motor (light green), cingulo-opercular task control (patriarch), auditory (fuchsia), default mode (dark red, red, light red, and pink), fronto-parietal task control (yellow), salience (purple), memory retrieval (gray), ventral attention (blue green), and dorsal attention (navy) functions. These results are consistent with the functional brain systems reported in the literature (Power et al., 2011). Note that the modules with "unknown" labels correspond to several subsystems identified by Power et al. (2011) to have fewer than four regions. The functional identities of these subsystems are unknown in the literature. Our method not only successfully separated these regions from other modules but also identified them to share similar functions.

Note that the above modules with different specialized functions are also called networks in the literature, for example, the default model network, cingulo-opercular task control network, and salience network. To keep terminology consistent in this chapter, we use modules instead of networks.



(a) Axial View (b) Sagittal View

Figure 3.2: The Identified Population-Mean Whole-Brain Directed Networks in Axial (a) and Sagittal (b) Views based on the First FMRI Data Set. The nodes in the same color are identified to be in the same module. The nodes with more than one color are identified to be in more than one module. Black edges represent directed connections between modules that have distinct functions. The directed connections selected have top 1% posterior probabilities.

Our method revealed several smaller modules in large functional brain systems, such as the visual and the default mode functional systems. These results align with the literature that the visual system (Zeki et al., 1991) and the default mode system (Buckner et al., 2008) consist of several functionally and anatomically different brain areas. Moreover, the identified small visual modules overlap with several known subdivisions in the visual system, including medial visual area (visual module A), occipital pole (visual module B), and lateral visual areas (visual modules C and D) (Ikeda et al., 2022). Our method is also able to identify modules of posterior cingulate and retrosplenial cortices (PCC & RSC), anterior cingulate and medial prefrontal cortices (ACC & mPFC), inferior parietal lobe and lateral temporal cortex (IPL & LTC), and other regions in the default mode system (Davey et al., 2016; Raichle, 2015). The correspondence between identified modules with known functional brain systems and the high overlap between identified small modules in the large visual and default mode systems with known subdivisions of these two systems all provide evidence that our method can successfully detect subtle functional differences between subdivisions in a large functional system and reveal the hierarchical modular organization of the brain.

3.5.2 Identification of Directed Connections

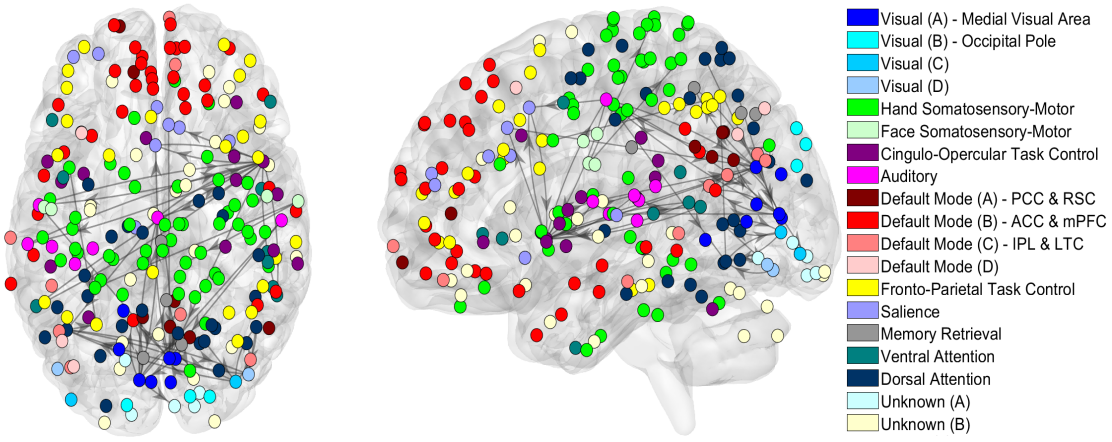
Most of the identified directed connections are between regions in the same module or between modules with similar brain functions (e.g., between the four visual modules). These connections are dense, as expected. For easy visualization of directed connections between different functionally specialized modules, we show only directed connections between modules with different specialized functions in Figure 3.2.

We discovered that the strongest between-module directed connections are between the auditory module and somatosensory-motor modules. Although existing studies have already reported strong functional connectivity between motor and auditory brain areas (De Luca et al., 2006; He et al., 2009; Mesulam, 1998), our results further suggest directed connections are between the face somatosensory-motor module and the auditory module. We also observed additional connections between the cingulo-opercular task control module and the salience module. This result is in accordance with the finding that the salience module engages the cingulo-opercular task-control regions (Seeley, 2019). In summary, our method can reliably detect directed connections between functionally specialized brain modules based on whole-brain resting-state fMRI data. In contrast, existing studies typically rely on tasked-based fMRI data to evaluate directed connections between only a few regions of interest with different specialized functions.

Another interesting finding, regarding directed connections between modules, is that the default mode module has no connection with other modules. This result is consistent with the abundant literature (Smith et al., 2009) that the default-mode network tends to be nonactive when the brain is during the performance of various goal-directed tasks (Gusnard and Raichle, 2001; Raichle et al., 2001).

3.5.3 Variation of Directed Brain Networks across Subjects

We examined the variation of directed brain networks across subjects. Figure 3.3 shows the whole-brain directed network of one subject. Identified modules in subject-specific directed brain networks are generally similar to those in the population-mean directed networks, although small modules in large functional brain systems, such as the default mode

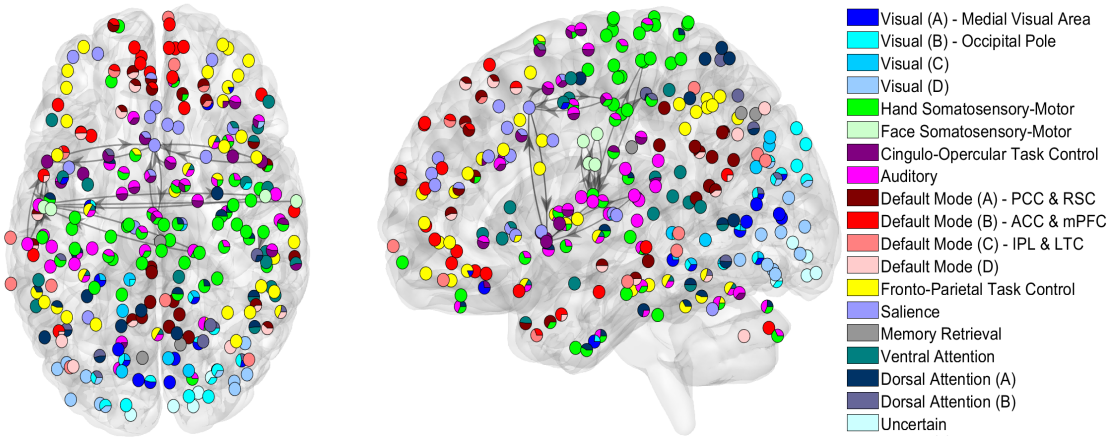


(a) Axial View (b) Sagittal View

Figure 3.3: The Identified Whole-Brain Directed Networks of One Subject in Axial (a) and Sagittal (b) Views based on the First FMRI Data Set. The nodes in the same color are identified to be in the same module. Black edges represent directed connections between modules that have distinct functions. The directed connections selected have top 1% posterior probabilities.

and somatosensory-motor modules, have moderate variations across subjects. We also found that regions in auditory, visual, somatosensory-motor, cingulo-opercular task control, and salience modules can fall into different modules in different subjects' networks, as demonstrated by nodes with more than one color in Figure 3.2. These results are consistent with the findings in the literature (Bushara et al., 2001; Deshpande et al., 2008; Power et al., 2011; Riedl et al., 2016; Seeley et al., 2007) that these modules have strong functional connectivity between them. Our results additionally suggest that regions in these modules can be involved in different brain functions.

The most considerable variation in directed brain networks across subjects lies in between-module directed connections. As shown in Figure 3.3, subject-specific directed brain networks have more between-module connections than the population-mean directed network.We consider several potential reasons for these results. First, the specialized functions of brain regions tend to be consistent across healthy subjects, while connectivity between regions vary dramatically across subjects during resting state. Second, fMRI data of each subject have a weak signal-to-noise ratio, leading to large variances of estimated subject-specific directed



(a) Axial View (b) Sagittal View

Figure 3.4: The Identified Population-Mean Whole-Brain Directed Networks in Axial (a) and Sagittal (b) Views based on the Second FMRI Data Set. The nodes in the same color are identified to be in the same module. The nodes with more than one color are identified to be in more than one module. Black edges represent directed connections between modules that have distinct functions. The directed connections selected have top 1% posterior probabilities.

brain networks. Third, estimating directed connectivity between many regions is susceptible to multicollinearity, while identifying modules, similar to clustering, is much less affected by multicollinearity. Therefore, identified functionally specialized modules tend to be stable across subjects, while identified connections between modules have much greater variations across subjects.

3.6 Reproducibility

We applied the variational Bayesian method to the same subjects' second resting-state fMRI data set and obtained the second estimated population-mean directed brain network shown in Figure 3.4. The network is similar to the first population-mean brain network (shown in Figure 3.2) obtained by analyzing the same subjects' first fMRI data set.

We calculated overlap coefficients of identified modules in the two networks to assess the

reproducibility of our method. The overlap coefficient is defined as

$$\text{overlap}(S_1, S_2) = \frac{|S_1 \cap S_2|}{\min(|S_1|, |S_2|)},$$

where S_1 and S_2 are two sets, for example, modules of regions. Let $\mathbb{S}_1$ and $\mathbb{S}_2$ be the collection of all the modules identified in the first and second population-mean directed brain networks, respectively. For each module $S_2 \in \mathbb{S}_2$, its overlap coefficient with $\mathbb{S}_1$ is defined as $\max_{S_1 \in \mathbb{S}_1} \text{overlap}(S_1, S_2)$. Similarly, we define the overlap coefficient of each module $S_1 \in \mathbb{S}_1$ with $\mathbb{S}_2$ as $\max_{S_2 \in \mathbb{S}_2} \text{overlap}(S_1, S_2)$. The mean of the overlap coefficients of modules in S2 with S1 is 80%, and the mean of the overlap coefficients of modules in $\mathbb{S}_1$ with $\mathbb{S}_2$ is 82%. The overlap coefficient of identified directed connections in the two population-mean networks is 92

We also examined the similarity between two estimated whole-brain directed networks for each subject. The average overlap coefficient of identified modules in subject-specific brain networks is 81%, and the average overlap coefficient of identified directed connections is 76%. Again, directed connections have more variations than modules across runs for reasons given above.

3.7 Discussion

We propose a new high-dimensional directed network method for analyzing resting-state fMRI data of many subjects. The advantages of our new method lie in three aspects. First, our model building exploits the principles of the brain's functional organization by characterizing both modules and directed connections in brain networks. Second, the new Bayesian model accommodates the variation of brain networks across subjects while enabling integration of many subjects' data to estimate whole-brain directed networks. Third, the developed new variational Bayesian method can simultaneously identify modules and directed connections with both computational efficiency and estimation accuracy.

Setting the lower bound, l_0, for prior probabilities of within-module connections at a high value of 0.9 is necessary for several reasons. First, it is documented in the literature that

regions in the same subnetwork (called modules in our analysis) are coactive (Cole et al., 2010). This coactivation leads to very strong correlations (at values of almost 1) between these regions' fMRI data. Second, fMRI preprocessing steps can increase correlations of fMRI data in different regions (Gargouri et al., 2018). Third, the large number of regions' fMRI data under study brings the multicollinearity issue when using a model to identify connections. Then, setting a high value for l_0 can enable us to reduce the false selections due to the high correlations caused by the second and third issues and identify truly strongly connected regions. Fourth, we found that using a smaller value of l_0 can render regions specialized for different functions incorrectly merged together because of the second and third issues. Fifth, our choice of l_0 has been implemented in Chapter 2.

We used the first-order MARSS, instead of higher-order ones, to identify directed connections for several reasons. First, the purpose of this study is to identify directed connections by detecting the existence of temporal dependence between regions' temporal activities rather than explaining fMRI data variation, fitting the data perfectly, or examining the extent of temporal dependence between regional activity. The first-order MARSS is efficient in capturing the presence of temporal dependence. Second, though a high-order MARSS may fit the data better, it contains many more free parameters. Estimating these more parameters brings significantly more variances and uncertainty in identifying directed connections. Third, simulations performed in Section 2.3 have demonstrated that the first-order MARSS can detect directed connections with high accuracy for data generated from high-order MARSS. We did similar simulations and obtained the same results. However, since the DCM is more distinct from the MARSS and, arguably, a generative model for fMRI, we presented simulation results based on the DCM. On the other hand, since our method is focused on detecting temporal dependence using a parsimonious model, the method does not differentiate between negative inhibitory relationships and positive excitatory relationships between regions. This analysis requires using more detailed models.

Evaluation of directed connections between functionally distinct areas is mainly through low-dimensional directed network analysis of task-based fMRI data in only a few regions of interest. Thus, these directed connectivity results are restricted to fMRI studies with specifically designed tasks. In contrast, our method can reliably detect directed connections

between modules with different functions based on whole-brain resting-state fMRI data. Our network results enhance our understanding of the brain's functional organization.

In future research we will extend our method to model dynamic connectivity by allowing indicators for directed connectivity to vary over time or assuming transition probabilities for directed connectivity. We will also develop the model for task-based fMRI data, compare resting-state and task-based whole-brain directed networks, and further investigate the variation of directed brain networks across different tasks and conditions.

4.0 Future Research

Due to recent technological advances, brain data collected from large populations at different time stages in various modalities are available. For example, the Adolescent Brain Cognitive Development (ABCD) study (Casey et al., 2018) follows more than $10,000$ children aged $9-10$ years old and tracks their biological and behavioral development through adolescence into young adulthood. The ABCD study collects these children's multimodal neuroimaging data (Hagler Jr et al., 2019). With our experience in analyzing massive neuroimaging data, we aim to develop new statistical models for these new data to address new statistical and scientific questions.

In the future, we will develop statistical models to infer the relationships between different brain networks by utilizing these new datasets. "Inferring the relationships between different brain networks" falls into several real scenarios. First, we will develop new methods to infer the relationships between brain networks of different connectivity types estimated from multimodal brain data. For example, the relationships between brain networks of structural connectivity inferred from diffusion tensor imaging (DTI) data and brain networks of functional connectivity estimated from fMRI data, or the relationships between brain networks estimated from structural MRI data and positron emission tomography (PET) data. Understanding the relationship between brain networks estimated from multimodal brain data will facilitate borrowing information across different data types (Zhu et al., 2017). Second, we will infer the relationships between the brain networks of different human populations. For example, the young adults in HCP and the adolescents in the ABCD study. Third, we will model the longitudinal brain networks estimated from longitudinal brain data. For the ABCD data, we aim to reveal normal and abnormal brain network changes and identify biological, social, and environmental factors that affect the development of healthy brain functions. For aging data (Bookheimer et al., 2019; Mueller et al., 2005), we aim to reveal normal brain network changes due to aging and abnormal brain network changes due to Alzheimer's disease progression.

Appendix A Appendix of Chapter 2

A.1 Joint Distribution

All the parameters to be estimated in the proposed Bayesian framework are

$$\boldsymbol{\Theta} = \{\boldsymbol{\Gamma}, \mathbf{B}, \mathbf{M}, \mathbf{A}, \boldsymbol{c}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{p}\}.$$

Let $\mathbf{X} = \{\boldsymbol{x}(0), \ldots, \boldsymbol{x}(T)\}$ and $\mathbf{Y} = \{\boldsymbol{y}(1), \ldots, \boldsymbol{x}(T)\}$. The joint posterior distribution of $\mathbf{X}$ and $\boldsymbol{\Theta}$ is:

$$\begin{aligned}
&p(\mathbf{X}, \boldsymbol{\Theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Theta}) \cdot p(\mathbf{X}|\boldsymbol{\Theta}) \cdot p(\boldsymbol{\Theta}) \\
\propto &\prod_{t=1}^{T} \frac{1}{|\mathbf{R}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{y}(t) - \mathbf{C}\boldsymbol{x}(t)\right]' \mathbf{R}^{-1} \left[\boldsymbol{y}(t) - \mathbf{C}\boldsymbol{x}(t)\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left[\boldsymbol{x}(0) - \boldsymbol{\mu}\right]' \left[\boldsymbol{x}(0) - \boldsymbol{\mu}\right]\right\} \\
&\cdot \prod_{t=1}^{T} \exp\left\{-\frac{1}{2}\left[\boldsymbol{x}(t) - (\boldsymbol{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)\right]' \left[\boldsymbol{x}(t) - (\boldsymbol{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)\right]\right\} \\
&\cdot \prod_{i,j=1}^{d} (\boldsymbol{m}_i' \boldsymbol{B} \boldsymbol{m}_j)^{\gamma_{ij}} (1 - \boldsymbol{m}_i' \boldsymbol{B} \boldsymbol{m}_j)^{1-\gamma_{ij}} \cdot \prod_{k_1,k_2=1, k_1 \neq k_2}^{K} 1\{0 < B_{k_1 k_2} < u_0\} \prod_{k=1}^{K} 1\{l_0 < B_{kk} < 1\} \\
&\cdot \prod_{i=1}^{d} \prod_{k=1}^{K} p_k^{m_{ik}} \cdot \prod_{k=1}^{K} 1\{p_k < 1\} \cdot 1\{\sum_{k=1}^{K} p_k = 1\} \\
&\cdot \prod_{i=1}^{d} \phi(\frac{c_i}{\xi_1}) \cdot \phi(\frac{\mu_i}{\xi_1}) \cdot \frac{1}{\tau_i^{1+\rho_0}} \exp\{-\frac{\rho_0}{\tau_i}\} \cdot \prod_{i,j=1}^{d} \phi(\frac{A_{ij}}{\xi_0}),
\end{aligned}$$

where $\circ$ denotes the Hadamard product, $\mathbf{C}$ is a $d \times d$ diagonal matrix with diagonal elements as $\boldsymbol{c}$, and $\mathbf{R}$ is a $d \times d$ diagonal matrix with diagonal entries as $\boldsymbol{\tau}$.

A.2 Partially Collapsed Gibbs Sampler

In each MCMC step, the parameters are simulated in the same order as their posterior conditional distributions presented below.

Unless specified, all the $\mathbf{I}$s in the following refer to identity matrices.

A.2.1 Update $\mathbf{\Gamma}$

We first integrate out $\mathbf{A}$ and simulate $\mathbf{\Gamma}$ from its ensuing posterior conditional distribution. Denote the ith row of $\mathbf{\Gamma}$ by $\mathbf{\Gamma}_i$ and define

$$\mathbf{U}^i = \sum_{t=1}^{T} [\mathbf{\Gamma}_i^{'} \circ \boldsymbol{x}(t-1)][\mathbf{\Gamma}_i \circ \boldsymbol{x}(t-1)^{'}] + \frac{1}{\xi_0^2}\mathbf{I} \text{ and } \mathbf{V}^i = \sum_{t=1}^{T} x_i(t) \cdot (\mathbf{\Gamma}_i \circ \boldsymbol{x}(t-1)^{'}).$$

Let $\mathbf{\Theta}_{-\mathrm{A}}$ represent all parameters excluding $\mathbf{A}$. The joint posterior probability $p(\mathbf{X}, \mathbf{\Theta}|\mathbf{Y})$ can be formulated as

$$f(\mathbf{Y}, \mathbf{X}, \mathbf{\Theta}_{-\mathrm{A}}) \cdot \prod_{t=1}^{T} \exp\Big\{ -\frac{1}{2}[\boldsymbol{x}(t) - (\mathbf{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)]^{'}[\boldsymbol{x}(t) - (\mathbf{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)]\Big\} \cdot \prod_{i,j=1}^{d} \phi(\frac{A_{ij}}{\xi_0}),$$

where $f(\mathbf{Y}, \mathbf{X}, \mathbf{\Theta}_{-\mathrm{A}})$ is the part of $p(\mathbf{X}, \mathbf{\Theta}|\mathbf{Y})$ without $\mathbf{A}$.

Denoting the ith row of $\mathbf{A}$ by $\mathbf{A}_i$, we have

$$\begin{aligned}
&\int f(\mathbf{Y}, \mathbf{X}, \mathbf{\Theta}_{-\mathrm{A}}) \cdot \prod_{t=1}^{T} \exp\Big\{ -\frac{1}{2}[\boldsymbol{x}(t) - (\mathbf{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)]^{'}[\boldsymbol{x}(t) - (\mathbf{\Gamma} \circ \mathbf{A})\boldsymbol{x}(t-1)]\Big\} \\
&\cdot \prod_{i,j=1}^{d} \phi(\frac{A_{ij}}{\xi_0}) d\mathbf{A} \\
\propto\ & f(\mathbf{Y}, \mathbf{X}, \mathbf{\Theta}_{-\mathrm{A}}) \cdot \int \prod_{i=1}^{d} \frac{1}{|(\mathbf{U}^i)^{-1}|^{\frac{1}{2}}} \cdot \exp\Big\{ -\frac{1}{2}[\mathbf{A}_i^{'} - (\mathbf{U}^i)^{-1}(\mathbf{V}^i)^{'}]^{'}\mathbf{U}^i[\mathbf{A}_i^{'} - (\mathbf{U}^i)^{-1}(\mathbf{V}^i)^{'}]\Big\} \\
&\cdot \prod_{i=1}^{d} |(\mathbf{U}^i)^{-1}|^{\frac{1}{2}} \cdot \prod_{i=1}^{d} \exp\Big\{ \frac{1}{2}(\mathbf{V}^i)(\mathbf{U}^i)^{-1}(\mathbf{V}^i)^{'}\Big\} d\mathbf{A} \\
\propto\ & f(\mathbf{Y}, \mathbf{X}, \mathbf{\Theta}_{-\mathrm{A}}) \cdot \prod_{i=1}^{d} |(\mathbf{U}^i)^{-1}|^{\frac{1}{2}} \cdot \prod_{i=1}^{d} \exp\Big\{\frac{1}{2}(\mathbf{V}^i)(\mathbf{U}^i)^{-1}(\mathbf{V}^i)^{'}\Big\}.
\end{aligned}$$

We update one γ_{ij} a time for $i, j = 1, \ldots, d$. At the step for updating γ_{ij}, let $\mathbf{\Gamma}_{-ij} = \mathbf{\Gamma} \backslash \{\gamma_{ij}\}$, the values of all γ_{ij}s excluding γ_{ij}. Let $\mathbf{\Gamma}_i^{(ij)=1}$ and $\mathbf{\Gamma}_i^{(ij)=0}$ be the vectors the same as $\mathbf{\Gamma}_i$ except that the jth entry in the two vectors equals 1 and 0, respectively.

Replace the vector $\mathbf{\Gamma}_i$ in the matrices $\mathbf{U}^i$ and $\mathbf{V}^i$ with $\mathbf{\Gamma}_i^{(ij)=1}$ or $\mathbf{\Gamma}_i^{(ij)=0}$, and define the ensuing matrices to be $\mathbf{U}_1^i$, $\mathbf{V}_1^i$, $\mathbf{U}_0^i$, and $\mathbf{V}_0^i$, respectively.

After calculating $\log w_{ij}^1 = \frac{1}{2}\log|(\mathbf{U}_1^i)^{-1}| + \frac{1}{2}(\mathbf{V}_1^i)(\mathbf{U}_1^i)^{-1}(\mathbf{V}_1^i)^{'} + \log(\boldsymbol{m}_i^{'}\mathbf{B}\boldsymbol{m}_j)$, and $\log w_{ij}^0 =$

$\frac{1}{2}\log|(\mathbf{U}_0^i)^{-1}|+\frac{1}{2}(\mathbf{V}_0^i)(\mathbf{U}_0^i)^{-1}(\mathbf{V}_0^i)' + \log(1-\boldsymbol{m}_i'\mathbf{B}\boldsymbol{m}_j)$, we sample γ_{ij} from

$$p(\gamma_{ij}=1|\boldsymbol{\Gamma}_{-ij},\mathbf{X},\mathbf{B},\boldsymbol{m}) = \left[1+\exp\left\{\log w_{ij}^0 - \log w_{ij}^1\right\}\right]^{-1}.$$

A.2.2 Update A

We update one $\mathbf{A}_i$ a time for $i=1,\ldots,d$. Sample $\mathbf{A}_i$ for $i=1,\ldots,d$ from

$$\mathbf{A}_i|\mathbf{X},\boldsymbol{\Gamma}_i \sim \text{MVN}((\mathbf{U}^i)^{-1}(\mathbf{V}^i)', (\mathbf{U}^i)^{-1}).$$

A.2.3 Update X

We update $\boldsymbol{x}(0)$, $\boldsymbol{x}(1)$, $\ldots$, $\boldsymbol{x}(T)$ one by one.

- For $\boldsymbol{x}(0)$: $p(\boldsymbol{x}(0)|\boldsymbol{x}(1),\boldsymbol{\Gamma},\mathbf{A},\boldsymbol{\mu})$ is proportional to

$$\begin{aligned}
&\exp\Big\{-\frac{1}{2}[\boldsymbol{x}(0)-\boldsymbol{\mu}]'[\boldsymbol{x}(0)-\boldsymbol{\mu}]\Big\}\cdot\exp\Big\{-\frac{1}{2}[\boldsymbol{x}(1)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(0)]'[\boldsymbol{x}(1)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(0)]\Big\}\\
\propto\ &\exp\left\{-\frac{1}{2}\Big[\boldsymbol{x}(0)'((\boldsymbol{\Gamma}\circ\mathbf{A})'(\boldsymbol{\Gamma}\circ\mathbf{A})+\mathbf{I})\boldsymbol{x}(0) - 2(\boldsymbol{x}(1)'(\boldsymbol{\Gamma}\circ\mathbf{A})+\mu')\boldsymbol{x}(0)\Big]\right\}.
\end{aligned}$$

Let $\mathbf{U}_0 = (\boldsymbol{\Gamma}\circ\mathbf{A})'(\boldsymbol{\Gamma}\circ\mathbf{A})+\mathbf{I}$, and $\mathbf{V}_0 = \boldsymbol{x}(1)'(\boldsymbol{\Gamma}\circ\mathbf{A})+\boldsymbol{\mu}'$. We sample $\boldsymbol{x}(0)$ from

$$\boldsymbol{x}(0)|\boldsymbol{x}(1),\boldsymbol{\Gamma},\mathbf{A},\boldsymbol{\mu} \sim \text{MVN}((\mathbf{U}_0)^{-1}(\mathbf{V}_0)', (\mathbf{U}_0)^{-1}).$$

- For $\boldsymbol{x}(t)$, $t=1,\ldots,T-1$: $p(\boldsymbol{x}(t)|\boldsymbol{x}(t-1),\boldsymbol{x}(t+1),\boldsymbol{y}(t),\boldsymbol{\Gamma},\mathbf{A},\mathbf{C},\mathbf{R})$ is proportional to

$$\begin{aligned}
&\exp\Big\{-\frac{1}{2}[\boldsymbol{x}(t)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(t-1)]'[\boldsymbol{x}(t)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(t-1)]\Big\}\\
&\cdot\exp\Big\{-\frac{1}{2}[\boldsymbol{x}(t+1)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(t)]'[\boldsymbol{x}(t+1)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(t)]\Big\}\\
&\cdot\exp\Big\{-\frac{1}{2}[\boldsymbol{y}(t)-\mathbf{C}\boldsymbol{x}(t)]'\mathbf{R}^{-1}[\boldsymbol{y}(t)-\mathbf{C}\boldsymbol{x}(t)]\Big\}\\
\propto\ &\exp\left\{-\frac{1}{2}\Big[\boldsymbol{x}(t)'((\boldsymbol{\Gamma}\circ\mathbf{A})'(\boldsymbol{\Gamma}\circ\mathbf{A})+\mathbf{I}+\mathbf{C}'\mathbf{R}^{-1}\mathbf{C})\boldsymbol{x}(t)\right.\\
&\qquad\left.-2(\boldsymbol{x}(t-1)'(\boldsymbol{\Gamma}\circ\mathbf{A})'+\boldsymbol{x}(t+1)'(\boldsymbol{\Gamma}\circ\mathbf{A})+\boldsymbol{y}(t)'\mathbf{R}^{-1}\mathbf{C})\boldsymbol{x}(t)\Big]\right\}.
\end{aligned}$$

Let $\mathbf{U}_t = (\boldsymbol{\Gamma}\circ\mathbf{A})'(\boldsymbol{\Gamma}\circ\mathbf{A})+\mathbf{I}+\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}$ and $\mathbf{V}_t = \boldsymbol{x}(t-1)'(\boldsymbol{\Gamma}\circ\mathbf{A})'+\boldsymbol{x}(t+1)'(\boldsymbol{\Gamma}\circ\mathbf{A})+$

$\boldsymbol{y}(t)^{'}\mathbf{R}^{-1}\mathbf{C}$. We sample $\boldsymbol{x}(t)$ from

$$\boldsymbol{x}(t)|\boldsymbol{x}(t-1),\boldsymbol{x}(t+1),\boldsymbol{y}(t),\boldsymbol{\Gamma},\mathbf{A},\mathbf{C},\mathbf{R}\sim \text{MVN}((\mathbf{U}_t)^{-1}(\mathbf{V}_t)^{'},(\mathbf{U}_t)^{-1}).$$

- For $\boldsymbol{x}(T)$: $p(\boldsymbol{x}(T)|\boldsymbol{x}(T-1),\boldsymbol{y}(T),\boldsymbol{\Gamma},\mathbf{A},\mathbf{C},\mathbf{R})$ is proportional to

$$\begin{aligned}&\exp\Big\{-\frac{1}{2}[\boldsymbol{x}(T)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(T-1)]^{'}[\boldsymbol{x}(T)-(\boldsymbol{\Gamma}\circ\mathbf{A})\boldsymbol{x}(T-1)]\Big\}\\ &\cdot\exp\Big\{-\frac{1}{2}[\boldsymbol{y}(T)-\mathbf{C}\boldsymbol{x}(T)]^{'}\mathbf{R}^{-1}[\boldsymbol{y}(T)-\mathbf{C}\boldsymbol{x}(T)]\Big\}\\ &\propto\ \exp\left\{-\frac{1}{2}\Big[\boldsymbol{x}(T)^{'}(\mathbf{I}+\mathbf{C}^{'}\mathbf{R}^{-1}\mathbf{C})\boldsymbol{x}(T)-2(\boldsymbol{x}(T-1)^{'}(\boldsymbol{\Gamma}\circ\mathbf{A})^{'}+\boldsymbol{y}(T)^{'}\mathbf{R}^{-1}\mathbf{C})\boldsymbol{x}(T)\Big]\right\}\end{aligned}$$

  Let $\mathbf{U}_T=\mathbf{I}+\mathbf{C}^{'}\mathbf{R}^{-1}\mathbf{C}$ and $\mathbf{V}_T=\boldsymbol{x}(T-1)^{'}(\boldsymbol{\Gamma}\circ\mathbf{A})^{'}+\boldsymbol{y}(T)^{'}\mathbf{R}^{-1}\mathbf{C}$. We sample $\boldsymbol{x}(T)$ from

$$\boldsymbol{x}(T)|\boldsymbol{x}(T-1),\boldsymbol{y}(T),\boldsymbol{\Gamma},\mathbf{A},\mathbf{C},\mathbf{R}\sim \text{MVN}((\mathbf{U}_T)^{-1}(\mathbf{V}_T)^{'},(\mathbf{U}_T)^{-1}).$$

A.2.4 Update $\boldsymbol{c}$

For each $c_i,\ \ i=1,\ldots,d$: $p(c_i|\mathbf{X},\mathbf{Y},\mathbf{R})$ is proportional to

$$\begin{aligned}&\exp\Big\{-\frac{1}{2}\frac{\sum_{t=1}^{T}[y_i(t)-c_ix_i(t)]^2}{\tau_i}\Big\}\cdot\phi(\frac{c_i}{\xi_1})\\ &\cdot\exp\Big\{-\frac{1}{2}\Big[(\sum_{t=1}^{T}\frac{x_i(t)^2}{\tau_i}+\frac{1}{\xi_1^2})c_i^2-2\sum_{t=1}^{T}\frac{x_i(t)y_i(t)}{\tau_i}c_i\Big]\Big\}.\end{aligned}$$

Let $u_i=\sum_{t=1}^{T}\frac{x_i(t)^2}{\tau_i}+\frac{1}{\xi_1^2},\ \ v_i=\sum_{t=1}^{T}\frac{x_i(t)y_i(t)}{\tau_i}$. We sample c_i from $c_i|\mathbf{X},\mathbf{Y},\mathbf{R}\sim \text{N}(\frac{v_i}{u_i},\frac{1}{u_i})$.

A.2.5 Update R

For each $\tau_i,\ \ i=1,\ldots,d$:

$$p(\tau_i|\mathbf{X},\mathbf{Y},\mathbf{C})\propto\ (\tau_i)^{-\frac{T}{2}-\rho_0-1}\exp\Big\{-\frac{1}{2\tau_i}\sum_{t=1}^{T}[y_i(t)-c_ix_i(t)]^2-\frac{\rho_0}{\tau_i}\Big\}.$$

We sample τ_i from

$$\tau_i|\mathbf{X},\mathbf{Y},\mathbf{C}\sim \text{Inv-Gamma}\,(\frac{T}{2}+\rho_0,\frac{1}{2}\sum_{t=1}^{T}[y_i(t)-c_ix_i(t)]^2+\rho_0).$$

A.2.6 Update $\boldsymbol{\mu}$

Let $\mathbf{U}_\mu = (1 + \frac{1}{\xi_1^2})\mathbf{I}$, $\mathbf{V}_\mu = \boldsymbol{x}(0)'$. We sample $\boldsymbol{\mu}$ from

$$\boldsymbol{\mu}|\boldsymbol{x}(0) \sim \text{MVN}((\mathbf{U}_\mu)^{-1}(\mathbf{V}_\mu)', (\mathbf{U}_\mu)^{-1}).$$

A.2.7 Update $\boldsymbol{p}$

Define a K-dimensional vector $\boldsymbol{\iota}$: the kth entry of $\boldsymbol{\iota}$ is $\iota_k = \sum_{i=1}^d m_{ik}$ for $k = 1, \ldots, K$. We sample $\boldsymbol{p}$ from

$$\boldsymbol{p}|\mathbf{M} \sim \text{Dirichlet}\,(\boldsymbol{\iota} + \mathbf{1}_K),$$

where $\mathbf{1}_K$ is a K-dimensional vector with all entries equal to 1.

A.2.8 Update $\boldsymbol{m}_i$

We update $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_d$ in sequence.

For $\boldsymbol{m}_i$, $i = 1, \ldots, d$, let $J_k = \prod_{j=1}^d (\boldsymbol{m}_i' \mathbf{B} \boldsymbol{m}_j)^{\gamma_{ij}} (1 - \boldsymbol{m}_i' \mathbf{B} \boldsymbol{m}_j)^{1-\gamma_{ij}} \cdot \prod_{\substack{j=1 \\ j \neq i}}^d (\boldsymbol{m}_j' \mathbf{B} \boldsymbol{m}_i)^{\gamma_{ji}} (1 - \boldsymbol{m}_j' \mathbf{B} \boldsymbol{m}_i)^{1-\gamma_{ji}} \cdot p_l|_{\boldsymbol{m}_i = \boldsymbol{e}_k}$, where $\boldsymbol{e}_k$ is a K-dimensional vector with its kth entry equals to 1 and the rest equal to 0.

We sample $\boldsymbol{m}_i$ from

$$p(\boldsymbol{m}_i = \boldsymbol{e}_k | \mathbf{M}_{-i}, \mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{p}) = \frac{J_k}{\sum_{k=1}^K J_k},$$

where $\mathbf{M}_{-i}$ denotes all the columns of $\mathbf{M}$ excluding $\boldsymbol{m}_i$.

A.2.9 Update B

- For $k = 1, \ldots, K$, we sample B_{kk} from

$$B_{kk}|\boldsymbol{m}, \boldsymbol{\Gamma} \sim \text{Beta}\Big(1 + \sum_{\{i,j:\boldsymbol{m}_i = \boldsymbol{e}_k, \boldsymbol{m}_j = \boldsymbol{e}_k\}} \gamma_{ij},\ 1 + \sum_{\{i,j:\boldsymbol{m}_i = \boldsymbol{e}_k, \boldsymbol{m}_j = \boldsymbol{e}_k\}} (1 - \gamma_{ij})\Big) \cdot 1\{B_{kk} \in (l_0, 1)\}$$

- For $k_1, k_2 = 1, \ldots, K,\ \ k_1 \neq k_2$, we sample $B_{k_1k_2}$ from

$$B_{k_1k_2}|\boldsymbol{m}, \boldsymbol{\Gamma} \sim \text{Beta}\Big(1 + \sum_{\{i,j:\boldsymbol{m}_i=\boldsymbol{e}_{k_1},\boldsymbol{m}_j=\boldsymbol{e}_{k_2}\}} \gamma_{ij},\ \ 1 + \sum_{\{i,j:\boldsymbol{m}_i=\boldsymbol{e}_{k_1},\boldsymbol{m}_j=\boldsymbol{e}_{k_2}\}} (1-\gamma_{ij})\Big) \cdot 1\{B_{k_1k_2} \in (0, u_0)\}$$

A.3 Hypothesis Testing on Network Changes at Seizure Onset

This section explains how we evaluated the statistical significance of the increase in the number of directed connections after seizure onset for the SOZ. We performed a hypothesis test for comparing the number of directed connections of the SOZ in the immediate postictal window (0-25 seconds after seizure onset) versus those in the interictal windows. For each patient, we first calculated the numbers of directed connections of the SOZ in all the interictal windows (25-second length, 1-second overlap) before seizure onset and used these numbers to create the null distribution for the number of directed connections of the SOZ. Against this null distribution, we obtained the p-value of the number of directed connections of the SOZ in the immediate postictal window. Similarly, we obtained the p-value of comparing the size of the SOZ module in the immediate postictal window versus those in the interictal windows. Our analysis showed that for all six patients, both increases in the number of directed connections and the size of the SOZ right after seizure onset were statistically significant (p-values $\leq$3%).

A.4 Videos for the Brain Network Evolution

In the following link, we provide videos for brain network changes of Patient 1 and 2 from 300 seconds before to 300 seconds after seizure onset: https://github.com/Yaotian-Wang/High-Dimensional-Directional-Brain-Network-Analysis-for-Focal-Epileptic-Seizures.

Appendix B Appendix of Chapter 3

B.1 Posterior Distribution with Given X

We provide explicit formulas of the posterior distribution, $p(\mathbf{\Theta}|\mathbf{X})$, below.

$$
\begin{aligned}
p(\mathbf{\Theta}|\mathbf{X}) &\propto \prod_{s=1}^{S}\prod_{l=1}^{L}\Big\{\prod_{t=1}^{T_l} p\Big(\boldsymbol{x}^{s,l}(t)\Big|\boldsymbol{x}^{s,l}(t-1),\mathbf{\Theta}\Big)\Big\}\cdot p(\mathbf{\Theta}) \\
&\propto \prod_{s=1}^{S}\prod_{l=1}^{L}\Big\{\prod_{t=1}^{T_l}\exp\Big\{-\frac{1}{2}[\boldsymbol{x}^{s,l}(t)-(\mathbf{\Gamma}^s\circ\mathbf{A}^{s,l})\boldsymbol{x}^{s,l}(t-1)]'[\boldsymbol{x}^{s,l}(t)-(\mathbf{\Gamma}^s\circ\mathbf{A}^{s,l})\boldsymbol{x}^{s,l}(t-1)]\Big\}\Big\} \\
&\cdot\prod_{s=1}^{S}\Big\{\prod_{i,j=1,i\neq j}^{d}\Big((\boldsymbol{m}_i^s)'\ \mathbf{B}\ \boldsymbol{m}_j^s\Big)^{\gamma_{ij}^s}\Big(1-(\boldsymbol{m}_i^s)'\ \mathbf{B}\ \boldsymbol{m}_j^s\Big)^{1-\gamma_{ij}^s}\Big\}\cdot\prod_{s=1}^{S}\Big\{\prod_{i=1}^{d}\prod_{k=1}^{K}p_{ik}^{m_{ik}^s}\Big\} \\
&\cdot\prod_{i=1}^{d}\Big\{\prod_{k=1}^{K}p_{ik}^{\frac{1}{K}-1}\cdot 1_{\{0<p_{ik}<1\}}\cdot 1_{\{\sum_{k=1}^{K}p_{ik}=1\}}\Big\} \\
&\cdot\prod_{k=1}^{K}1_{\{l_0<B_{kk}<1\}}\cdot\prod_{k_1,k_2=1,k_1\neq k_2}^{K}1_{\{0<B_{k_1k_2}<u_0\}}\cdot\prod_{s=1}^{S}\prod_{i,j=1,i\neq j}^{d}\prod_{l=1}^{L}\exp\Big\{-\frac{1}{2\xi_0^2}(A_{ij}^{s,l})^2\Big\}\Big\}.
\end{aligned}
$$

B.2 Variational Bayesian Algorithm

We explain detailed steps in the optimization algorithm for minimizing the KL-divergence $\mathrm{KL}\left(q(\mathbf{\Theta}|\mathbb{V})||p(\mathbf{\Theta}|\mathbf{X})\right)$ below.

We let $K=d$. In the initial step, we set the initial values of $\Phi_k^{\boldsymbol{m}_i^s}$ to be one if $k=i$, and to be zero if $k\neq i$. We set the initial values of $\Phi_k^{\boldsymbol{p}_i}$ to be $(S+1)/K$. In the initial step, we let $\alpha_{ij}^s, u_{ij}^{s,1},\ldots,u_{ij}^{s,L}$ equal 0 and $w_{ij}^{s,1},\ldots,w_{ij}^{s,L}$ equal ξ_0^2 for $i,j=1,\ldots,d,\ i\neq j$. We also let $\mathbb{E}_{q_4(B_{k_1k_2})}[\log B_{k_1k_2}]=\log l_0$ for $k_1=k_2$ and $\mathbb{E}_{q_4(B_{k_1k_2})}[\log B_{k_1k_2}]=\log u_0$ for $k_1\neq k_2$; $\mathbb{E}_{q_4(B_{k_1k_2})}[\log(1-B_{k_1k_2})]=\log(1-l_0)$ for $k_1=k_2$ and $\mathbb{E}_{q_4(B_{k_1k_2})}[\log(1-B_{k_1k_2})]=\log(1-u_0)$ for $k_1\neq k_2$.

In the following, for each parameter θ, we use $\mathbb{E}_{-\theta}$ to denote the integration operation that integrates out all the parameters except for θ under the parameters' variational distributions.

B.2.1 Update $\mathbf{\Phi}^{\boldsymbol{m}_i^s}$s

For each subject s, we update $\mathbf{\Phi}^{\boldsymbol{m}_i^s}, i = 1, \ldots, d$ in sequence. For arbitrary $k = 1, \ldots, K$, we define

$$g_i(k,s) = \sum_{j=1, j\neq i}^{d} \Bigg\{ \alpha_{ij}^s \sum_{h=1}^{K} \Phi_h^{\boldsymbol{m}_j^s} \cdot \mathbb{E}_{q_4(B_{kh})}[\log(B_{kh})] + (1-\alpha_{ij}^s) \sum_{h=1}^{K} \Phi_h^{\boldsymbol{m}_j^s} \cdot \mathbb{E}_{q_4(B_{kh})}[\log(1-B_{kh})] \\ + \alpha_{ji}^s \sum_{h=1}^{K} \Phi_h^{\boldsymbol{m}_j^s} \cdot \mathbb{E}_{q_4(B_{hk})}[\log(B_{hk})] + (1-\alpha_{ji}^s) \sum_{h=1}^{K} \Phi_h^{\boldsymbol{m}_j^s} \cdot \mathbb{E}_{q_4(B_{hk})}[\log(1-B_{hk})] \Bigg\} + \psi(\Phi_k^{\boldsymbol{p}_i}),$$

where $\psi(\cdot)$ is the digamma function.

We update $\Phi_k^{\boldsymbol{m}_i^s}$ to be $\dfrac{\exp\Big\{ g_i(k,s) \Big\}}{\sum_{h=1}^{K} \exp\Big\{ g_i(h,s) \Big\}}$.

B.2.2 Update $\mathbf{\Phi}_{ij}^s$s

For each subject s, we update $\mathbf{\Phi}_{ij}^s, i, j = 1, \ldots, d.\ i \neq j$ in sequence. We define

$$\mathcal{L} = -\int q_1(A_{ij}^{s,1}, \ldots, A_{ij}^{s,L}, \gamma_{ij}^s | \mathbf{\Phi}_{ij}^s) \Big(\mathbb{E}_{-(A_{ij}^{s,1}, \ldots, A_{ij}^{s,L}, \gamma_{ij}^s)}[\log p(\mathbf{\Theta}, \mathbf{X})] \Big) \mathrm{d}A_{ij}^{s,1} \cdots \mathrm{d}A_{ij}^{s,L} \mathrm{d}\gamma_{ij}^s \\ + \int q_1(A_{ij}^{s,1}, \ldots, A_{ij}^{s,L}, \gamma_{ij}^s | \mathbf{\Phi}_{ij}^s) \log q_1(A_{ij}^{s,1}, \ldots, A_{ij}^{s,L}, \gamma_{ij}^s | \mathbf{\Phi}_{ij}^s) \mathrm{d}A_{ij}^{s,1} \cdots \mathrm{d}A_{ij}^{s,L} \mathrm{d}\gamma_{ij}^s. \tag{B.1}$$

We look for the optimal values of $\mathbf{\Phi}_{ij}^s$s which minimize $\mathcal{L}$. We start from calculating the quantity in the brackets of the first term of (B.1).

$$\begin{aligned} & \mathbb{E}_{-(A_{ij}^{s,1}, \ldots, A_{ij}^{s,L}, \gamma_{ij}^s)}[\log p(\mathbf{\Theta}, \mathbf{X})] \\ = & -\frac{1}{2}\gamma_{ij}^s \sum_{l=1}^{L} \Big\{ (A_{ij}^{s,l})^2 \sum_{t=1}^{T_l} x_j^{s,l}(t-1)^2 \Big\} - \gamma_{ij}^s \sum_{l=1}^{L} \Big\{ A_{ij}^{s,l} \sum_{k=1, k\neq i,j}^{d} \alpha_{ik}^s u_{ik}^{s,l} \Big(\sum_{t=1}^{T_l} x_j^{s,l}(t-1) x_k^{s,l}(t-1) \Big) \Big\} \\ & + \gamma_{ij}^s \sum_{l=1}^{L} \Big\{ A_{ij}^{s,l} \sum_{t=1}^{T_l} x_i^{s,l}(t) x_j^{s,l}(t-1) \Big\} + \gamma_{ij}^s \cdot f_i(j,s) - \frac{1}{2\xi_0^2} \sum_{l=1}^{L} (A_{ij}^{s,l})^2 + \text{Constant}. \end{aligned} \tag{B.2}$$

$f_i(j,s)$ in (B.2) takes the following form:

$$f_i(j,s) = \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \Phi_{k_1}^{\boldsymbol{m}_i^s}\Phi_{k_2}^{\boldsymbol{m}_j^s}\Big(\mathbb{E}_{q_4(B_{k_1k_2})}[\log B_{k_1k_2}] - \mathbb{E}_{q_4(B_{k_1k_2})}[\log(1-B_{k_1k_2})]\Big). \tag{B.3}$$

Substitute (B.2), (B.3) to (B.1), we have

$$\begin{aligned}
\mathcal{L} =& \alpha_{ij}^s \cdot \Bigg\{ \frac{1}{2}\sum_{l=1}^{L}\Bigg\{\Big((u_{ij}^{s,l})^2 + w_{ij}^{s,l}\Big)\Big(\sum_{t=1}^{T_l} x_j^{s,l}(t-1)^2\Big)\Bigg\} \\
&+ \sum_{l=1}^{L}\Bigg\{u_{ij}^{s,l}\sum_{k=1,k\neq i,j}^{d}\alpha_{ik}^s u_{ik}^{s,l}\Big(\sum_{t=1}^{T_l} x_j^{s,l}(t-1)x_k^{s,l}(t-1)\Big)\Bigg\} \\
&- \sum_{l=1}^{L}\Bigg\{u_{ij}^{s,l}\Big(\sum_{t=1}^{T_l} x_i^{s,l}(t)x_j^{s,l}(t-1)\Big)\Bigg\} - f_i(j,s) + \frac{1}{2\xi_0^2}\sum_{l=1}^{L}\Big((u_{ij}^{s,l})^2 + w_{ij}^{s,l}\Big)\Bigg\} + \frac{L}{2}(1-\alpha_{ij}^s) \\
&+ \alpha_{ij}^s\cdot\Big(\log(\alpha_{ij}^s) - \frac{1}{2}\sum_{l=1}^{L}\log(w_{ij}^{s,l})\Big) + (1-\alpha_{ij}^s)\cdot\Big(\log(1-\alpha_{ij}^s) - \frac{1}{2}\sum_{l=1}^{L}\log(\xi_0^2)\Big) + \text{Constant}.
\end{aligned} \tag{B.4}$$

Let the partial derivatives of $\mathcal{L}$ with respect to $\alpha_{ij}^s, u_{ij}^{s,l}, w_{ij}^{s,l}$ be 0. We update $\boldsymbol{\Phi}_{ij}^s$ to be

$$\begin{aligned}
w_{ij}^{s,l} &= \frac{1}{\sum_{t=1}^{T_l} x_j^{s,l}(t-1)^2 + 1/\xi_0^2}; \\
u_{ij}^{s,l} &= w_{ij}^{s,l}\cdot\Bigg\{\sum_{t=1}^{T_l} x_i^{s,l}(t)x_j^{s,l}(t-1) - \sum_{k=1,k\neq i,j}^{d}\alpha_{ik}^s u_{ik}^{s,l}\Big(\sum_{t=1}^{T_l} x_j^{s,l}(t-1)x_k^{s,l}(t-1)\Big)\Bigg\}; \\
\frac{\alpha_{ij}^s}{1-\alpha_{ij}^s} &= \prod_{l=1}^{L}\Big(\frac{w_{ij}^{s,l}}{\xi_0^2}\Big)^{\frac{1}{2}}\cdot\exp\Bigg\{\sum_{l=1}^{L}\frac{(u_{ij}^{s,l})^2}{2w_{ij}^{s,l}} + f_i(j,s)\Bigg\}.
\end{aligned} \tag{B.5}$$

From (B.5), $w_{ij}^{s,l}$ is a constant. Therefore, we only update $u_{ij}^{s,l}$ and α_{ij}^s in our iterative algorithm.

B.2.3 Update $\boldsymbol{\Phi}^{\boldsymbol{p}_i}$s

We update $\boldsymbol{\Phi}^{\boldsymbol{p}_i}, i = 1,\ldots,d$ in sequence.

$$\mathbb{E}_{-\boldsymbol{p}_i}\Big[\sum_{s=1}^{S}(\boldsymbol{m}_i^s)'\log(\boldsymbol{p}_i) + (\frac{1}{K}-1)(\mathbf{1}_K)'\log(\boldsymbol{p}_i)\Big] = \Big(\sum_{s=1}^{S}\boldsymbol{\Phi}^{\boldsymbol{m}_i^s} + (\frac{1}{K}-1)\mathbf{1}_K\Big)'\cdot\log(\boldsymbol{p}_i).$$

We update $\boldsymbol{\Phi}^{\boldsymbol{p}_i}$ to be $\frac{1}{K}\mathbf{1}_K + \sum_{s=1}^{S}\boldsymbol{\Phi}^{\boldsymbol{m}_i^s}$, where $\mathbf{1}_K$ is a K-dimensional all-ones vector.

B.2.4 Update $\mathbf{\Phi}^{B_{k_1k_2}}$s

We update $\mathbf{\Phi}^{B_{k_1k_2}}, k_1, k_2 = 1, \ldots, K$ in sequence.

- $k_1 = k_2$:

$$\beta_{1,k_1} = 1 + \sum_{s=1}^{S} \sum_{i,j=1, i\neq j}^{d} \alpha_{ij}^s \Phi_{k_1}^{\boldsymbol{m}_i^s} \Phi_{k_1}^{\boldsymbol{m}_j^s};$$
$$\beta_{2,k_1} = 1 + \sum_{s=1}^{S} \sum_{i,j=1, i\neq j}^{d} (1 - \alpha_{ij}^s) \Phi_{k_1}^{\boldsymbol{m}_i^s} \Phi_{k_1}^{\boldsymbol{m}_j^s}.$$

- $k_1 \neq k_2$:

$$\beta_{1,k_1k_2} = 1 + \sum_{s=1}^{S} \sum_{i,j=1, i\neq j}^{d} \alpha_{ij}^s \Phi_{k_1}^{\boldsymbol{m}_i^s} \Phi_{k_2}^{\boldsymbol{m}_j^s};$$
$$\beta_{2,k_1k_2} = 1 + \sum_{s=1}^{S} \sum_{i,j=1, i\neq j}^{d} (1 - \alpha_{ij}^s) \Phi_{k_1}^{\boldsymbol{m}_i^s} \Phi_{k_2}^{\boldsymbol{m}_j^s}.$$

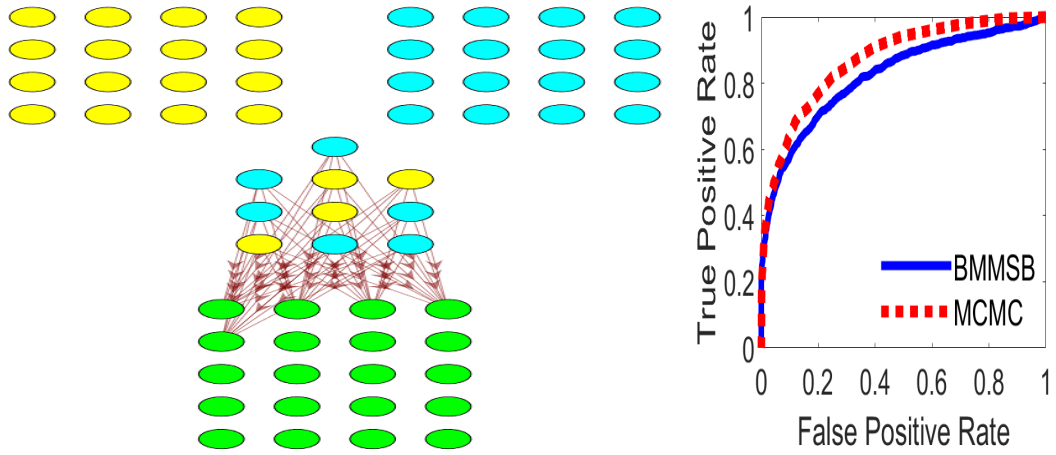
After updating all $\mathbf{\Phi}^{B_{k_1k_2}}$s, we first calculate $\{\mathbb{E}_{q_4(B_{k_1k_2})}[\log B_{k_1k_2}], \mathbb{E}_{q_4(B_{k_1k_2})}[\log(1-B_{k_1k_2})]\}_{k_1,k_2=1}^K$ right before updating $\mathbf{\Phi}^{\boldsymbol{m}_i^s}$s and $\mathbf{\Phi}_{ij}^s$s in the next iteration of our variational Bayesian algorithm, which are illustrated in Appendix B.2.1, B.2.2.

B.3 Comparison with MCMC Simulations

We performed a simulation study to compare the proposed variational Bayesian method with a fully Bayesian approach and examined the efficacy and efficiency of the proposed method in identifying directed connections.

Since an MCMC algorithm is extremely slow even for analyzing a single subject's single fMRI run in more than 50 regions, we compare the proposed variational Bayesian method with a fully Bayesian approach (where the state functions $\boldsymbol{x}^s(t)$ are also unknown) based on a single fMRI run of a single subject. We use BSBM proposed in Chapter 2 to implement the fully Bayesian approach.

Figure B.1a shows simulated network patterns with $d = 62$. Network nodes with the



(a) Simulated Network (b) ROC Curves

Figure B.1: Comparison of MCMC and Variational Bayesian Algorithms based on One Subject's Data Generated from the DCM. (a) The simulated network pattern. Nodes in the same color are in the same module in the subject's brain network. (b) ROC curves for directed connections identified by using the fully Bayesian method (MCMC) and the proposed variational method (BMMSB).

same color are in the same module in the subject's network. All network nodes in the same module are pairwise connected. We show only between-module connections in the figure for easy visualization.

Figure B.1b shows that the ROC curve of directed connection selection by our proposed variational Bayesian approach (BMMSB) is almost identical to that of a fully Bayesian approach (MCMC). The AUC of the former method is 0.82, and the AUC of the latter method is 0.87. This result suggests that the variational method can effectively approximate the target posterior distribution.

Bibliography

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

Alarcón, G. and Valentín, A. (2012), *Introduction to epilepsy*, Cambridge University Press.

Arroyo-Relión, J. D., Kessler, D., Levina, E. and Taylor, S. F. (2017), 'Network classification with applications to brain connectomics', *arXiv preprint arXiv:1701.08140* .

Baccalá, L. A. and Sameshima, K. (2001), 'Partial directed coherence: a new concept in neural structure determination', *Biological cybernetics* **84**(6), 463–474.

Basu, S. and Michailidis, G. (2015), 'Regularized estimation in sparse high-dimensional time series models'.

Bernhardt, B. C., Bonilha, L. and Gross, D. W. (2015), 'Network analysis for a network disorder: the emerging role of graph theory in the study of epilepsy', *Epilepsy & Behavior* **50**, 162–170.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017), 'Variational inference: A review for statisticians', *Journal of the American statistical Association* **112**(518), 859–877.

Bookheimer, S. Y., Salat, D. H., Terpstra, M., Ances, B. M., Barch, D. M., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Diaz-Santos, M., Elam, J. S. et al. (2019), 'The lifespan human connectome project in aging: an overview', *Neuroimage* **185**, 335–348.

Brown, P. J., Vannucci, M. and Fearn, T. (1998), 'Multivariate bayesian variable selection and prediction', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**(3), 627–641.

Buckner, R. L., Andrews-Hanna, J. R. and Schacter, D. L. (2008), 'The brain's default network: anatomy, function, and relevance to disease.'.

Burns, S. P., Santaniello, S., Yaffe, R. B., Jouny, C. C., Crone, N. E., Bergey, G. K., Anderson, W. S. and Sarma, S. V. (2014), 'Network dynamics of the brain and influence of the epileptic seizure onset zone', *Proceedings of the National Academy of Sciences* **111**(49), E5321–E5330.

Bushara, K. O., Grafman, J. and Hallett, M. (2001), 'Neural correlates of auditory–visual stimulus onset asynchrony detection', *Journal of Neuroscience* **21**(1), 300–304.

Calhoun, V. D. and Adali, T. (2012), 'Multisubject independent component analysis of fmri: a decade of intrinsic networks, default mode, and neurodiagnostic discovery', *IEEE reviews in biomedical engineering* **5**, 60–73.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H. et al. (2018), 'The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites', *Developmental cognitive neuroscience* **32**, 43–54.

Cervenka, M., Franaszczuk, P., Crone, N., Hong, B., Caffo, B., Bhatt, P., Lenz, F. and Boatman-Reich, D. (2013), 'Reliability of early cortical auditory gamma-band responses', *Clinical Neurophysiology* **124**(1), 70–82.

Chaikin, P. M., Lubensky, T. C. and Witten, T. A. (1995), *Principles of condensed matter physics*, Vol. 10, Cambridge university press Cambridge.

Cheung, B., Riedner, B., Tononi, G. and Van, V. B. (2010), 'Estimation of cortical connectivity from eeg using state-space models', *IEEE Trans Biomed Eng.* **57**(9), 2122–34.

Chiang, S., Guindani, M., Yeh, H. J., Haneef, Z., Stern, J. M. and Vannucci, M. (2017), 'Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data', *Human brain mapping* **38**(3), 1311–1332.

Cole, D. M., Smith, S. M. and Beckmann, C. F. (2010), 'Advances and pitfalls in the analysis and interpretation of resting-state fmri data', *Frontiers in systems neuroscience* p. 8.

Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P. and Mayberg, H. S. (2012), 'A whole brain fmri atlas generated via spatially constrained spectral clustering', *Human brain mapping* **33**(8), 1914–1928.

Davey, C. G., Pujol, J. and Harrison, B. J. (2016), 'Mapping the self in the brain's default mode network', *Neuroimage* **132**, 390–397.

David, O. and Friston, K. J. (2003), 'A neural mass model for meg/eeg:: coupling and neuronal dynamics', *NeuroImage* **20**(3), 1743–1755.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M. and Smith, S. M. (2006), 'fmri resting state networks define distinct modes of long-distance interactions in the human brain', *Neuroimage* **29**(4), 1359–1367.

Deshpande, G., Hu, X., Stilla, R. and Sathian, K. (2008), 'Effective connectivity during haptic perception: a study using granger causality analysis of functional magnetic resonance imaging data', *Neuroimage* **40**(4), 1807–1814.

Durante, D. and Dunson, D. B. (2014), 'Nonparametric bayes dynamic modelling of relational data', *Biometrika* **101**(4), 883–898.

Englot, D. J., Hinkley, L. B., Kort, N. S., Imber, B. S., Mizuiri, D., Honma, S. M., Findlay, A. M., Garrett, C., Cheung, P. L., Mantle, M. et al. (2015), 'Global and regional functional connectivity maps of neural oscillations in focal epilepsy', *Brain* **138**(8), 2249–2262.

Englot, D. J., Konrad, P. E. and Morgan, V. L. (2016), 'Regional and global connectivity disturbances in focal epilepsy, related neurocognitive sequelae, and potential mechanistic underpinnings', *Epilepsia* **57**(10), 1546–1557.

Euán, C., Ombao, H. and Ortega, J. (2018), 'Spectral synchronicity in brain signals', *Statistics in medicine* **37**(19), 2855–2873.

Fienberg, S. E., Meyer, M. M. and Wasserman, S. S. (1985), 'Statistical analysis of multiple sociometric relations', *Journal of the american Statistical association* **80**(389), 51–67.

Fisher, R., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P. and Engel, J. (2005), 'Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe)', *Epilepsia* **46**(4), 470–472.

Fodor, J. A. (1983), *The modularity of mind*, MIT press.

Frässle, S., Lomakina, E. I., Kasper, L., Manjaly, Z. M., Leff, A., Pruessmann, K. P., Buhmann, J. M. and Stephan, K. E. (2018), 'A generative model of whole-brain effective connectivity', *Neuroimage* **179**, 505–529.

Friedman, J., Hastie, T. and Tibshirani, R. (2014), 'glasso: Graphical lasso-estimation of gaussian graphical models', *R package version* **1**(8).

Friston, K. J. (1994), 'Functional and effective connectivity in neuroimaging: a synthesis', *Human brain mapping* **2**(1-2), 56–78.

Friston, K. J. (2011), 'Functional and effective connectivity: a review', *Brain connectivity* **1**(1), 13–36.

Friston, K. J., Harrison, L. and Penny, W. (2003), 'Dynamic causal modelling', *Neuroimage* **19**(4), 1273–1302.

Gargouri, F., Kallel, F., Delphine, S., Ben Hamida, A., Lehéricy, S. and Valabregue, R. (2018), 'The influence of preprocessing steps on graph theory measures derived from resting state fmri', *Frontiers in computational neuroscience* **12**, 8.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2013), *Bayesian data analysis*, Chapman and Hall/CRC.

Gelman, A. and Rubin, D. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical Science* **7**, 457–511.

Geng, J., Bhattacharya, A. and Pati, D. (2019), 'Probabilistic community detection with unknown number of communities', *Journal of the American Statistical Association* **114**(526), 893–905.

George, E. and McCulloch, R. (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **88**, 881–889.

George, E. and McCulloch, R. (1997), 'Approaches for bayesian variable selection', *Statistica Sinica* **7**, 339–373.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R. et al. (2013), 'The minimal preprocessing pipelines for the human connectome project', *Neuroimage* **80**, 105–124.

Goebel, R., Roebroeck, A., Kim, D.-S. and Formisano, E. (2003), 'Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping', *Magnetic resonance imaging* **21**(10), 1251–1261.

Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. M. (2010), 'A survey of statistical network models'.

Gusnard, D. A. and Raichle, M. E. (2001), 'Searching for a baseline: functional imaging and the resting human brain', *Nature reviews neuroscience* **2**(10), 685–694.

Hagler Jr, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B., Barch, D. M., Harms, M. P. et al. (2019), 'Image processing and analysis methods for the adolescent brain cognitive development study', *Neuroimage* **202**, 116091.

Hahn, G., Skeide, M. A., Mantini, D., Ganzetti, M., Destexhe, A., Friederici, A. D. and Deco, G. (2019), 'A new computational approach to estimate whole-brain effective connectivity from functional and structural mri, applied to language development', *Scientific reports* **9**.

Harrison, L., Penny, W. D. and Friston, K. (2003), 'Multivariate autoregressive modeling of fmri time series', *Neuroimage* **19**(4), 1477–1491.

Harroud, A., Bouthillier, A., Weil, A. and Nguyen, D. (2012), 'Temporal lobe epilepsy surgery failures: A review', *Epilepsy Research and Treatment* **2012**, 201651.

Hayden, D., Chang, Y. H., Goncalves, J. and Tomlin, C. J. (2016), 'Sparse network identifiability via compressed sensing', *Automatica* **68**, 9–17.

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., Tang, H., Zhu, C., Gong, Q., Zang, Y. et al. (2009), 'Uncovering intrinsic modular organization of spontaneous brain activity in humans', *PloS one* **4**(4), e5226.

Hinrichs, H., Heinze, H.-J. and Schoenfeld, M. A. (2006), 'Causal visual interactions as revealed by an information theoretic measure and fmri', *NeuroImage* **31**(3), 1051–1060.

Hoffman, M. and Blei, D. (2015), Stochastic structured variational inference, *in* 'Artificial Intelligence and Statistics', PMLR, pp. 361–369.

Holmes, E. E., Ward, E. J. and Wills, K. (2012), 'Marss: multivariate autoregressive state-space models for analyzing time-series data.', *R journal* **4**(1).

Ikeda, S., Kawano, K., Watanabe, S., Yamashita, O. and Kawahara, Y. (2022), 'Predicting behavior through dynamic modes in resting-state fmri data', *NeuroImage* **247**, 118801.

Ishwaran, H. and Rao, J. S. (2005), 'Spike and slab variable selection: frequentist and bayesian strategies', *The Annals of Statistics* **33**(2), 730–773.

Jacobs, J., Staba, R., Asano, E., Otsubo, H., Wu, J., Zijlmans, M., Mohamed, I., Kahane, P., Dubeau, F., Navarro, V. and Gotman, J. (2012), 'High-frequency oscillations (hfos) in clinical epilepsy', *Progress in Neurobiology* **98**, 302–315.

Jirsa, V. K., Stacey, W. C., Quilichini, P. P., Ivanov, A. I. and Bernard, C. (2014), 'On the nature of seizure dynamics', *Brain* **137**(8), 2210–2230.

Khambhati, A. N., Davis, K. A., Oommen, B. S., Chen, S. H., Lucas, T. H., Litt, B. and Bassett, D. S. (2015), 'Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy', *PLoS computational biology* **11**(12), e1004608.

Kiebel, S. J., Garrido, M. I., Moran, R. J. and Friston, K. J. (2008), 'Dynamic causal modelling for eeg and meg', *Cognitive neurodynamics* **2**(2), 121–136.

Kontoghiorghes, E. J. (2005), *Handbook of parallel computing and statistics*, CRC Press.

Kook, J. H., Vaughn, K. A., DeMaster, D. M., Ewing-Cobbs, L. and Vannucci, M. (2021), 'Bvar-connect: A variational bayes approach to multi-subject vector autoregressive models for inference on brain connectivity networks', *Neuroinformatics* **19**, 39–56.

Korzeniewska, A., Cervenka, M. C., Jouny, C. C., Perilla, J. R., Harezlak, J., Bergey, G. K., Franaszczuk, P. J. and Crone, N. E. (2014), 'Ictal propagation of high frequency activity is recapitulated in interictal recordings: effective connectivity of epileptogenic networks recorded with intracranial eeg', *Neuroimage* **101**, 96–113.

Korzeniewska, A., Crainiceanu, C. M., Kuś, R., Franaszczuk, P. J. and Crone, N. E. (2008), 'Dynamics of event-related causality in brain electrical activity', *Human brain mapping* **29**(10), 1170–1192.

Kramer, M. A. and Cash, S. S. (2012), 'Epilepsy as a disorder of cortical network organization', *The Neuroscientist* **18**(4), 360–372.

Kramer, M. A., Kolaczyk, E. D. and Kirsch, H. E. (2008), 'Emergent network topology at seizure onset in humans', *Epilepsy research* **79**(2-3), 173–186.

Li, F., Zhang, T., Wang, Q., Gonzalez, M., Maresh, E. and Coan, J. (2015), 'Spatial bayesian variable selection and grouping in high-dimensional covariate spaces with application to fmri', *Annals of Applied Statistics* **9**, 687–713.

Li, H., Wang, Y., Yan, G., Sun, Y., Tanabe, S., Liu, C.-C., Quigg, M. S. and Zhang, T. (2021), 'A bayesian state-space approach to mapping directional brain networks', *Journal of the American Statistical Association* **116**(536), 1637–1647.

Lindquist, M. A. (2008), 'The statistical analysis of fmri data', *Statistical science* pp. 439–464.

Liu, Y. and Aviyente, S. (2012), 'Quantification of effective connectivity in the brain using a measure of directed information', *Computational and mathematical methods in medicine* **2012**.

McCann, L. and Welsch, R. E. (2007), 'Robust variable selection using least angle regression and elemental set sampling', *Computational Statistics & Data Analysis* **52**(1), 249–257.

Mejia, A. F., Nebel, M. B., Wang, Y., Caffo, B. S. and Guo, Y. (2020), 'Template independent component analysis: Targeted and reliable estimation of subject-level brain networks using big data population priors', *Journal of the American Statistical Association* **115**(531), 1151–1177.

Mennes, M., Kelly, C., Zuo, X.-N., Di Martino, A., Biswal, B. B., Castellanos, F. X. and Milham, M. P. (2010), 'Inter-individual differences in resting-state functional connectivity predict task-induced bold activity', *Neuroimage* **50**(4), 1690–1701.

Mesulam, M.-M. (1998), 'From sensation to cognition.', *Brain: a journal of neurology* **121**(6), 1013–1052.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. and Bullmore, E. T. (2009), 'Hierarchical modularity in human brain functional networks', *Frontiers in neuroinformatics* **3**, 37.

Mill, R. D., Bagic, A., Bostan, A., Schneider, W. and Cole, M. W. (2017), 'Empirical validation of directed functional connectivity', *NeuroImage* **146**, 275–287.

Miller, A. (2002), *Subset selection in regression*, CRC Press.

Moussa, M. N., Steen, M. R., Laurienti, P. J. and Hayasaka, S. (2012), 'Consistency of network modules in resting-state fmri connectome data'.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W. and Beckett, L. (2005), 'The alzheimer's disease neuroimaging initiative', *Neuroimaging Clinics* **15**(4), 869–877.

Newman, M. E. (2006), 'Modularity and community structure in networks', *Proceedings of the national academy of sciences* **103**(23), 8577–8582.

Nicholson, W. B., Matteson, D. S. and Bien, J. (2017), 'Varx-l: Structured regularization for large vector autoregressions with exogenous variables', *International Journal of Forecasting* **33**(3), 627–651.

Nowicki, K. and Snijders, T. A. B. (2001), 'Estimation and prediction for stochastic blockstructures', *Journal of the American statistical association* **96**(455), 1077–1087.

Park, H.-J. and Friston, K. (2013), 'Structural and functional brain networks: from connections to cognition', *Science* **342**(6158), 1238411.

Paul, S. and Chen, Y. (2018), 'A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging', *arXiv preprint arXiv:1805.02292* .

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J. and Nichols, T. E. (2011), *Statistical parametric mapping: the analysis of functional brain images*, Elsevier.

Petersen, S. E. and Sporns, O. (2015), 'Brain networks and cognitive architectures', *Neuron* **88**(1), 207–219.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L. et al. (2011), 'Functional network organization of the human brain', *Neuron* **72**(4), 665–678.

Raichle, M. E. (2015), 'The brain's default mode network', *Annual review of neuroscience* **38**, 433–447.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. and Shulman, G. L. (2001), 'A default mode of brain function', *Proceedings of the National Academy of Sciences* **98**(2), 676–682.

Riedl, V., Utz, L., Castrillón, G., Grimmer, T., Rauschecker, J. P., Ploner, M., Friston, K. J., Drzezga, A. and Sorg, C. (2016), 'Metabolic connectivity mapping reveals effective connectivity in the resting human brain', *Proceedings of the National Academy of Sciences* **113**(2), 428–433.

Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T. and Kawashima, R. (2004), 'A state-space model of the hemodynamic approach: nonlinear filtering of bold signals', *NeuroImage* **21**(2), 547–567.

Rosenow, F. and Lüders, H. (2001), 'Presurgical evaluation of epilepsy', *Brain* **124**(9), 1683–1700.

Rosenthal, J. S. (2000), 'Parallel computing and monte carlo algorithms', *Far east journal of theoretical statistics* **4**(2), 207–236.

Sabesan, S., Good, L. B., Tsakalis, K. S., Spanias, A., Treiman, D. M. and Iasemidis, L. D. (2009), 'Information flow and application to epileptogenic focus localization from intracranial eeg', *IEEE transactions on neural systems and rehabilitation engineering* **17**(3), 244–253.

Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., Brammer, M. J. and Amaro Jr, E. (2010), 'Analyzing the connectivity between regions of interest: an approach based on cluster granger causality for fmri data analysis', *Neuroimage* **52**(4), 1444–1455.

Schiff, S. J., Sauer, T., Kumar, R. and Weinstein, S. L. (2005), 'Neuronal spatiotemporal pattern discrimination: the dynamical evolution of seizures', *Neuroimage* **28**(4), 1043–1055.

Schreiber, T. (2000), 'Measuring information transfer', *Physical review letters* **85**(2), 461.

Schröder, A. L. and Ombao, H. (2019), 'Fresped: Frequency-specific change-point detection in epileptic seizure multi-channel eeg data', *Journal of the American Statistical Association* **114**(525), 115–128.

Seeley, W. W. (2019), 'The salience network: a neural system for perceiving and responding to homeostatic demands', *Journal of Neuroscience* **39**(50), 9878–9882.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L. and Greicius, M. D. (2007), 'Dissociable intrinsic connectivity networks for salience processing and executive control', *Journal of Neuroscience* **27**(9), 2349–2356.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P. et al. (2013), 'Resting-state fmri in the human connectome project', *Neuroimage* **80**, 144–168.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R. et al. (2009), 'Correspondence of the brain's functional architecture during activation and rest', *Proceedings of the national academy of sciences* **106**(31), 13040–13045.

Sporns, O. (2010), *Networks of the Brain*, MIT press.

Sporns, O. (2013), 'Network attributes for segregation and integration in the human brain', *Current opinion in neurobiology* **23**(2), 162–171.

Sporns, O. and Betzel, R. F. (2016), 'Modular brain networks', *Annual review of psychology* **67**, 613–640.

Sporns, O., Honey, C. J. and Kötter, R. (2007), 'Identification and classification of hubs in brain networks', *PloS one* **2**(10), e1049.

Stacey, W., Kramer, M., Gunnarsdottir, K., Gonzalez-Martinez, J., Zaghloul, K., Inati, S., Sarma, S., Stiso, J., Khambhati, A. N., Bassett, D. S. et al. (2019), 'Emerging roles of network analysis for epilepsy', *Epilepsy Research* p. 106255.

Surges, R. and Elger, C. (2013), 'Reoperation after failed resective epilepsy surgery', *Seizure* **22**(7), 493 – 501.

Theo, H. and Mike, E. (2004), 'Mapping multiple qtl using linkage disequilibrium and linkage analysis information and multitrait data', *Genet. Sel. Evol* **36**, 261–279.

van de Ven, V. G., Formisano, E., Prvulovic, D., Roeder, C. H. and Linden, D. E. (2004), 'Functional connectivity as revealed by spatial independent component analysis of fmri measurements during rest', *Human brain mapping* **22**(3), 165–178.

Van Den Heuvel, M. P. and Pol, H. E. H. (2010), 'Exploring the brain network: a review on resting-state fmri functional connectivity', *European neuropsychopharmacology* **20**(8), 519–534.

Van Dyk, D. A. and Park, T. (2008), 'Partially collapsed gibbs samplers: Theory and methods', *Journal of the American Statistical Association* **103**(482), 790–796.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H. et al. (2013), 'The wu-minn human connectome project: an overview', *Neuroimage* **80**, 62–79.

Van Mierlo, P., Carrette, E., Hallez, H., Raedt, R., Meurs, A., Vandenberghe, S., Van Roost, D., Boon, P., Staelens, S. and Vonck, K. (2013), 'Ictal-onset localization through connectivity analysis of intracranial eeg signals in patients with refractory epilepsy', *Epilepsia* **54**(8), 1409–1418.

Varela, F., Lachaux, J.-P., Rodriguez, E. and Martinerie, J. (2001), 'The brainweb: phase synchronization and large-scale integration', *Nature reviews neuroscience* **2**(4), 229–239.

Vicente, R., Wibral, M., Lindner, M. and Pipa, G. (2011), 'Transfer entropy—a model-free measure of effective connectivity for the neurosciences', *Journal of computational neuroscience* **30**(1), 45–67.

Wainwright, M. J. and Jordan, M. I. (2008), *Graphical models, exponential families, and variational inference*, Now Publishers Inc.

Wang, Y., Yan, G., Tanabe, S., Liu, C.-C., Moosa, S., Quigg, M. S. and Zhang, T. (2022), 'High-dimensional directional brain network analysis for focal epileptic seizures', *arXiv preprint arXiv:2208.07991* .

Wang, Y., Yan, G., Wang, X., Li, S., Peng, L., Tudorascu, D. L. and Zhang, T. (2023), 'A variational bayesian approach to identifying whole-brain directed networks with fmri data', *The Annals of Applied Statistics* **17**(1), 518–538.

Wilke, C., Ding, L. and He, B. (2008), 'Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function', *IEEE transactions on biomedical engineering* **55**(11), 2557–2564.

Wilke, C., Worrell, G. and He, B. (2011), 'Graph analysis of epileptogenic networks in human partial epilepsy', *Epilepsia* **52**(1), 84–93.

Witten, D. M., Friedman, J. H. and Simon, N. (2011), 'New insights and faster computations for the graphical lasso', *Journal of Computational and Graphical Statistics* **20**(4), 892–900.

Xia, M., Wang, J. and He, Y. (2013), 'Brainnet viewer: a network visualization tool for human brain connectomics', *PloS one* **8**(7), e68910.

Yan, C.-G., Wang, X.-D., Zuo, X.-N. and Zang, Y.-F. (2016), 'Dpabi: data processing & analysis for (resting-state) brain imaging', *Neuroinformatics* **14**(3), 339–351.

Yi, N., George, V. and Allison, D. B. (2003), 'Stochastic search variable selection for identifying multiple quantitative trait loci', *Genetics* **164**(3), 1129–1138.

Zaveri, H. P., Pincus, S. M., Goncharova, I. I., Duckrow, R. B., Spencer, D. D. and Spencer, S. S. (2009), 'Localization-related epilepsy exhibits significant connectivity away from the seizure-onset area', *Neuroreport* **20**(9), 891–895.

Zeki, S., Watson, J., Lueck, C., Friston, K. J., Kennard, C. and Frackowiak, R. (1991), 'A direct demonstration of functional specialization in human visual cortex', *Journal of neuroscience* **11**(3), 641–649.

Zhang, T., Sun, Y., Li, H., Yan, G., Tanabe, S., Miao, R., Wang, Y., Caffo, B. and Quigg, M. (2020), 'Bayesian inference of a directional brain network for intracranial eeg data', *Computational Statistics and Data Analysis* **144**, 106847.

Zhang, T., Wu, J., Li, F., Caffo, B. and Boatman-Reich, D. (2015), 'A dynamic directional model for effective brain connectivity using electrocorticographic (ecog) time series', *Journal of the American Statistical Association* **110**(509), 93–106.

Zhang, T., Yin, Q., Caffo, B., Sun, Y. and Boatman-Reich, D. (2017), 'Bayesian inference of high-dimensional, cluster-structured ordinary differential equation models with applications to brain connectivity studies'.

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**(4), 2266–2292.

Zhu, X., Thung, K.-H., Adeli, E., Zhang, Y. and Shen, D. (2017), Maximum mean discrepancy based multiple kernel learning for incomplete multimodality neuroimaging data, *in* 'Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III', Springer, pp. 72–80.