Submitted to the Annals of Applied Statistics

A VARIATIONAL BAYESIAN APPROACH TO IDENTIFYING WHOLE-BRAIN DIRECTED NETWORKS WITH FMRI DATA

By Yaotian Wang[1],Guofen Yan[2],Xiaofeng Wang[3],Shuoran Li[1],Lingyi Peng[4],Dana L Tudorascu[4], and Tingting Zhang[1,*]

[1]*Department of Statistics, University of Pittsburgh,* [*]*tiz67@pitt.edu*

[2]*Department of Public Health Sciences, University of Virginia,*

[3]*Department of Quantitative Health Sciences, Cleveland Clinic,*

[4]*Department of Biostatistics, University of Pittsburgh,*

The brain is a high-dimensional directed network system as it consists of many regions as network nodes that exert influence on each other. The directed influence exerted by one region on another is referred to as directed connectivity. We aim to reveal whole-brain directed networks based on resting-state functional magnetic resonance imaging (fMRI) data of many subjects. However, it is both statistically and computationally challenging to produce scientifically meaningful estimates of whole-brain directed networks. To address the statistical modeling challenge, we assume modular brain networks, which reflect functional specialization and functional integration of the brain. We address the computational challenge by developing a variational Bayesian method to estimate the new model. We apply our method to resting-state fMRI data of many subjects and identify modules and directed connections in whole-brain directed networks. The identified modules are accordant with functional brain systems specialized for different functions. We also detect directed connections between functionally specialized modules, which is not attainable by existing network methods based on functional connectivity. In summary, this paper presents a new computationally efficient and flexible method for directed network studies of the brain as well as new scientific findings regarding the functional organization of the human brain.

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 effective connectivity (Friston, 2011)). 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;

Keywords and phrases: Directed connectivity, brain network, variational Bayes, resting-state fMRI.

Sporns and Betzel, 2016; Park and Friston, 2013) have suggested that the modular organization (also called modularity) of networks is tied with 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, i.e., 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, 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, Kolaczyk and Kirsch, 2008; Schiff et al., 2005), cross-coherence (Schröder and Ombao, 2018), transfer entropy (Vicente et al., 2011), directed transinformation (Hinrichs, Heinze and Schoenfeld, 2006), directed information (Liu and Aviyente, 2012), and many others (van Mierlo et al., 2013; Wilke, Worrell and He, 2011), quantify pairwise connectivity between regions and cannot be directly used to identify modules. Popular models such as dynamic causal modeling (DCM, Friston, Harrison and Penny, 2003; Frässle et al., 2018) and neural mass models (David and Friston, 2003) characterize directed connectivity but not modules. Methods such as independent component analysis (van de Ven et al., 2004; Calhoun and Adali, 2012; Mejia et al., 2020) 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, Honey and Kötter, 2007) and directed connections (Friston, 2011; Chiang et al., 2017; Kook et al., 2020) separately with different approaches, resulting in two different and hard-to-track errors in the estimated directed network. Despite the recent development of models (Zhang et al., 2015, 2017, 2019; Li et al., 2021) 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 (Li et al., 2021). At the population level, we assign a mixed-membership stochastic blockmodel (MMSB) as a prior for 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 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 organization. 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 (Van Essen et al., 2013, HCP). 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 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 article is organized as follows. In Section 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, we develop a variational Bayesian approach to estimate the new Bayesian model. In Section 4, we examine the robustness and effectiveness of the proposed method compared to existing network methods through a simulation study. Section 5 presents the analysis results of resting-state fMRI data of many subjects. Section 6 concludes with a discussion.

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.

2.1. *The Multivariate Autoregressive State-Space Model.* Let $\boldsymbol{y}^{s,l}(t) = (y_1^{s,l}(t), \dots, y_d^{s,l}(t))'$ 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 $s = 1, \dots, S$, $t = 1, \dots, T$ and $l = 1, \dots, L$. 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, $\{y_j^{s,l}(1), \dots, y_j^{s,l}(T)\}$, is standardized to have mean zero and variance one. Let $\boldsymbol{x}^{s,l}(t) = (x_1^{s,l}(t), \dots, x_d^{s,l}(t))'$ be the state functions of the d brain regions at time t in the lth 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 scan for subject s. We model directed connections between the d regions of each subject s using a multivariate autoregressive state-space model (MARSS):

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

$$(2) \qquad x_i^{s,l}(t) = \sum_{j=1}^{d} \gamma_{ij}^{s} \cdot A_{ij}^{s,l} \cdot x_j^{s,l}(t-1) + \eta_i^{s,l}(t), \;\; t = 1, \dots, T_l,$$

where $c_i^{s,l}$ 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 $\eta_i^{s,l}(t)$ and $\epsilon_i^{s,l}(t)$ 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 (Li et al., 2021).

We use indicators, γ_{ij}^{s}s, to distinguish nonzero directed connections from zero ones. Model (1) and (2) distinguishes 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 (Sato et al., 2010; Hayden et al., 2016), we fix $\gamma_{ii}^{s} = 0$ for $i = 1, \dots, d, \;\; s = 1, \dots, S$. 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, (1) and (2), we focus on identifying nonzero γ_{ij}^{s}s 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 the literature (Li et al., 2021) and the simulation study (Section 4). For mathematical simplicity and computational efficiency, we let $\eta_i^{s,l}(t) \overset{\text{i.i.d}}{\sim} \mathrm{N}(0,1)$ and $\epsilon_i^{s,l}(t) \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \tau_i^2)$.

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, 2011), we impose modularity on γ_{ij}^ss by using a mixed membership stochastic blockmodel (MMSB) (Fienberg, Meyer and Wasserman, 1985; Airoldi et al., 2008; Nowicki and Snijders, 2001; Durante and Dunson, 2014) 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 $k_1, k_2 = 1, \ldots, K$.

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:

$$\gamma_{ij}^s | \boldsymbol{m}_i^s, \boldsymbol{m}_j^s, \mathbf{B} \overset{\text{ind}}{\sim} \text{Bernoulli}((\boldsymbol{m}_i^s)' \, \mathbf{B} \, \boldsymbol{m}_j^s), \; i, j = 1, \ldots, d; \tag{3}$$

$$\boldsymbol{m}_i^s \overset{\text{i.i.d}}{\sim} \text{Multinomial}(1; p_{i1}, \ldots, p_{iK}) \text{ and } (p_{i1}, \ldots, p_{iK}) \sim \text{Dirichlet}(\tfrac{1}{K}\mathbf{1}_K); \tag{4}$$

$$B_{kk} \overset{\text{i.i.d}}{\sim} \text{Uniform}(l_0, 1) \text{ and } B_{k_1k_2} \overset{\text{i.i.d}}{\sim} \text{Uniform}(0, u_0), \; k_1, k_2 = 1, \ldots, K, \; k_1 \neq k_2; \tag{5}$$

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) 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 $(\boldsymbol{m}_i^s)' \, \mathbf{B} \, \boldsymbol{m}_j^s = B_{k_1k_2}$.

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)-(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,

(1) and (2), with the MMSB prior (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. 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 thousand, 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 $\boldsymbol{x}^{s,l}(t)$ using the standard MARSS (Holmes, Ward and Wills, 2012) (where γ_{ij}^ss in (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, Ward and Wills, 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 $l = 1, \ldots, L$, $X^{s,l} = \{\boldsymbol{x}^{s,l}(0), \ldots, \boldsymbol{x}^{s,l}(T)\}$, and $\mathbf{X} = \{X^{s,l}, s = 1, \ldots, S, l = 1, \ldots, L\}$. Let $\boldsymbol{\Theta}$ denote all the unknown parameters:

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

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

$$p(\boldsymbol{\Theta}|\mathbf{X}) \propto \prod_{s=1}^{S} \prod_{l=1}^{L} \left\{ \prod_{t=1}^{T_l} p\left(\boldsymbol{x}^{s,l}(t) \middle| \boldsymbol{x}^{s,l}(t-1), \boldsymbol{\Theta}\right) \right\} \cdot p(\boldsymbol{\Theta}), \tag{6}$$

where $p\left(\boldsymbol{x}^{s,l}(t) \middle| \boldsymbol{x}^{s,l}(t-1), \boldsymbol{\Theta}\right)$ is derived using the state model (2). The prior distribution for the parameters γ_{ij}^s, $\boldsymbol{m}_i^s$, and $\mathbf{B}$ is the MMSB prior, (3), (4), and (5). We assign normal priors to A_{ij}^ss:

$$A_{ij}^{s,l} \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \xi_0^2), \tag{7}$$

where ξ_0 is a pre-specified positive constant. Explicit formulas of the posterior distribution, $p(\boldsymbol{\Theta}|\mathbf{X})$ are provided in Section 1 of the Supplementary Material (Wang et al., 2022).

We use a variational method to approximate the posterior distribution $p(\boldsymbol{\Theta}|\mathbf{X})$ in (6). Variational methods (Blei, Kucukelbir and McAuliffe, 2017) have received enormous popularity in estimating graphical models and network models (Fienberg, Meyer and Wasserman, 1985; Airoldi et al., 2008; Nowicki and Snijders, 2001; Durante and Dunson, 2014; 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 $p(\boldsymbol{\Theta}|\mathbf{X})$. The factorized distribution is given as follows:

(8)

$$q(\boldsymbol{\Theta}|\mathbb{V}) = \prod_{s=1}^{S} \prod_{i,j=1,i\neq j}^{d} q_1(\mathrm{A}_{ij}^{s,1}, \ldots, \mathrm{A}_{ij}^{s,L}, \gamma_{ij}^{s}|\boldsymbol{\Phi}_{ij}^{s}) \cdot \prod_{s=1}^{S} \prod_{i=1}^{d} q_2(\boldsymbol{m}_i^s|\boldsymbol{\Phi}^{\boldsymbol{m}_i^s}) \cdot \prod_{i=1}^{d} q_3(\boldsymbol{p}_i|\boldsymbol{\Phi}^{\boldsymbol{p}_i})$$

$$\cdot \prod_{k_1,k_2=1}^{K} q_4(\mathrm{B}_{k_1k_2}|\boldsymbol{\Phi}^{\mathrm{B}_{k_1k_2}}),$$

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

The variational distribution factors in the factorized distribution (8) and their variational parameters are given below:

$$q_1(\gamma_{ij}^s|\boldsymbol{\Phi}_{ij}^s) = \text{Bernoulli}(\gamma_{ij}^s|\alpha_{ij}^s);$$

$$q_1(\mathrm{A}_{ij}^{s,1}, \ldots, \mathrm{A}_{ij}^{s,L}|\gamma_{ij}^s, \boldsymbol{\Phi}_{ij}^s) = \prod_{l=1}^{L} q_1(\mathrm{A}_{ij}^{s,l}|\gamma_{ij}^s, u_{ij}^{s,l}, w_{ij}^{s,l}),$$

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

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

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

$$q_4(\mathrm{B}_{k_1k_2}|\boldsymbol{\Phi}^{\mathrm{B}_{k_1k_2}}) = \begin{cases} \text{Beta}(\mathrm{B}_{k_1k_1}|\beta_{1,k_1}, \beta_{2,k_1}) \cdot 1_{\{l_0<\mathrm{B}_{k_1k_2}<1\}} & \text{if } k_1 = k_2, \\ \text{Beta}(\mathrm{B}_{k_1k_2}|\beta_{1,k_1k_2}, \beta_{2,k_1k_2}) \cdot 1_{\{0<\mathrm{B}_{k_1k_2}<u_0\}} & \text{if } k_1 \neq k_2; \end{cases}$$

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 which 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 (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, Lubensky and Witten, 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 (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 (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 approximation distribution $q(\boldsymbol{\Theta}|\mathbb{V})$ and the posterior distribution $p(\boldsymbol{\Theta}|\mathbf{X})$:

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

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 $k \neq i$, $i = 1, \dots, d$, and $s = 1, \dots, S$. The initial values of the other variational parameters and detailed steps in the iterative optimization algorithm for evaluating variational parameters are provided in Section 2 of the Supplementary Material (Wang et al., 2022).

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

Algorithm 1 Pseudocode for 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 $s = 1, \dots, S$ and $i = 1, \dots, d$:
        Update $\boldsymbol{\Phi}^{\boldsymbol{m}_i^s}$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.
2. For $s = 1, \dots, S$ and $i, j = 1, \dots, d$:
        Update $\boldsymbol{\Phi}_{ij}^s$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.
3. For $i = 1, \dots, d$:
        Update $\boldsymbol{\Phi}^{\boldsymbol{p}_i}$ in $\mathbb{V}$ based on the rest parameters in $\mathbb{V}$.
4. For $k_1, k_2 = 1, \dots, K$:
        Update $\boldsymbol{\Phi}^{B_{k_1 k_2}}$ 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
```

We use $KL^t - KL^{t-1}$ to examine the convergence of the iterative optimization algorithm, because the KL-divergence can be evaluated only up to an arbitrary additive constant, and $KL^t - KL^{t-1}$ does not involve this constant. The algorithm terminates when $KL^t - KL^{t-1}$ is smaller than 1% of the maximum possible change in the KL-divergence, i.e., M_{KL}.

We employ parallel computing (Rosenthal, 2000; Kontoghiorghes, 2005) 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.

3.1. *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.

3.1.1. *Identification of Modules in Subject-Specific Brain Networks.* Intuitively, given an appropriate number of modules K, one can use the variational parameters $\boldsymbol{\Phi}^{\boldsymbol{m}_i^s}$ 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, Kucukelbir and McAuliffe (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, $\Phi_k^{\boldsymbol{m}_i^s}$, 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 number of modules and reevaluate posterior probabilities of each region i in different modules. We first identify the regions that are in the same module in most 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; and (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 (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

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

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.

3.1.2. *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 2 different modules.

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 (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 module in the directed brain network of subject s (i.e., $\hat{\Phi}_{k_{(1)}}^{\boldsymbol{m}_i^s} < 50\%$), $\mathcal{C}_{i,\xi_0^2}^s = \emptyset$. 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

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

We choose the ξ_0^2 that leads to the smallest $\text{BIC}(\xi_0^2)$ 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. *Directed Connection Identification.*

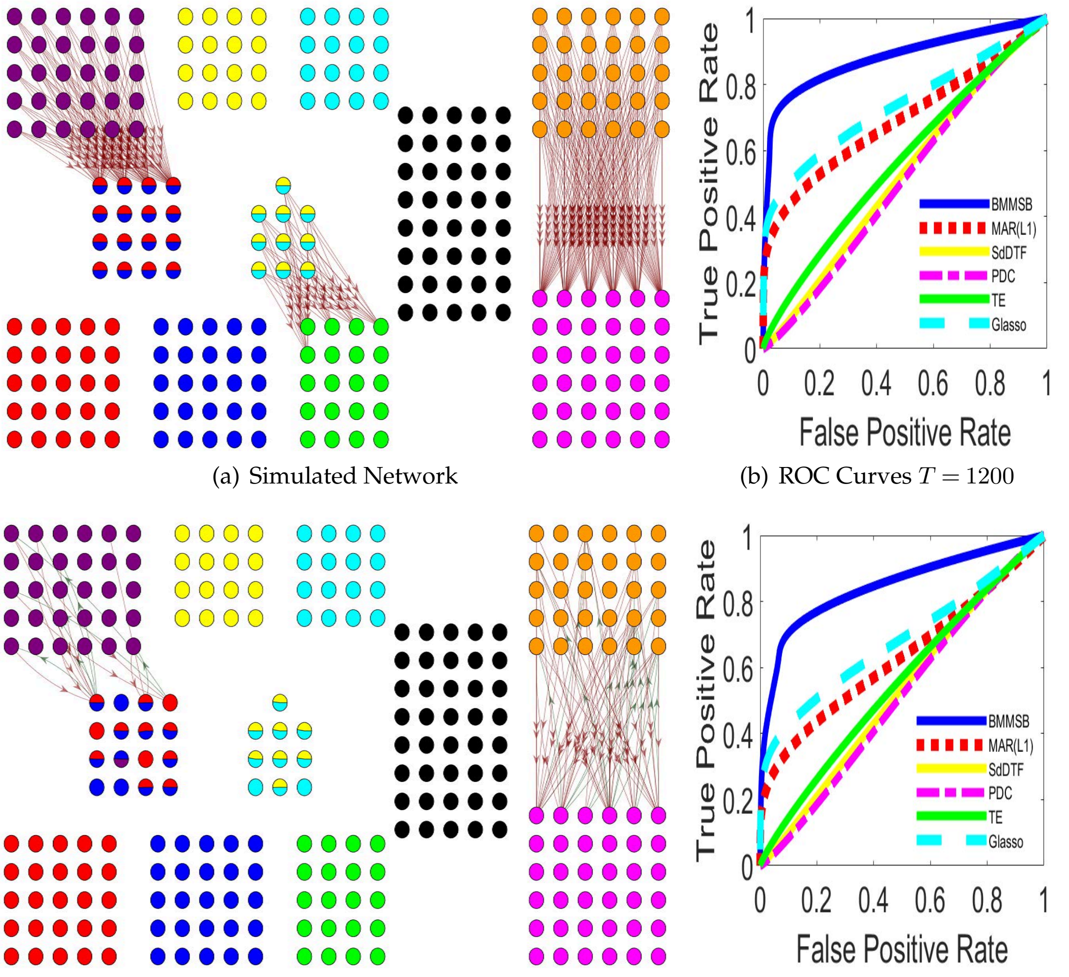
We use α_{ij}^s to identify directed connections in the subject-specific directed network for each subject s and use average posterior probabilities $\bar{\alpha}_{ij} = \sum_{s=1}^{S} \alpha_{ij}^s / S$, $i, j = 1, \ldots, d$ to identify directed connections in the population-mean directed network.

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.

4. Simulation Studies.

We used SPM software (Penny et al., 2011) to simulate fMRI data from the DCM (Friston, Harrison and Penny, 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 $d = 264$ regions 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 $\boldsymbol{y}^{s,l}(t)$, 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 0.72s as our simulated data, the same as the TR of real fMRI data under study.

Figure 1(a) shows simulated network patterns. We used the BrainNet Viewer (Xia, Wang and He, 2013) to visualize networks. The number of modules and the sizes of



(a) Simulated Network

(b) ROC Curves $T = 1200$

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

(d) ROC Curves $T = 600$

Figure 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 network methods. (c) The estimated population-mean directed network. (d) ROC curves for directed connections identified by six network methods based on data with $T = 600$ time points.

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 with 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, Matteson and Bien, 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, Hastie and Tibshirani, 2014; Witten, Friedman and Simon, 2011). Figure 1(b) 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.

Figure 1(c) 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 towards 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 1(d) 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

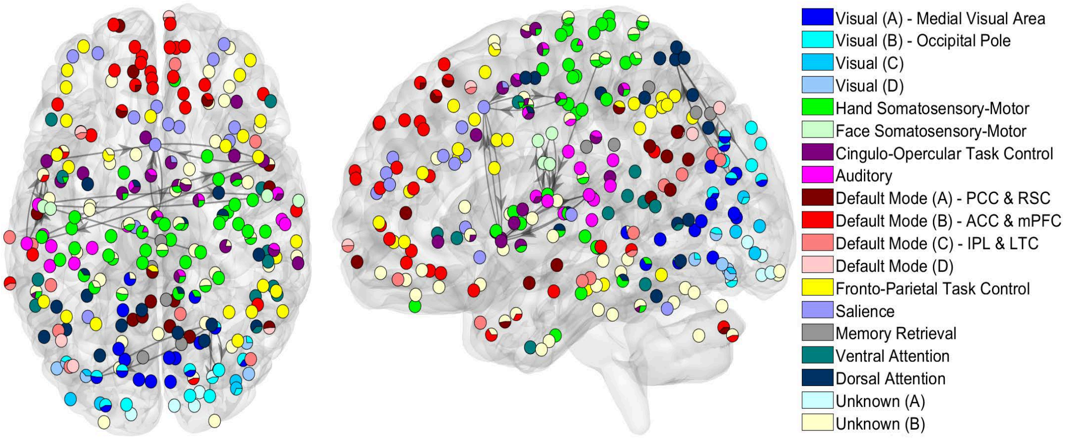
$d = 62$ regions of a single subject. The ROC curve of the variational method is only slightly lower than that of the fully Bayesian approach: The AUC of the former method is 0.82, and the AUC of the latter method is 0.87. This result suggests that the variational method can effectively approximate the target posterior distribution. More details of this simulation study can be found in Section 3 of the Supplementary Material (Wang et al., 2022).

5. An Application to an fMRI Study. 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 1-hour (in total) resting-state fMRI scanning at 3T (Smith et al., 2013) in two pairs of 15-min runs on each of two separate days. The data of each subject per run consist of functional images at $T = 1200$ 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 paper by Glasser et al. (2013); Smith et al. (2013). Following the practice by Power et al. (2011), we extracted fMRI time series from the 10mm-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 systems 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.

Identification of modules. Our method identified modules specialized for different functions, though the method did not use spatial information of regions. Figure 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


Visual (A) - Medial Visual Area
Visual (B) - Occipital Pole
Visual (C)
Visual (D)
Hand Somatosensory-Motor
Face Somatosensory-Motor
Cingulo-Opercular Task Control
Auditory
Default Mode (A) - PCC & RSC
Default Mode (B) - ACC & mPFC
Default Mode (C) - IPL & LTC
Default Mode (D)
Fronto-Parietal Task Control
Salience
Memory Retrieval
Ventral Attention
Dorsal Attention
Unknown (A)
Unknown (B)

(a) Axial View

(b) Sagittal View

Figure 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.

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 paper, we use modules instead of networks.

Our method revealed several smaller modules in large functional brain systems, such as the visual network and the default mode network. These results align with the literature that the visual network (Zeki et al., 1991) and the default mode network (Buckner, Andrews-Hanna and Schacter, 2008) consist of several functionally and anatomically different brain areas. Moreover, the identified small visual modules overlap with several known functional subsystems in the visual network, 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 network (Raichle, 2015; Davey, Pujol and Harrison, 2016). 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.

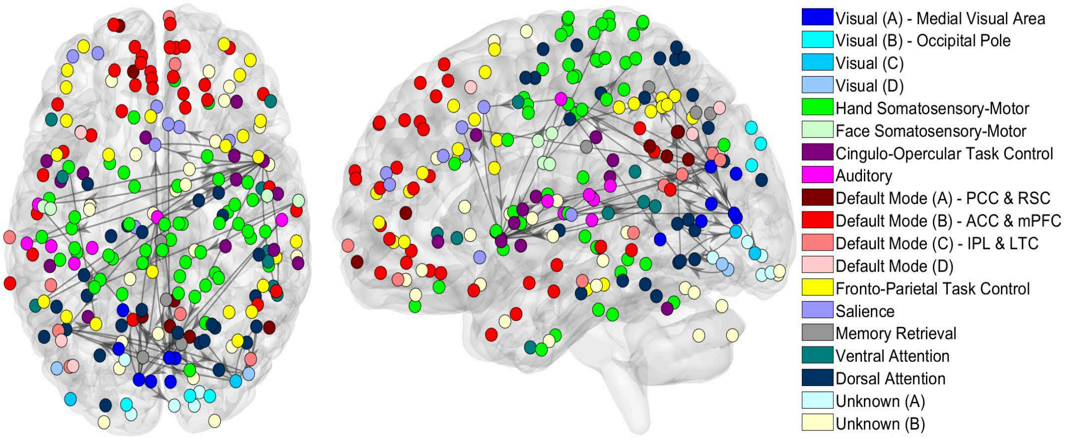
Identification of directed connections. Most of the identified directed connections are between regions in the same module or between modules with similar brain functions (e.g., between the four visual modules). These connections are dense, as expected. For easy visualization of directed connections between different functionally specialized modules, we show only directed connections between modules with different specialized functions in Figure 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 (He et al., 2009; Mesulam, 1998; De Luca et al., 2006), 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).

Variation of directed brain networks across subjects. We examined the variation of directed brain networks across subjects. Figure 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 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 2. These results are consistent with the findings in the literature (Power et al., 2011; Riedl et al., 2016; Seeley et al., 2007; Deshpande et al., 2008; Bushara, Grafman and Hallett, 2001) 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, 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 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.


Visual (A) - Medial Visual Area
Visual (B) - Occipital Pole
Visual (C)
Visual (D)
Hand Somatosensory-Motor
Face Somatosensory-Motor
Cingulo-Opercular Task Control
Auditory
Default Mode (A) - PCC & RSC
Default Mode (B) - ACC & mPFC
Default Mode (C) - IPL & LTC
Default Mode (D)
Fronto-Parietal Task Control
Salience
Memory Retrieval
Ventral Attention
Dorsal Attention
Unknown (A)
Unknown (B)

(a) Axial View

(b) Sagittal View

Figure 3: The Identified Whole-Brain Directed Networks of One Subject in Axial (a) and Sagittal (b) Views. 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.

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 4. The network is similar to the first population-mean brain network (shown in Figure 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

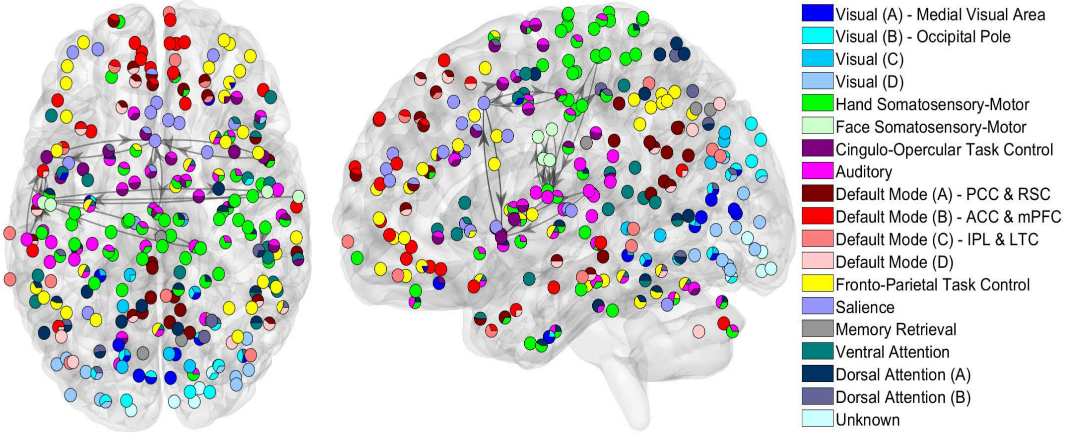
$$\text{overlap}(S_1, S_2) = \frac{|S_1 \cap S_2|}{\min(|S_1|, |S_2|)},$$

where S_1 and S_2 are two sets, e.g., 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 $\mathbb{S}_2$ with $\mathbb{S}_1$ 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.

6. 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 enables integration of many subjects'


Visual (A) - Medial Visual Area
Visual (B) - Occipital Pole
Visual (C)
Visual (D)
Hand Somatosensory-Motor
Face Somatosensory-Motor
Cingulo-Opercular Task Control
Auditory
Default Mode (A) - PCC & RSC
Default Mode (B) - ACC & mPFC
Default Mode (C) - IPL & LTC
Default Mode (D)
Fronto-Parietal Task Control
Salience
Memory Retrieval
Ventral Attention
Dorsal Attention (A)
Dorsal Attention (B)
Unknown

(a) Axial View

(b) Sagittal View

Figure 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.

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, Smith and Beckmann, 2010). This co-activation 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 the literature (Li et al., 2021).

We used the first-order MARSS instead of higher-order ones to identify directed directions 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 by Li et al. (2021) 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.

Acknowledgments. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Funding. Tingting Zhang gratefully acknowledges the support by NSF-2048991 and the computing support provided by the Pitt Center for Research Computing.

Dana L Tudorascu gratefully acknowledges the support by R01 AG063752.

SUPPLEMENTARY MATERIAL

The Variational Bayesian Algorithm

This supplementary file explains the optimization steps for implementing the proposed variational Bayesian algorithm.

Codes for Variational Bayesian Algorithm

This supplementary file contains MATLAB codes and the manual for using our toolbox to implement the proposed variational Bayesian algorithm.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9** 1981–2014.

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

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association* **112** 859–877.

BUCKNER, R. L., ANDREWS-HANNA, J. R. and SCHACTER, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease.

BUSHARA, K. O., GRAFMAN, J. and HALLETT, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience* **21** 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.

CHAIKIN, P. M., LUBENSKY, T. C. and WITTEN, T. A. (1995). *Principles of condensed matter physics* **10**.

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** 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* **4** 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** 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. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* **20** 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** 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** 1807–1814.

DURANTE, D. and DUNSON, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101** 883-898.

FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the american Statistical association* **80** 51-67.

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**.

FRISTON, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Humman Brain Mapping* **2** 56-78.

FRISTON, K. (2011). Functional and effective connectivity: a review. *Brain Connectivity* **1** 13-36.

FRISTON, K., HARRISON, L. and PENNY, W. (2003). Dynamic causal modelling. *NeuroImage* **19** 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.

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.

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

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** e5226.

HINRICHS, H., HEINZE, H. and SCHOENFELD, M. (2006). Causal visual interactions as revealed by an information theoretic measure and fMRI. *Neuroimage* **31** 1051-60.

HOFFMAN, M. and BLEI, D. (2015). Stochastic structured variational inference. In *Artificial Intelligence and Statistics* 361–369. PMLR.

HOLMES, E. E., WARD, E. J. and WILLS, K. (2012). MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data. *R journal* **4**.

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.

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. (2020). BVAR-Connect: A Variational Bayes Approach to Multi-Subject Vector Autoregressive Models for Inference on Brain Connectivity Networks. *Neuroinformatics* 1–18.

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.

KRAMER, M., KOLACZYK, E. and KIRSCH, H. (2008). Emergent network topology at seizure onset in humans. *Epilepsy Research* **79** 173-186.

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** 1637–1647.

LINDQUIST, M. (2008). The statistical analysis of fMRI data. *Statistical Science* **23** 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** 16.

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** 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** 1690–1701.

MESULAM, M.-M. (1998). From sensation to cognition. *Brain: a journal of neurology* **121** 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.

MOUSSA, M. N., STEEN, M. R., LAURIENTI, P. J. and HAYASAKA, S. (2012). Consistency of network modules in resting-state FMRI connectome data. *PloS one* **7** e44428.

NEWMAN, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* **103** 8577-8696.

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** 627-651.

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association* **96** 1077-1087.

PARK, H.-J. and FRISTON, K. (2013). Structural and Functional Brain Networks: From Connections to Cognition. *Science* **342**.

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.

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** 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** 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** 428–433.

ROSENTHAL, J. S. (2000). Parallel computing and Monte Carlo algorithms. *Far east journal of theoretical statistics* **4** 207–236.

SABESAN, S., GOOD, L., TSAKALIS, K., SPANIAS, A., TREIMAN, D. and IASEMIDIS, L. (2009). Information flow and application to epileptogenic focus localization from intracranial EEG. *IEEE Trans Neural Syst Rehabil Eng.* **17** 244-53.

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** 1444–1455.

SCHIFF, S., SAUER, T., KUMAR, R. and WEINSTEIN, S. (2005). Neuronal spatiotemporal pattern discrimination: The dynamical evolution of seizures. *Neuroimage* **28** 1043-1055.

SCHREIBER, T. (2000). Measuring Information Transfer. *Phys. Rev. Lett.* **85** 461.

SCHRÖDER, A. L. and OMBAO, H. (2018). FreSpeD: Frequency-specific change-point detection in epileptic seizure multi-channel EEG data. *Journal of the American Statistical Association* 1-14.

SEELEY, W. W. (2019). The salience network: a neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience* **39** 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** 2349–2356.

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** 13040–13045.

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.

SPORNS, O. (2011). *Networks of the Brain*. The MIT Press, Cambridge, Massachusetts.

SPORNS, O. (2013). Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology* **23** 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** e1049.

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** 165–178.

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** 1409-18.

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** 45–67.

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

WANG, Y., YAN, G., WANG, X., LI, S., PENG, L., TUDORASCU, D. L. and ZHANG, T. (2022). Supplement to "A Variational Bayesian Approach to Identifying Whole-Brain Directed Networks with fMRI Data". *The Annals of Applied Statistics*.

WILKE, C., WORRELL, G. and HE, B. (2011). Graph analysis of epileptogenic networks in human partial epilepsy. *Epilepsia* **52** 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** 892-900.

XIA, M., WANG, J. and HE, Y. (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one* **8** 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** 339–351.

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** 641–649.

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** 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. *The Annals of Applied Statistics* **11** 868-897.

ZHANG, T., SUN, Y., YAN, G., YIN, Q., LI, H., TANABE, S., CAFFO, B. and QUIGG, M. (2019). Bayesian inference of a directional brain network for intracranial EEG data. *Computational Statistics and Data Analysis* **106847**.