

Bayesian Low-rank Graphic Regression Models for Mapping Human Connectome Data

Eunjee Lee, Joseph G. Ibrahim, Yong Fan, Hongtu Zhu
and for the Alzheimer's Disease Neuroimaging Initiative* *

December 15, 2017

Abstract

We propose a Bayesian low-rank graphic regression model (BLGRM) framework for the regression analysis of matrix response data across subjects. This work is motivated by detailed comparisons of functional and structural connectivity data across subjects, groups, and time and relating neural connections to particular behavioral measures. The BLGRM can be regarded as a novel integration of principal component analysis, tensor decomposition, and regression models. In BLGRM, we find a common low-dimensional subspace for efficiently representing all matrix responses. Based on such low-dimensional representation, we can easily quantify the effects of various predictors of interest, such as age and diagnosis, and then perform regression analysis in the common subspace, leading to both substantial dimension reduction and much better prediction. We adapt a parameter expansion approach to our graphic regression model to address weak identifiability and high posterior dependence among parameters in our decomposition model. Posterior computation proceeds via an efficient Markov chain Monte Carlo algorithm. A simulation study is performed to evaluate the finite sample performance of BLGRM and to compare it with several competing approaches. We apply BLGRM to the resting state functional magnetic resonance imaging data set obtained from the Alzheimer's Disease Neuroimaging Initiative.

Keywords: Connection matrix; Covariate; Human Connectome; Low rank graph regression; Markov chain Monte Carlo.

*Address for correspondence and reprints: Hongtu Zhu, Ph.D., E-mail: hzhu5@mdanderson.org; Phone No: 346-814-0191. Eunjee Lee (Email: eunjee@umich.edu) was a Ph.D. student under the supervision of Drs. Ibrahim and Zhu in the Department of Statistics and Operational Research. Joseph G. Ibrahim is Alumni Distinguished Professor of Biostatistics (Email: ibrahim@bios.unc.edu) and Hongtu Zhu is Professor of Biostatistics, Department of Biostatistics, University of North Carolina, Chapel Hill, NC, 27599. Hongtu Zhu is also Professor of Biostatistics, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030. Yong Fan is Assistant Professor of Radiology, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104-6116 (Email: yong.fan@uphs.upenn.edu). This material was based on work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. This work was also partially supported by NIH grants MH086633, NSF grants SES-1357666 and DMS-1407655, and a grant from the Cancer Prevention Research Institute of Texas. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in this analysis or writing of this report. A complete list of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1 Introduction

The Alzheimer’s Disease Neuroimaging Initiative (ADNI) study (<http://www.adni-info.org/>), which represents a groundbreaking “big data” project for Alzheimer’s disease (AD), has collected imaging, genetic, clinical, and cognitive data from thousands of subjects since 2004. An important objective for ADNI is to quantify the clinical, cognitive, imaging, genetic and biochemical biomarker characteristics associated with the entire spectrum of AD as the pathology evolves from normal aging of the brain (NC), to mild cognitive impairment (MCI), to dementia or AD. This paper is motivated by the joint analysis of resting state functional magnetic resonance imaging (rfMRI) data and clinical and behavioral variables from $n = 153$ subjects in the ADNI study. After applying a standard preprocessing pipeline, we obtained a 116×116 rfMRI correlation matrix from each of the 153 subjects. We are particularly interested in addressing two objectives:

- (Obj1) the first is to derive functional-connectivity-based brain biomarkers for classifying AD, NC, and MCI groups;
- (Obj2) the second is to identify a common low-dimensional subspace that characterizes the major variations of the brain functional network across subjects and groups.

Statistically, these objectives can be formulated as the use of a vector of predictors (e.g., diagnosis), denoted as $\mathbf{x} = (x_1, \dots, x_p)^T$, to predict a $V \times V$ matrix response, denoted as $L = (L_{(g,g')})_{1 \leq g,g' \leq V}$, where g is a vertex and V is the total number of vertices. In this case, L is the rfMRI connectivity matrix and \mathbf{x} may include age, gender, and diagnosis (AD, NC, or MCI). The rfMRI data have been widely used in behavioral and cognitive neuroscience to understand functional segregation and integration of different brain regions in a single subject and across different populations [17, 26, 22, 18, 7, 45].

To achieve the objectives (Obj1) and (Obj2), we develop a framework for a Bayesian low-rank graphic regression model (BLGRM or Bayesian LGRM) to deal with three challenges arising from the use of \mathbf{x} to predict high-dimensional L . Such challenges include (i) complex spatial information, (ii) high-dimensional data, and (iii) the remarkable variability of brain functional connectivity across subjects and groups. It is commonly believed that the organization of brain networks is governed by both short- and long-range connections among different brain regions. Moreover, in most neuroimaging studies, the dimension of the functional and structural connectivity data (or L) can be much larger than the number of subjects, which varies from several dozens to a few thousands. Specifically, for the ADNI rfMRI data set, we have $n = 153$, and each connectivity matrix L contains $V(V - 1)/2 = 6670$ unique elements for $V = 116$. There is considerable inter-subject

spatial variability due to natural variability, striking neuroanatomical variations, different conditions, and different subject groups.

In the current literature, there are two major approaches to the group analysis of functional connectivity data, including univariate methods and graphic theoretic methods [44, 13, 42]. The graphic theoretic methods consist of calculating graphic theoretic summaries (i.e., girth, diameter, modularity, small-worldness) of the entire graph for each subject and fitting linear (or nonlinear) regression models with these summaries as responses [8, 44, 29, 18, 15]. A key limitation of graphic theoretic methods is that they cannot reveal subtle differences at both the sub-network and nodal levels. The most popular univariate approaches involve fitting a regression model, such as linear regression, to association measures (e.g., correlation and partial correlation) from all subjects at each edge, and then generating a statistical network map of test statistics and p -values across all edges [44, 54]. One fundamental issue for such univariate methods is to correct for multiple comparisons due to the large number of network edges when V is relatively large. Therefore, these univariate methods often suffer from low statistical power associated with detecting weak to moderate signals [28], even though there are some network analogues of cluster-based thresholding methods, such as the network-based statistic and spatial pairwise clustering methods [44, 54]. So, alternative methods of connectivity analysis are critically needed for understanding the complex organization of brain networks, while achieving dimension reduction.

There is great interest in developing statistical models that explicitly delineate the conditional distribution of L given \mathbf{x} , that is, $p(L|\mathbf{x})$. Ideally, such a model should account for the complex topological structure of networks, while flexibly assessing the effects of multiple variables of interest and local network features. For a single network, popular models, including exponential random graphic models, stochastic block models, and latent space models, have been developed primarily for binary networks, whereas their extensions to weighted networks are in their infancy [43, 12]. Little has been done relating weighted networks with various covariates of interest, such as disease status and time, due to additional computational and methodological challenges.

The aim of this paper is to propose a BLGRM with matrix responses and clinical covariates. Our BLGRM can be regarded as a novel extension of the hierarchical eigenmodel for pooled covariance matrices [25], tensor decomposition [30], and regression models. The key idea of BLGRM is to find an intrinsic low-dimensional subspace for all subjects, denoted as B . Such B not only allows us to dramatically reduce the dimension, but also characterizes the organization of brain connection maps across subjects at a systemic level. We further represent each graphic response by using a $R \times R$ subject-specific coefficient matrix, denoted

Λ_i , while preserving an individual network structure of the low-dimensional eigenspace. One more intriguing part is that we introduce a hierarchical structure of Λ_i in order to incorporate the effects of clinical/demographic covariates on graphic responses. Due to weak identifiability and high posterior dependence among the parameters of our decomposition model, BLGRM suffers from poor mixing and slow convergence of the Markov chain Monte Carlo (MCMC) sampler. To address these issues, we adapt a parameter expansion approach to our graphic regression model.

Based on this formulation, an efficient MCMC algorithm is used to perform posterior computation. We take a Bayesian approach to estimate the parameters involved in B and the regression coefficients associated with the covariates of interest. We examine the effects of the covariates of interest on graph responses by utilizing the highest posterior density (HPD) intervals. We have developed the BLGRM package using MATLAB and will release it through the website <https://www.nitrc.org/> and our group Github website.

The rest of this paper is organized as follows. In Section 2, we introduce BLGRM in detail and present its Bayesian estimation procedure. In Section 3, we present simulation results to evaluate the finite sample performance of the estimation procedure. In Section 4, we apply BLGRM to the ADNI data set we introduced to address (Obj1)-(Obj2). We present concluding remarks in Section 5.

2 Bayesian Low-rank Graph Regression Models

2.1 Model Specification

We consider network data from n independent subjects in ADNI. For the i -th subject, we observe a $p \times 1$ vector of predictors, denoted as \mathbf{x}_i , and a graph, denoted as $L_i = (L_{i(g,g')})_{g,g' \leq V}$, corresponding to V vertices. Without loss of generality, it is assumed that L_i is symmetric, that is, $L_{i(g,g')} = L_{i(g',g)}$ holds for all $g, g' \leq V$.

Our BLGRM consists of two key components: a common component model and a regression model. The *common component model* is given by

$$L_{i(g,g')} = \sum_{r=1}^R \sum_{s=1}^R \beta_{r,g} \lambda_{i(r,s)} \beta_{s,g'} + \epsilon_{i(g,g')}, \quad (1)$$

where $\epsilon_{i(g,g')}$ are measurement errors, and $\lambda_{i(r,s)}$ are subject-specific coefficients that can be non-zero even for $r \neq s$. Moreover, $B = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_R]$ as a $V \times R$ orthogonal matrix is a common eigenmap across all subjects, where $\boldsymbol{\beta}_s = (\beta_{s,g})_{g \leq V}$ is an orthonormal basis for $s = 1, \dots, R$. The subject-specific matrix $\Lambda_i = [\lambda_{i(r,s)}]_{r,s=1,\dots,R}$ preserves an intrinsic network structure in the low-dimensional space spanned by the columns of B . Equation

(1) can be written in a matrix form as follows:

$$L_i = B\Lambda_i B^T + \boldsymbol{\epsilon}_i, \quad (2)$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i(g,g')})$. For the symmetric graph, our *regression model* assumes that

$$\lambda_{i(r,s)} = \lambda_{i(s,r)} = f_{r,s}(\mathbf{x}_i) + \delta_{i,(r,s)}, \quad (3)$$

where $\delta_{i,(r,s)}$ are measurement errors and $f_{r,s}(\mathbf{x}_i)$ is a nonparametric or parametric function of \mathbf{x}_i . A simple example is to set $f_{r,s}(\mathbf{x}_i) = \mathbf{x}_i^T \gamma_{r,s}$. In general, one may choose a set of basis functions, denoted as $\{\phi_k(\mathbf{x}_i)\}_{k \leq K}$, and then approximate $f_{r,s}(\mathbf{x}_i)$ by $\sum_{k=1}^K \phi_k(\mathbf{x}_i) \gamma_{k(r,s)}$.

Our BLGRM based on (2) and (3) has at least four unique features. First, it achieves substantial dimension reduction by reducing from $V(V - 1)/2$ to around $VR + R(R + 1)/2$ for each subject to deal with the high-dimensionality of the connectivity matrices. Second, the R eigenmaps $\beta_r = \{\beta_{r,g} : g \in \mathcal{G}\}$ can be considered as independent networks of vertices that characterize the latent organization of the connectivity structures across subjects at a systemic level. Then, the subject-specific coefficient matrix Λ_i preserves an individual network structure in the low-dimensional space spanned by the eigenmaps. This enables us to construct an underlying relational structure among brain areas and to reduce the heterogeneity of the graph structure in the low-dimensional space. Third, $\eta_{i,r} = \{\eta_{i,r}(g) = \sum_{s=1}^R \beta_{s,g} \lambda_{i,(s,r)} : g \in \mathcal{G}\}$ delineate the individual organization of connection maps among vertices across R different latent networks. One more intriguing part is that we assume a hierarchical structure of Λ_i to delineate the effect of clinical/demographic covariates on the graph. Our method facilitates the estimation of the effects of covariates on Λ_i , while reducing the dimension of the graph. For instance, for ADNI, we can examine local differences in functional connectivity across diagnostic groups by including diagnostic groups as covariates of interest. Moreover, the hierarchical structure allows the decomposition to be supervised by the covariates of interest. Thus, the estimated eigenmap B explains variations in the graph across subjects, while accounting for the interrelationship between covariates and the graph.

2.2 Sampling Distributions

For model (1), it is assumed that $\epsilon_{i(g,g')}$ s for $g \leq g'$ are identically and independently distributed as $N(0, \sigma^2)$. For a given R , the likelihood function of (L_1, \dots, L_n) is given by

$$p(L_1, \dots, L_n | B, \Lambda_1, \dots, \Lambda_n) \propto \prod_{i=1}^n \exp \left[-\frac{1}{2\sigma^2} \text{tr} \left\{ (L_i - B\Lambda_i B^T)^2 \right\} \right].$$

It is assumed that the eigenmap B is a lower triangular matrix in order to ensure the identifiability of the decomposition model (2) [20]. Moreover, R should be chosen such that $VR - R(R - 1)/2 + nR(R + 1)/2$ is smaller than or equal to $nV(V + 1)/2$.

For model (3), it is assumed that $\delta_{i,(r,s)}$'s for $r \leq s$ are identically and independently distributed as $N(0, \sigma_0^2)$. The likelihood function of $(\Lambda_1, \dots, \Lambda_n)$ is given by

$$\pi(\Lambda_1, \dots, \Lambda_n | \Delta_1, \dots, \Delta_n, \sigma_0^2) \propto \prod_{i=1}^n \exp \left[-\frac{1}{2\sigma_0^2} \text{tr} \{ (\Lambda_i - \Delta_i)^2 \} \right],$$

where $\text{vech}(\Delta_i) = \Gamma^T \mathbf{x}_i = [\mathbf{x}_i^T \gamma_1, \mathbf{x}_i^T \gamma_2, \dots, \mathbf{x}_i^T \gamma_q]^T$, in which $q = R(R+1)/2$ and $\text{vech}(\cdot)$ is the half-vectorization of a symmetric matrix. For better identification of parameters, we fix σ_0 , the scale of Λ_i , at 1. Moreover, $\Gamma = [\gamma_1, \dots, \gamma_q]$ is a $p \times q$ matrix, where $\gamma_j = (\gamma_{1j}, \dots, \gamma_{pj})^T$. Throughout the paper, it is assumed that the first element of \mathbf{x}_i is equal to 1 for all i .

We may take a Bayesian approach to estimate all the parameters of interest. We place the following priors on all parameters, including B and $\{\Lambda_1, \dots, \Lambda_n\}$. The full posterior distribution is proportional to

$$\pi(\Lambda, B, \sigma^2, \sigma_0^2 | \mathbf{L}; \mathbf{X}) \propto p(\mathbf{L} | \Lambda, B, \sigma^2) \pi(\Lambda | \Gamma, \sigma_0^2; \mathbf{X}) \pi(\Gamma | \sigma_\gamma^2) \pi(B) \pi(\sigma^2) \pi(\sigma_0^2) \pi(\sigma_\gamma^2).$$

However, model (1) suffers from poor mixing and slow convergence of the MCMC due to the weak identifiability and high posterior dependence of B and the Λ_i 's.

2.3 Parameter-Expanded Common Component Model

We develop a parameter expansion method to model (1) in order to reduce the posterior dependence between B and the Λ_i 's. Such a parameter expansion method not only increases the computational efficiency of the expectation-maximization algorithm [34], but also improves the convergence and mixing of MCMC chains [35, 19, 21]. For instance, for Bayesian factor analysis, Ghosh and Dunson [21] proposed a parameter expansion method to reduce the posterior dependence of parameters, leading to fast mixing of MCMC chains.

The parameter-expanded version of model (1) is given by

$$\begin{aligned} L_i &= B^* \Lambda_i^* B^{*T} + \epsilon_i, \\ B &= s(B) \circ B^* \Psi^{-1/2} \quad \text{and} \quad \Lambda_i = d(B) \Psi^{1/2} \Lambda_i^* \Psi^{1/2} d(B), \end{aligned} \tag{4}$$

where $s(B) = \mathbf{1}_V (\text{sign}(\beta_{11}), \dots, \text{sign}(\beta_{RR}))$ is a $V \times R$ matrix, in which $\mathbf{1}_V$ is a $V \times 1$ vector of ones, $\Psi = \text{diag}(\psi_1, \dots, \psi_R)$, and $d(B) = \text{diag}(\text{sign}(\beta_{11}), \dots, \text{sign}(\beta_{RR}))$. The sign of the diagonal elements in B is multiplied accordingly for the purpose of identifiability. Subsequently, we define $\Gamma = \Gamma^* (d(B) \Psi^{1/2}) \otimes_s (d(B) \Psi^{1/2})$, where \otimes_s denotes the symmetric Kronecker product introduced by [1]. Let K be a linear operator on \mathcal{S}^n defined by the map

$$K \mapsto \frac{1}{2} (AKB^T + BK A^T),$$

where $A, B \in \mathbb{R}^{n \times n}$, and \mathcal{S}^n is the vector space of real symmetric $n \times n$ matrices. The symmetric kronecker product of A and B is defined by the identity

$$(A \otimes_s B)\text{svec}(K) = \text{svec}\left(\frac{1}{2}(AKB^T + BKA^T)\right),$$

where svec maps \mathcal{S}^n to $\mathbb{R}^{\frac{1}{2}n(n+1)}$ by

$$\text{svec}(K) = \left[K_{11}, \sqrt{2}K_{12}, \dots, K_{11}, \sqrt{2}K_{12}, \dots, K_{22}, \dots, \sqrt{2}K_{2n}, \dots, K_{nn} \right]^T.$$

Therefore, the sampling distribution of $(\Lambda_1^*, \dots, \Lambda_n^*)$ is given by

$$\pi(\Lambda_1^*, \dots, \Lambda_n^* | \Delta_1^*, \dots, \Delta_n^*, \Psi) \propto \prod_{i=1}^n \exp\left(-0.5\text{tr}(\{(\Lambda_i^* - \Delta_i^*)\Psi\}^2)\right).$$

where $\text{vech}(\Delta_i^*) = \Gamma^{*T} \mathbf{x}_i = (\mathbf{x}_i^T \boldsymbol{\gamma}_1^*, \mathbf{x}_i^T \boldsymbol{\gamma}_2^*, \dots, \mathbf{x}_i^T \boldsymbol{\gamma}_q^*)^T$.

2.4 Priors

For lower triangular elements of B^* , we adapt the Bayesian lasso prior [40] for each column in order to induce sparse loadings for each eigenvector. We also impose conjugate priors on Γ^* , ψ_i s, and σ^2 . Specifically, these priors are given as follows:

$$\begin{aligned} \beta_{r,g}^* &\sim N(0, \tau_r^{-1} \phi_{rg}^{-1}), \quad 1 \leq g \leq V, \quad 1 \leq r \leq \min(g, R) \\ \phi_{rg} &\sim \text{Gamma}(1, a_0/2), \quad a_0 \sim \text{Gamma}(a_1, a_2), \quad \pi(\tau_r) = 1/\tau_r, \\ \pi(\Gamma^* | \sigma_\gamma^2) &\propto \exp\left[-\frac{1}{2\sigma_\gamma^2} \text{tr}(\Gamma^{*T} \Gamma^*)\right], \quad \sigma^2 \sim \text{IG}(b_1, b_2), \quad \sigma_\gamma^2 \sim \text{IG}(c_1, c_2), \\ \psi_i &\sim \text{Gamma}(\nu_a, \nu_a), \quad i = 1, \dots, R, \end{aligned}$$

where σ_γ^2 is the dispersion parameter, IG denotes the inverse gamma distribution, and ν_a , a_1, a_2 , b_1, b_2 , c_1 , and c_2 are hyper-parameters.

2.5 Posterior Computation

We propose a straightforward Gibbs sampler for posterior computation based on the full conditional distributions of all parameters as follows. The full conditional distribution of B^* is proportional to

$$\exp\left(-0.5\sigma^{-2} \sum_{i=1}^n \text{tr}((L_i - B^* \Lambda_i^* B^{*T})^2) - 0.5 \sum_{r=1}^R \sum_{g=r}^V (\beta_{r,g}^*)^2 \tau_r \phi_{rg}\right). \quad (5)$$

We apply the slice sampling approach to draw random samples from $\pi(\beta_g^* | B_{(-g)}^*, \bullet)$ in (5), where $B_{(-g)}^*$ denotes the matrix B^* with the g -th column deleted.

The full conditional distribution of $\text{vech}(\Lambda_i^*)$ is given by

$$(\text{vech}(\Lambda_i^*) | \bullet) \sim N(\boldsymbol{\mu}_i, \mathbf{S}), \quad (6)$$

where $\boldsymbol{\mu}_i = \mathbf{S}\mathbf{D} \times \text{svec}(\mathbf{P}_i)$ and $\mathbf{S}^{-1} = \mathbf{D}(\mathbf{Q} \otimes_s \mathbf{Q})\mathbf{D}/\sigma^2 + \mathbf{D}(\Psi \otimes_s \Psi)\mathbf{D}$, in which

$$\begin{aligned}\mathbf{Q} &= B^{*T}B^*, \quad \mathbf{P}_i = \sigma^{-2}B^{*T}L_iB^* + \Psi\Delta_i^*\Psi, \\ \mathbf{D} &= \text{diag}(\underbrace{1, \sqrt{2}, \dots, \sqrt{2}}_{R \text{ elements}}, \underbrace{1, \sqrt{2}, \dots, \sqrt{2}}_{(R-1) \text{ elements}}, \underbrace{1}_{1 \text{ element}}).\end{aligned}$$

The full conditional distribution of $\text{vec}(\boldsymbol{\Gamma}^*)$ is given by

$$\text{vec}(\boldsymbol{\Gamma}^*)|\bullet \sim N(\boldsymbol{\mu}_\gamma, \mathbf{S}_\gamma), \quad (7)$$

where $\boldsymbol{\mu}_\gamma = \mathbf{S}_\gamma \sum_{i=1}^n (\mathbf{x}_i^T \text{vech}(\Psi\Lambda_i^*\Psi)^T)$ and $\mathbf{S}_\gamma^{-1} = \sum_{i=1}^n \{(\mathbf{D}(\Psi \otimes_s \Psi)\mathbf{D}) \otimes \mathbf{x}_i\mathbf{x}_i^T\} + \sigma_\gamma^{-2}\mathbf{I}$.

The full conditional distributions of all other parameters are respectively given by

$$\phi_{rg}|\bullet \sim \text{Inverse-Gaussian}\left(\sqrt{a_{0r}/(\tau_r\beta_{rg}^{*2})}, a_{0r}\right), \quad (8)$$

$$\tau_r|\bullet \sim \text{Gamma}\left((V-r+1)/2, \sum_{g=r}^V \phi_{rg}\beta_{r,g}^{*2}/2\right), \quad (9)$$

$$a_{0r}|\bullet \sim \text{Gamma}\left(a_1 + (V-r+1), a_2 + \sum_{g=r}^V (2\phi_{rg})^{-1}\right), \quad (10)$$

$$\sigma^2|\bullet \sim IG\left(b_1 + 0.25nV(V+1), 0.5\text{tr}\left(\sum_{i=1}^n (L_i - B^*\Lambda_i^*B^{*T})^2\right) + b_2\right), \quad (11)$$

$$\sigma_\gamma^2|\bullet \sim IG(c_0 + 0.25pR(R+1), 0.5\text{tr}(\boldsymbol{\Gamma}^{*\prime}\boldsymbol{\Gamma}^*) + c_1). \quad (12)$$

The full conditional distribution of (ψ_1, \dots, ψ_R) is proportional to

$$p(\psi_1, \dots, \psi_R|\bullet) \propto p(\Psi)p(\Lambda_1^* \dots, \Lambda_n^*|\Psi). \quad (13)$$

We apply the slice sampling approach to draw random samples from the full conditional distribution of (ψ_1, \dots, ψ_R) .

2.6 Determining the rank of B

We use the Bayesian information criterion (BIC) for determining the number of eigenvectors R . The BIC is given by

$$\text{BIC} = -2p(L_1, \dots, L_n|\tilde{B}, \tilde{\Lambda}_1, \dots, \tilde{\Lambda}_n) + \log(n) \times (VR + nR(R+1)/2), \quad (14)$$

where $\tilde{B}, \tilde{\Lambda}_1, \dots, \tilde{\Lambda}_n$ are estimated by using the frequentist version of LGRM. The detailed algorithm is given in the Appendix. We also calculate the reconstruction error, which is measured as follows:

$$\text{error} = \frac{1}{n} \sum_{i=1}^n \frac{\|L_i - \hat{B}\hat{\Lambda}_i\hat{B}^T\|_F}{\|L_i\|_F}. \quad (15)$$

3 Simulation study

We conducted simulations to examine the finite-sample performance of the parameter-expanded BLGRM and its posterior computation. We considered two main scenarios to generate response variables L_i from model (2). The first scenario assumes only the decomposition model (2) without the regression part for Λ_i given in (3). The second scenario assumes both decomposition and regression models. For each scenario, we considered 4 different setups by varying the sample size, the vertex number, and the true rank of B . We applied our proposed method with 5,500 MCMC iterations and 500 burn-in iterations to estimate $B, \Lambda_1, \dots, \Lambda_n$ and then to recover L_1, \dots, L_n for both scenarios. For the second scenario, we also considered the estimation performance of Γ . In all eight simulation setups, we computed the reconstruction errors defined in (15) for each simulated data set.

For comparison, we considered three competing methods: a frequentist version of BLGRM, the so-called LGRM, CANDECOMP/PARAFAC (CP) decomposition [27], and three-way DEDICOM [24]. Among them, both CP decomposition and three-way DEDICOM decomposed multiple (symmetric) matrices into a common basis and individual coefficient matrices, but imposed more constraints than BLGRM and LGRM. More detailed information regarding CP and three-way DEDICOM can be found in the supplementary document. We used BIC defined in (14) to select the number of common eigenvectors for both BLGRM and LGRM. We used the core consistency diagnostic to choose the proper number of components for CP [5]. For the three-way DEDICOM, we predefined the number of components as either R or $2R$, since there is no good way to determine the number of components.

3.1 Scenario 1

We generated L_i according to model (20), in which we simulated B , ϵ_i and Λ_i as described below. Specifically, we generated them as follows:

$$\begin{aligned} \beta_{lk}^g &\sim N(0, 1), \quad l = 1, \dots, V, \quad k = 1, \dots, R, \\ \pi(\epsilon_1, \dots, \epsilon_n) &= \prod_{i=1}^n \exp\left(-\frac{1}{2}\text{tr}(\epsilon_i^2)\right), \quad \pi(\Lambda_1, \dots, \Lambda_n) = \prod_{i=1}^n \exp\left(-\frac{1}{2}\text{tr}(\Lambda_i^2)\right). \end{aligned} \tag{16}$$

In this scenario, we considered 4 different simulation setups: (a) $(V, R, n) = (50, 3, 50)$; (b) $(V, R, n) = (50, 3, 100)$; (c) $(V, R, n) = (100, 6, 100)$; and (d) $(V, R, n) = (100, 6, 200)$. For each setup, we generated 50 data sets.

For all four estimation methods, we calculated their reconstruction errors across different numbers of eigenvectors. Figure 1 shows the reconstruction errors that correspond to setup (a) as R increases. Moreover, the BIC for BLGRM and LGRM selects 3 in all 50

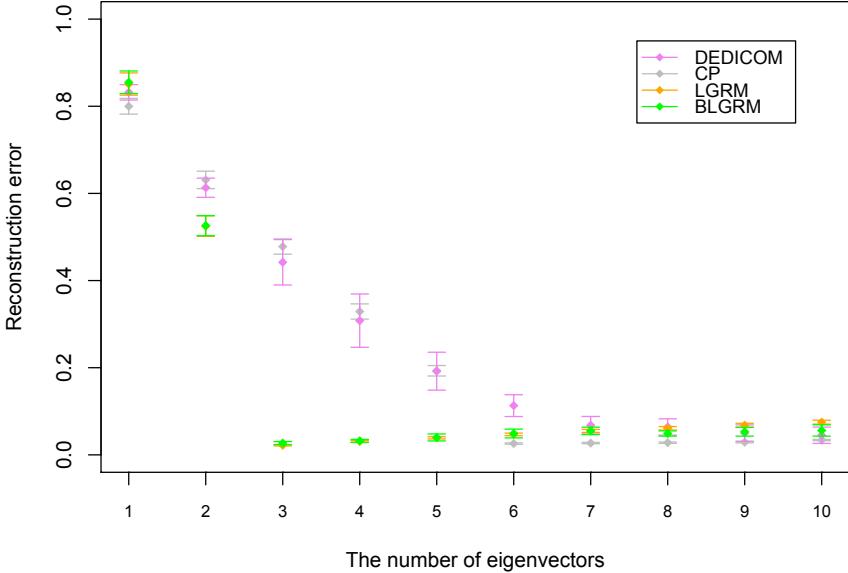


Figure 1: Reconstruction errors of the four decomposition models corresponding to setup (a) as R increases. Each dot and bar respectively represent the mean and Monte Carlo standard deviation of reconstruction errors based on the 50 simulations.

replications, whereas the core consistency diagnostic for the CP decomposition selects different optimal components with the true rank 3 as the median. Both BLGRM and LGRM have small reconstruction errors with the true rank, whereas their reconstruction errors slightly increase as R is greater than 3. Panels (a) and (b) of Figure 2 demonstrate that BLGRM leads to a good approximation of a randomly selected matrix. The CP decomposition model performs reasonably well when the number of components is greater than 5, whereas its core consistency diagnostic statistic often selects 3 as the optimal component. As expected, the reconstruction error of three-way DEDICOM gradually decreases as the number of components increases. Moreover, the variability of DEDICOM is relatively high compared with that of all the other competing methods.

Table 1 presents the reconstruction errors of all four estimation methods under the four setups. We use DEDICOM_R and DEDICOM_{2R} to represent DEDICOM with R and $2R$ factors, respectively, for each setup. Both LGRM and BLGRM outperform the other two competing methods. For setups (a) and (b), both LGRM and BLGRM perform similarly in terms of the approximation error, whereas BLGRM is more variable than LGRM. This may occur because BLGRM requires tuning the hyperparameters and the number of MCMC and slice sampling iterations. As expected, the approximation errors of LGRM and BLGRM improve as the sample size increases. CP is much more variable than the other

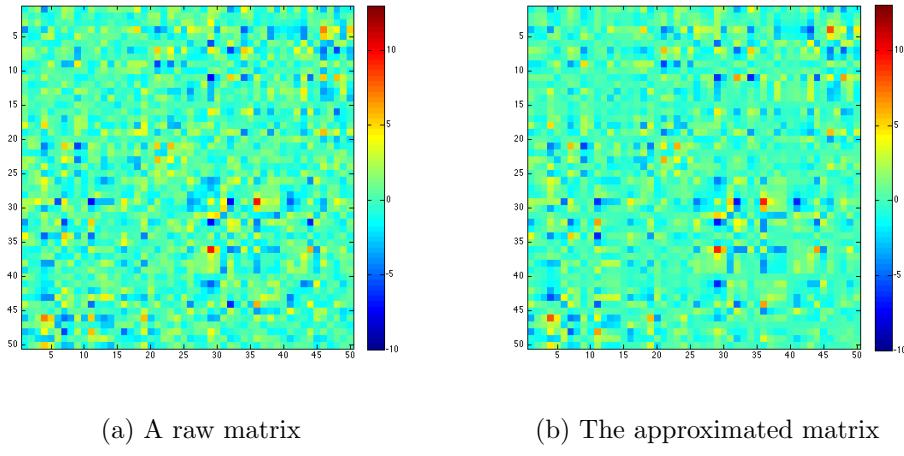


Figure 2: Panel (a) shows a randomly chosen raw data matrix \mathbf{L}_1 in the first simulation data set and panel (b) shows the matrix approximated by the proposed method.

methods, largely due to the inconsistency of its component selection method. In addition, DEDICOM_{2R} is much worse than LGRM and BLGRM, even with 2R factors.

3.2 Scenario 2

In this scenario, we examine the reconstruction error as well as the estimation accuracy of regression coefficients in BLGRM. We generated L_i according to model (20), in which we simulated B , $\boldsymbol{\epsilon}_i$ and Λ_i as described below. Specifically, we generated $\boldsymbol{\epsilon}_i$, with $B = (\beta_{lk})$ being a $V \times R$ matrix and Λ_i being a $R \times R$ matrix as follows:

$$\beta_{lk}^g \sim N(0, 1), \quad l = 1, \dots, V; \quad k = 1, \dots, R, \quad (17)$$

$$\pi(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n) = \prod_{i=1}^n \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{\epsilon}_i^2)\right), \quad (18)$$

$$\pi(\Lambda_1, \dots, \Lambda_n) = \prod_{i=1}^n \exp\left[-\frac{1}{2}\text{tr}((\Lambda_i - \Delta_i)^2)\right], \quad (19)$$

where $\text{vech}(\Delta_i) = \boldsymbol{\Gamma}^T \mathbf{x}_i = [\text{vech}(\boldsymbol{\Gamma}_0) \text{vech}(\boldsymbol{\Gamma}_1)]^T \mathbf{x}_i$. We assumed that $\mathbf{x}_i = [1, x_{1i}]$ contains an intercept and a continuous covariate with $x_{1i} \sim N(0.5, 1)$. In this scenario, we considered 4 different simulation setups: (a) $(V, R, n) = (50, 3, 50)$; (b) $(V, R, n) = (50, 3, 100)$; (c) $(V, R, n) = (100, 6, 100)$; and (d) $(V, R, n) = (100, 6, 200)$. We set $\boldsymbol{\Gamma}_0 = \mathbf{1}_3 \mathbf{1}_3^T$ and $\boldsymbol{\Gamma}_1 = [0, 4, 0; 4, 0, 4; 0, 4, 0]$ for setups (a) and (b), whereas we set $\boldsymbol{\Gamma}_0 = \mathbf{1}_6 \mathbf{1}_6^T$ and $\boldsymbol{\Gamma}_1 = [0, 4, 0; 4, 0, 4; 0, 4, 0] \otimes [1, 0; 0, 0]$ for setups (c) and (d).

For each setup, we generated 50 data sets. For each simulated data set, we calculated the estimation error of the BLGRM estimate $\widehat{\boldsymbol{\Gamma}}_j$ given by $\text{error}_{\gamma_j} = \|B\boldsymbol{\Gamma}_j B^T - \widehat{B}\widehat{\boldsymbol{\Gamma}}_j \widehat{B}^T\|_F / \|B\boldsymbol{\Gamma}_j B^T\|_F$ for $j = 0$ and 1.

Table 1: Scenario 1: Mean (Mn) and Monte Carlo standard deviation (SD) of reconstruction errors of all five estimation methods across four different values of (V, R, n) s based on 50 replications in setups (a)-(d). DEDICOM $_R$ and DEDICOM $_{2R}$ respectively represent DEDICOM with R and $2R$ factors for each setup.

Setup	(a) (b) (c) (d)							
	(V, R, n)							
	(50, 3, 50)		(50, 3, 100)		(100, 6, 100)		(100, 6, 200)	
Methods	Mn	SD	Mn	SD	Mn	SD	Mn	SD
CP	0.533	0.058	0.534	0.047	0.742	0.034	0.740	0.026
DEDICOM $_R$	0.442	0.044	0.463	0.038	0.649	0.029	0.657	0.024
DEDICOM $_{2R}$	0.113	0.037	0.115	0.031	0.393	0.029	0.412	0.028
LGRM	0.022	0.003	0.020	0.001	0.010	0.001	0.010	0.001
BLGRM	0.027	0.007	0.026	0.007	0.065	0.018	0.063	0.017

Figure 4 shows rapid and nice mixing and convergence of the MCMC chains for 10 randomly chosen elements in the B matrix under setup (b). Panels (a) and (b) of Figure 3 show that BLGRM provides a good approximation of the raw matrices.

Table 2 presents the reconstruction errors of all the estimation methods and the estimation errors of $\widehat{\Gamma}_j$ across different setups. LGRM performs slightly better than BLGRM in all setups in terms of approximation accuracy, but BLGRM can automatically estimate the effect of the covariates on Λ_i . As expected, the estimation accuracy of $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ and the approximation accuracy of LGRM and BLGRM improve as the sample size increases. Similar to Scenario 1, CP coupled with core consistency diagnostics is much more variable than the other methods, while the approximation accuracy of DEDICOM $_{2R}$ is much worse than those of LGRM and BLGRM.

Table 2: Scenario 2: Mean (Mn) and Monte Carlo standard deviation (SD) of reconstruction errors of all five estimation methods and estimation errors of $\widehat{\Gamma}_j$ across four different values of (V, R, n) s based on 50 replications in setups (a)-(d). DEDICOM _{R} and DEDICOM _{$2R$} respectively represent DEDICOM with R and $2R$ factors for each setup. Estimation accuracy for Γ_j is presented only for BLGRM.

Setup	(a)		(b)		(c)		(d)	
	(V, R, n)							
	(50, 3, 50)		(50, 3, 100)		(100, 6, 100)		(100, 6, 200)	
Methods	Mn	SD	Mn	SD	Mn	SD	Mn	SD
CP	0.273	0.044	0.263	0.033	0.518	0.061	0.486	0.074
DEDICOM _{R}	0.278	0.072	0.283	0.060	0.374	0.045	0.377	0.029
DEDICOM _{$2R$}	0.131	0.040	0.157	0.032	0.290	0.026	0.300	0.027
LGRM	0.009	0.001	0.008	0.001	0.005	0.000	0.005	0.000
BLGRM	0.041	0.021	0.040	0.017	0.057	0.017	0.057	0.017
error _{γ_0}	0.120	0.037	0.083	0.036	0.088	0.021	0.071	0.019
error _{γ_1}	0.046	0.012	0.034	0.011	0.063	0.013	0.045	0.011

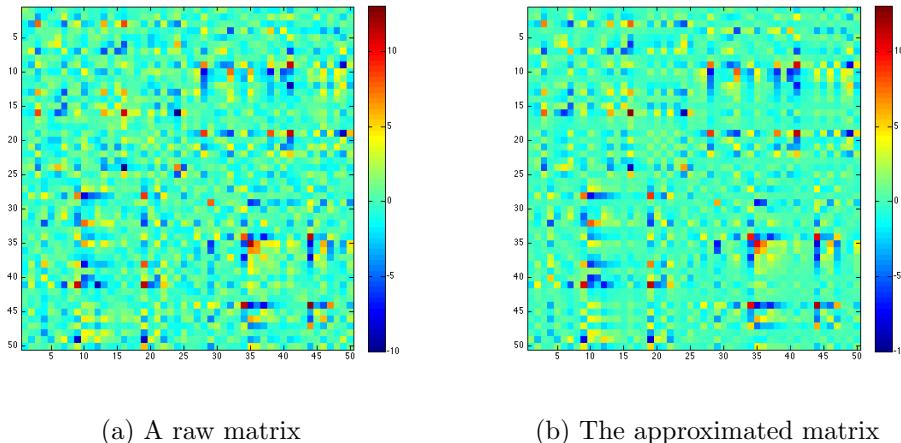


Figure 3: Panel (a) shows a randomly chosen raw data matrix \mathbf{L}_1 in the first simulation data set and panel (b) shows the matrix approximated by the proposed method.

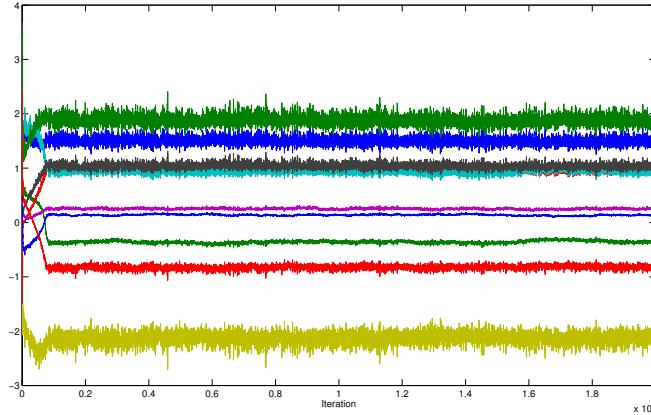


Figure 4: Trace plot for posterior samples of B elements, which shows rapid and nice mixing and convergence of MCMC chains.

4 The Alzheimer’s Disease Neuroimaging Initiative Data Analysis

4.1 Alzheimer’s Disease Neuroimaging Initiative (ADNI)

The development of the BLGRM is motivated by the analysis of rfMRI and clinical data collected by ADNI. “Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, M.D. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). For up-to-date information, see www.adni-info.org.”

Table 3: The demographic information of the 153 ADNI subjects at the baseline of ADNI. The mean and standard deviation (mean \pm sd) of the subjects ages are presented. For gender, the count and the percentage (in parentheses) of male subjects are shown.

	Total (N=153)	NC (N=54)	MCI (N=75)	AD (N=24)
Gender (male)	75 (49.02%)	25 (46.29%)	38 (50.67%)	12 (50.00%)
Age (in years)	72.44 ± 6.62	72.92 ± 6.06	71.78 ± 6.81	73.40 ± 7.27

We focused on 153 subjects from ADNI-1, ADNI-GO, and ADNI-2. There were four

baseline diagnostic categories: normal aging/cognitively normal (CN), significant memory concern (SMC), MCI, and AD. The ADNI reported that SMC subjects exhibited slight forgetfulness, but had cognitive scores within normal range. As their cognition was normal and forgetfulness was not consistent, SMC subjects were combined with CN subjects. Hereafter, we call this combined group “normal control (NC).” Among them, there were 75 male and 78 female subjects and the average age was 72.44 years, with a standard deviation of 6.62 years. There were 25 NC male subjects and the average age of the NC subjects was 72.92 years, with a standard deviation of 6.06 years. Among 75 MCI patients, 38 patients were male. The average age of the MCI subjects was 71.78 years, with a standard deviation of 6.81 years. For AD patients, 12 were male patients, while their average age was 73.40 years, with a standard deviation of 7.27 years. The demographic information at baseline is summarized in Table 3.

Altered brain functional connectivity has been considered an important factor in the process of cognitive decline in patients with AD [10, 11]. A seed-based approach that uses Fishers z-transformation and t-tests has indicated that patients with mild AD show abnormal functional connectivity in regions of the brain that include the medial prefrontal cortex (MPFC), ventral anterior cingulate cortex (vACC), right inferotemporal cortex, right cuneus extending into the precuneus, left cuneus, right superior and middle temporal gyri and posterior cingulate cortex (PCG or PCC) [49]. Wang et al. [48] used t-tests based on whole-brain regions of interest (ROIs) and showed that AD patients have decreased connectivity between the frontal and parietal lobes and increased within-lobe functional connectivity.

4.2 Resting State fMRI Data Acquisition and Pre-processing

The imaging protocol of rfMRI is given as follows: field strength=3.0 tesla; flip angle=80.0 degrees; manufacturer is Philips Medical Systems; matrix X=64.0 pixels; matrix Y=64.0 pixels; manufacturing model=Intera; pixel spacing X=3.3125 mm; pixel spacing Y=3.3125 mm; pulse sequence=GR; slices=6720.0; slice thickness=3.313 mm; echo time (TE)=30.001 ms; repetition time (TR)=3000.0 ms; and number of anatomical volumes=140.

The resting state fMRI data were pre-processed with the following steps: (i) discarding the first 10 time points, (ii) slice timing, (iii) head motion correction, (iv) intensity scaling of each fMRI scan after motion correction to yield a whole-brain mean value of 10000, (v) temporally band-pass filtering (0.01 Hz-0.08 Hz), (vi) regressing out a set of nuisance signals, including the signal averaged over the white matter, signal averaged over the cerebrospinal fluid, global signal averaged over the whole brain, and six motion parameters,

(vii) nonlinear normalization into the Montreal Neurological Institute space with resolution $3 \times 3 \times 3 \text{ mm}^3$ using SPM8; and (viii) spatially smoothing with a 6-mm full width at half maximum Gaussian kernel. The nonlinear normalization of fMRI data was implemented using DARTEL of SPM8 with the deformation fields of their co-registered T1-weighted images.

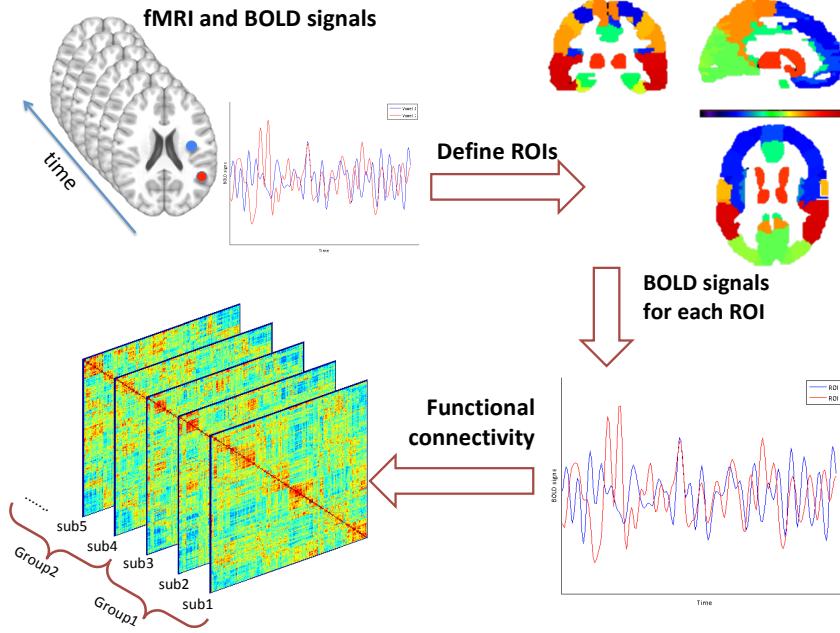


Figure 5: Process to estimate functional connectivity from rfMRI data.

Figure 5 shows the pipeline for calculating the resting state functional connectivity from fMRI data. We used the Automated Anatomical Labeling (AAL) atlas [46], a widely used manual macroanatomical parcellation, and obtained 116 ROIs for each subject. The AAL atlas divided the brains domain into 116 regions including 90 regions in the cerebra and 26 regions in the cerebella. We used the AFNI package of [9] to compute the average BOLD signal over an ROI of all voxel values and then calculated the Pearson correlation matrix of the rfMRI for each subject. Instead of using the raw correlation matrices, we computed Fisher's z-transformed correlation matrices as the responses of our BLGRM. The transformed correlation coefficients range from $-\infty$ to ∞ , while the original signs are preserved.

4.3 Data Analysis Results

We applied BLGRM to detect the group differences for the functional connectivity matrices among the NC, MCI, and AD patient groups. The value \mathbf{x}_i includes an intercept, gender (γ_2), age (γ_3), MCI=1(γ_4), and AD=1(γ_5). We used the same setting for hyperparameters as in the simulation studies. We ran BLGRM with 5,500 MCMC iterations and

500 burn-in iterations. We used BIC to select the number of components, which was 14. Finally, we estimated the highest posterior density (HPD) intervals of γ_4 , γ_5 , $\gamma_5 - \gamma_4$ in order to detect the differences among the three groups.

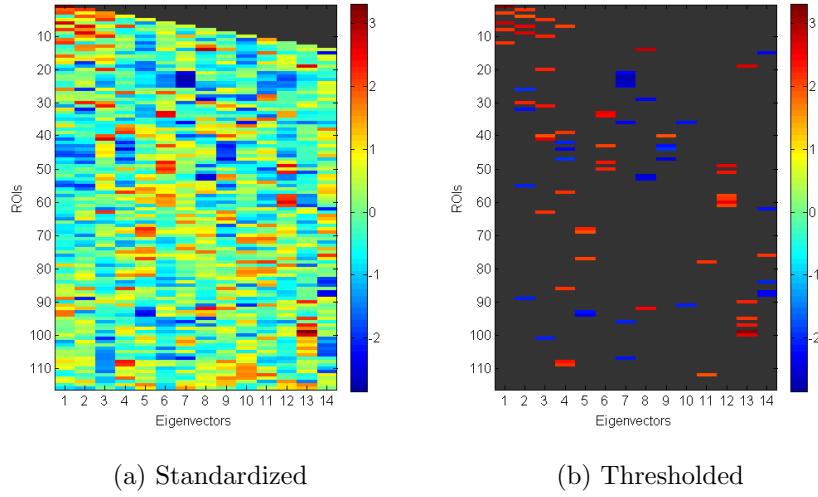


Figure 6: The B matrix (a) estimated by BLGRM with a standardized scale; (b) standardized B matrix after thresholding.

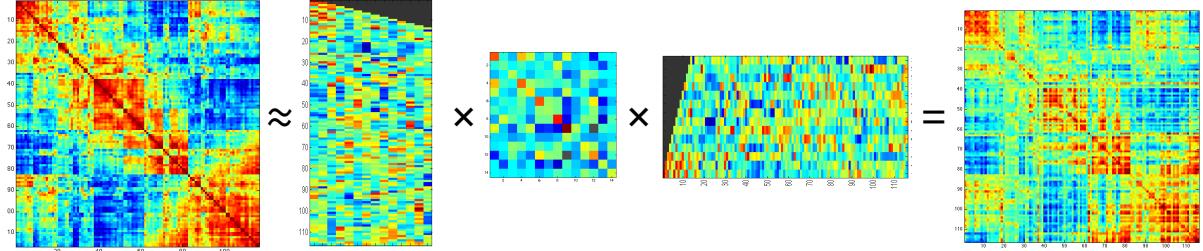


Figure 7: The raw correlation matrix for the first subject is decomposed and approximated by BLGRM.

Figure 6 shows the estimated B matrix. For better presentation, we plotted the standardized B matrix and the thresholded matrix (>1.96). Interestingly, the 1st eigenvector mainly consists of the right precentral gyrus, the right superior parietal gyrus, the right supramarginal gyrus, the right supplementary motor area, the right postcentral gyrus, the bilateral paracentral lobule, and the right inferior parietal gyrus. Most of these main components, including the superior parietal, supramarginal, postcentral, and inferior parietal gyri, and the paracentral lobule, are all within the parietal lobe. The parietal lobe merges sensory information from heterogeneous sources, such as proprioception, mechanoreception in the somatosensory cortex, and the dorsal stream of the visual system. It has been reported that specific parietal regions, such as the posterior parietal cortex, contribute to

retrieval tasks of episodic memory [47]. Therefore, the 1st eigenvector may represent brain functions of sensory information integration and episodic retrieval on the systemic level. Moreover, the 3rd eigenvector has large weights on the left middle frontal gyrus, the left triangular part of the inferior frontal gyrus, the left rolandic operculum, the left insula, the left opercular part of inferior frontal gyrus, the left supplementary motor area, the left superior frontal gyrus, the left supramarginal gyrus, and the right hippocampus. Many of the regions are located within the prefrontal lobe, which plays a central role in cognitive control, “the ability to take charge of one’s actions and direct them towards future, unseen goals” [37]. Moreover, its impairment has been associated with antisocial behavior [52]. Thus, this eigenvector mainly describes the individual’s cognitive control and behavior. The 4th eigenvector composes of the bilateral middle/inferior temporal gyrus, the bilateral fusiform gyri, the left superior temporal gyrus, the left cerebellum crus 1, and the left cerebellum 7b. The inferior temporal gyrus and fusiform gyrus are involved in the high-level visual processing, recognition memory face and body recognition, and color processing [6]. The cerebellum plays a significant role in motor control, and it is associated with attention and language functions [41, 36]. It implies that the 4th eigenvector represents visual and language processing in the higher levels. The most of elements of the 6th eigenvector have high weights on the bilateral inferior/middle/superior occipital gyri. They are secondary visual cortex areas, which are adjacent to the primary visual cortex. The 12th eigenvector mainly consists of the bilateral Calcarine fissure and surrounding cortex, the bilateral lingual gyrus, and the bilateral cuneus, which are all involved in visual processing. The lingual gyrus is especially related to processing of letters. The 13th eigenvector has large weights on the bilateral caudate nucleus, bilateral putamen, and the bilateral anterior cingulate gyrus. The putamen and the caudate nucleus form the dorsal striatum. It is suggested that the dorsal striatum mediates a learning process in which habits are learned by positive reinforcement [38, 39], and is involved in reward-related motor learning and decision making [50, 14]. The reward-related decision making is mediated by other brain regions including anterior cingulate cortices with their targets in the dorsal striatum [14]. Thus, the 13th eigenvector represents brain functions of reward-related learning and decision making.

We also investigated the accuracy of the proposed common component model in relation to the raw correlation matrix for all the subjects with the total reconstruction error=0.70. The total reconstruction error of the LGRM was 0.68. Figure 7 shows that the raw correlation matrix of the first subject can be approximated by the proposed BLGRM.

We examined the estimated regression coefficient $\gamma_5 - \gamma_4$ that represents the effect of a difference between MCI and AD patients on their functional connectivity maps. Figure 8 shows the regression coefficients estimated by the posterior means and the correspond-

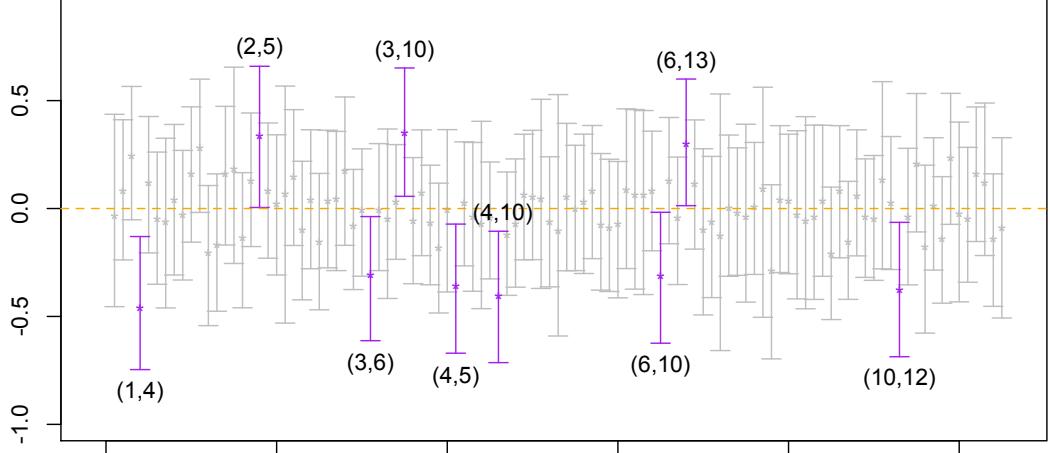


Figure 8: Regression coefficients estimated by the posterior means, marked by “*”, and the corresponding 95% HPD intervals (each vertical line) of $\text{vech}(\Gamma_4 - \Gamma_3)$. The purple lines indicate the 4 pairs of eigenvectors for which the HPD intervals do not include 0.

ing 95% HPD intervals. We identified important group connectivity differences in the eigenspace by examining the HPD intervals. There are 9 HPD intervals that do not include 0, which suggests that these 9 pairs of eigenvectors have important group differences. Based on the estimated B matrix described above, for example of the pair of the 1st and 4th eigenvectors, MCI and AD patients have different functional connectivity among brain regions of sensory information integration/episodic memory retrieval and the high-level vision/language functions. For the pair of the 3rd and 6th eigenvectors, there are group differences in the connections between cognitive control/behavior and visual processing. A group difference between the 6th and 13th eigenvectors implies that AD patients have altered connections between visual function and reward-related decision making.

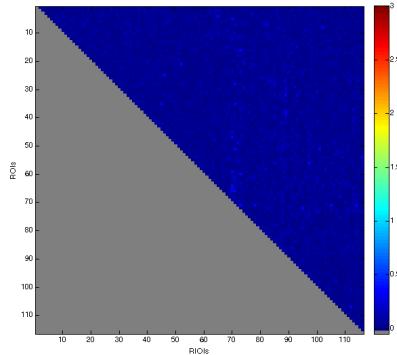


Figure 9: Comparison of the AD and MCI groups: the $-\log_{10}(p)$ of the univariate (ROI-wise) two-sample t-test for the two groups.

As a comparison, we did not find any significant group differences between AD and MCI patients by using an ROI-wise univariate two-sample t-test for Fisher’s z-transformed

correlation coefficients. Figure 9 presents the $-\log_{10}(p)$ of the two-sample t-tests between the AD and MCI groups. After Bonferroni correction, $-\log_{10}(p)$ should be greater than 5.13 at the 0.05 significance level, or it should be greater than 1.30 without any multiple testing correction. However, we did not identify any significant group difference with or without correction.

Table 4: Effect of a group difference between MCI and AD on the following pairs of brain regions. The average of the transformed correlation coefficients in the MCI group is positive.

Region1	Region2	Coeff	Region1	Region2	Coeff
PreCG.R	IOG.L	-0.198	CRBL45.L	FFG.L	-0.179
PreCG.R	MOG.L	-0.192	CRBL6.L	FFG.L	-0.215
PreCG.R	MTG.L	-0.202	CRBL8.L	FFG.L	-0.207
PreCG.R	ITG.L	-0.227	CRBL9.L	FFG.L	-0.201
PreCG.R	FFG.L	-0.206	CRBL6.L	ITG.L	-0.180
MFG.L	MTG.R	-0.174	CRBL8.L	ITG.L	-0.184
MFG.R	ITG.L	-0.205	CRBL9.L	ITG.L	-0.180
MFG.R	FFG.L	-0.172	CRBL6.L	IOG.L	-0.166
SFG.R	MTG.L	-0.201	CRBL6.L	CRBL45.L	-0.172
SFG.R	ITG.L	-0.190	CRBL7b.L	CRBL6.L	-0.187
ITG.L	FFG.L	-0.184	CRBL8.L	CRBL7b.L	-0.172
MTG.L	MFG.R	-0.174	CRBL6.L	CRBL9.L	-0.173
PCL.L	PreCG.L	0.178	CRBL8.L	CRBL9.L	-0.170
PoCG.R	ITG.L	-0.169	CRBL6.L	CERCR1.L	-0.210
CERCR1.L	MFG.R	-0.169	CRBL8.L	CERCR1.L	-0.188
CRBL45.L	MOPFC.L	0.172	CRBL9.L	CERCR1.L	-0.171

To investigate which ROIs have functional connectivity differences between the MCI and AD groups, we mapped the estimated coefficients $\widehat{\Gamma}$ from the eigenspace to the original ROI space. We used \widehat{B} and $\widehat{\Gamma}$ to estimate the coefficient matrix by calculating $\widehat{B}\widehat{\Gamma}\widehat{B}^T$ as the estimated coefficient matrix in the ROI space. After mapping the estimated coefficients from the eigenspace to the original ROI space, we plotted the estimated regression coefficients on the brain template using BrainNet Viewer of Xia et al. [51]. The first 1% of the largest effect sizes were selected to be shown in the figures. Figure 10 shows which brain regions have weaker (or stronger) positive connectivity for AD patients compared to MCI patients, whereas Figure 11 shows which brain regions have different negative connectivity patterns between the MCI and AD groups. In Figure 10, the blue line represents

Table 5: Effect of a group difference between MCI and AD on the following pairs of brain regions. The average of the transformed correlation coefficients in the MCI group is negative.

Region1	Region2	Coeff	Region1	Region2	Coeff
SFG.R	IOG.L	-0.174	CRBL8.L	MFG.L	0.201
SFG.R	MOG.L	-0.181	CRBL9.L	MFG.L	0.173
PreCG.R	MFG.L	0.176	CRBL7b.R	SFG.R	0.172
PreCG.R	MTG.R	0.171	CRBL8.R	SFG.R	0.230
FFG.L	MFG.L	0.176	CRBL9.R	SFG.R	0.176
FFG.L	MOPFC.L	0.178	CRBL10.R	SFG.R	0.186
FFG.L	MOPFC.R	0.187	CRBL8.L	SFG.L	0.195
FFG.R	PreCG.R	0.174	CRBL8.L	SFGM.L	0.176
ITG.L	PreCG.L	0.178	CRBL45.L	MOPFC.R	0.184
ITG.R	PoCG.R	0.183	CRBL6.L	MOPFC.R	0.188
ITG.R	PreCG.R	0.227	CRBL8.L	MOPFC.R	0.172
ITG.R	SFG.R	0.170	CRBL9.L	MOPFC.R	0.181
ITG.R	MFG.R	0.171	CRBL7b.R	PreCG.R	0.207
CRBL7b.R	PoCG.R	0.179	CRBL8.R	PreCG.R	0.204
CRBL8.R	PoCG.R	0.169	CRBL9.R	PreCG.R	0.201
CERCR1.R	PoCG.R	0.171	CERCR1.R	PreCG.R	0.182
CRBL8.L	ITG.R	0.169	CERCR2.R	PreCG.R	0.167

the estimated coefficient value $\gamma_5 - \gamma_4 < 0$, indicating that AD patients have a weaker positive connection than MCI subjects between the corresponding two ROIs. The red line represents the estimated coefficient value $\gamma_5 - \gamma_4 > 0$, which implies that AD patients have a stronger positive connection than MCI subjects between the corresponding two ROIs. For the negative connectivity, in Figure 11, the blue line indicates $\gamma_5 - \gamma_4 > 0$, which implies that AD patients have a weaker negative connection than MCI subjects between the corresponding two ROIs. Tred line implies that AD patients have a stronger negative connection than MCI subjects between the corresponding two ROIs. Various brain regions have weaker or stronger connections for AD patients, which implies that the functional networks of AD patients are altered compared to those of MCI subjects.

To examine functional connectivity differences between MCI and AD patients, we focus on Figures 10 and 11 and Tables 4 and 5. All the abbreviations used in this paper for the AAL template are shown in Table 6. The regression coefficients have different inter-

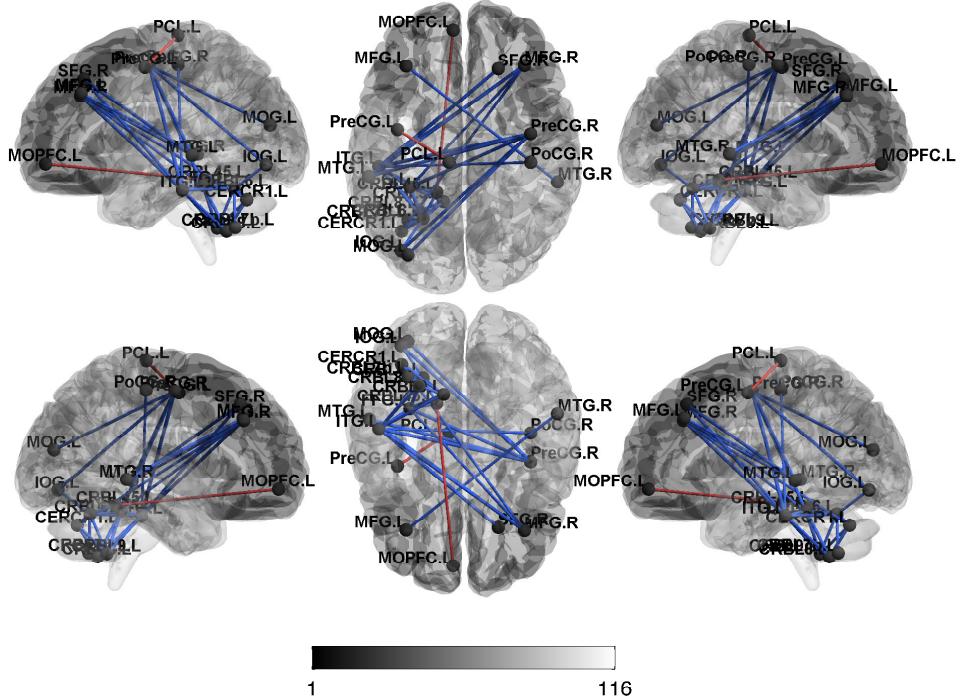


Figure 10: Pairs of ROIs that have different positive connectivity among the MCI group vs. the AD group.

pretations depending on the sign of the average transformed correlation coefficient in MCI patients. Table 4 shows selected regression coefficient estimates, where the average of the transformed correlation coefficient in MCI patients is positive; whereas Table 5 presents the selected coefficient estimates in the negative cases. We list some of the regression coefficients with sizes in the top 1%.

Table 4 and Figure 10 one can observe that AD patients have lower positive connections among some frontal, temporal, occipital brain areas, mostly between the left and the right hemispheres. It suggests that AD patients have weaker positive connections between the left and the right hemispheres compared to those for MCI patients. For example, the right precentral gyrus has a weaker positive connection with the left inferior/middle temporal gyrus, left inferior/middle occipital gyrus and the left fusiform gyrus for AD patients compared to those for MCI patients. The precentral gyrus controls the voluntary movements of skeletal muscles. Furthermore, many left cerebellum regions have weaker positive connections with the left inferior temporal gyri and the left fusiform gyrus for AD patients compared to those for MCI patients, while positive connections among many left cerebellum regions themselves are weaker for AD patients. Interestingly, AD patients have higher positive connections not only between the left paracentral lobule and the left precentral lobule but also between the left cerebellum area and the left medial orbital prefrontal cortex. This implies disrupted connections for AD patients among high-level visual processing,

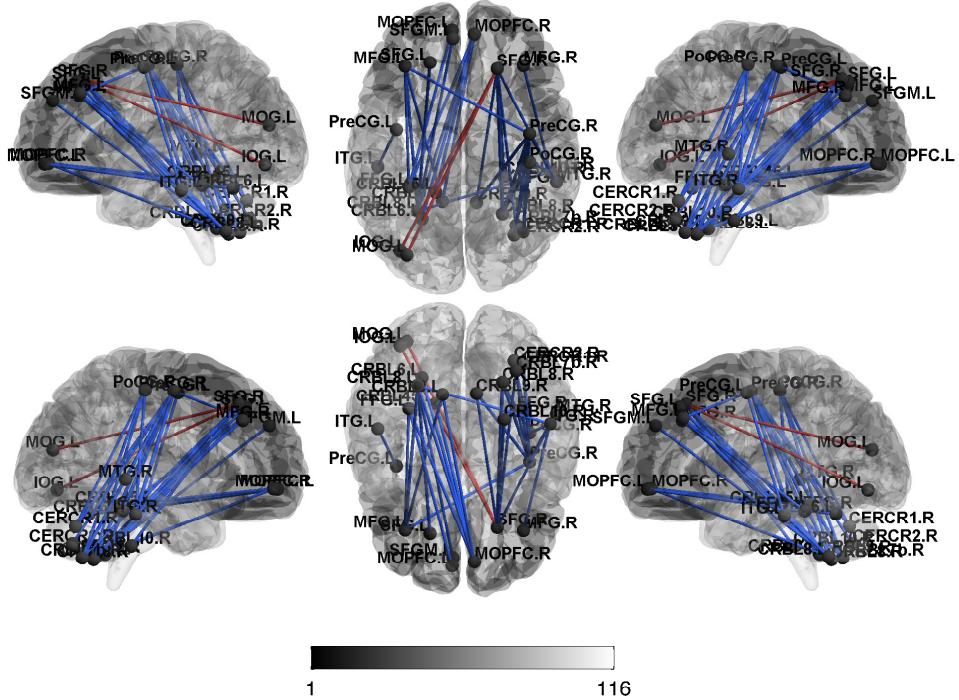


Figure 11: Pairs of ROIs that have different negative connectivity among the MCI group vs. the AD group.

motor control, voluntary movement control, and language functions.

Table 5 and Figure 11 present altered negative functional connectivity mostly among the frontal, temporal, occipital and cerebellum brain regions between the MCI and AD groups. We find that AD patients have weaker negative connections between the left cerebellum areas and the right middle/superior frontal gyri, the right medial orbital prefrontal cortex, the right postcentral gyrus, and the right precentral gyrus. Also, the right superior frontal gyrus has stronger negative connections with inferior/middle occipital gyrus for AD patients compared to those of MCI patients. This suggests that, compared to MCI subjects, AD patients have altered negative connectivity among sensory information integration, episodic memory retrieval on the systemic level, motor control, language functions and visual processing.

5 Discussion

In this study, we proposed a BLGRM by taking a global approach to analyze brain functional connectivity. Our model decomposes any symmetric data matrices with common eigenmaps across subjects and the subject-specific coefficient matrix. We see that the subject-specific coefficient matrix preserves an individual network structure in the low-dimensional space spanned by the common factors. We took a Bayesian approach to

Table 6: AAL parcellation of the entire brain and the corresponding abbreviations used in this paper.

Abbreviation	Name	Classification
SFG	Superior frontal gyrus	Prefrontal Lobe
SFGM	Superior frontal gyrus, medial	Prefrontal Lobe
MFG	Middle frontal gyrus	Prefrontal Lobe
MOPFC	Medial orbital prefrontal cortex	Other frontal
PreCG	Precentral gyrus	Other frontal
PCL	Paracentral lobule	Parietal lobe
PoCG	Postcentral gyrus	Parietal lobe
IOG	Inferior occipital gyrus	Occipital lobe
MOG	Middle occipital gyrus	Occipital lobe
MTG	Middle temporal gyrus	Temporal lobe
ITG	Inferior temporal gyrus	Temporal lobe
FFG	Fusiform gyrus	Temporal lobe
CRBL	Cerebellum	Cerebellum
CERCR	cerebellum crus	Cerebellum

estimate the underlying factors, individual coefficient matrix, and some parameters involved in the prediction model for clinical outcomes. We assumed a hierarchical structure within the prior of Λ_i so that we could automatically estimate the effects of covariates on the Λ_i matrices within the MCMC iterations. Furthermore, we took the parameter expansion approach on our graph regression model to reduce the posterior dependence between B and the Λ_i 's.

The simulation studies demonstrated that our method efficiently approximated the raw symmetric matrix data with good MCMC mixing properties. We compared the reconstruction errors of our method with those of the LGRM, CP decomposition, and three-way DEDICOM. Our method efficiently approximated the raw data matrices by fewer eigenvectors than the competing methods, and allowed the regression parameters for L_i to be recovered satisfactorily.

Our analysis of real data from the ADNI revealed that the bilateral cerebellum areas had weaker connections with the temporal/frontal/occipital lobe regions for AD patients than for MCI patients. Also, inter-connections within the left cerebellum areas of AD patients were weaker than those of MCI patients. It has been demonstrated that cerebellar volume is apparently reduced at the predementia stage, which supports the involvement

of the cerebellum in the progression of dementia [3]. Our study suggests that the cerebellum is an important brain region in terms of its inter/intra functional connectivity. For example, AD patients showed impaired resting-state functional connectivity within the cerebellum regions compared the healthy controls [48, 55]. The functional connectivity of healthy controls were significantly higher than that of MCI patients not only between the middle frontal gyrus and the fusiform gyrus but also between cerebellum and the fusiform gyrus [4]. Also, there was a connectivity difference between the precentral gyrus and other brain regions, including the middle frontal gyri, some occipital gyri, and some temporal gyri. This suggests that AD patients have altered connections among brain regions that are involved in sensory information integration, high-level visual and auditory processing, language functions, and motor control.

This study may provide a guideline for elucidating hidden pathology of neurological disorders through the perspective of brain connectivity. In clinical applications, our method can be used to (1) examine whether subjects with normal brain function and patients with altered brain function (or among disease subtypes) have different functional connectivity structures and where the differences originate. This exploratory analysis allows for better understanding the underlying mechanism of a disorder, which may help to develop future treatments that target some identified brain regions. Also, altered connectivity can provide (2) diagnostic and prognostic information [16]. To incorporate clinical outcomes to measure diagnostic and prognostic status, our method can be modified by reformulating the regression model (3) to use the clinical outcomes. Capturing disruption in functional connectivity can be important for better/earlier diagnosis of psychiatric disorders. For the example of attention deficit hyperactivity disorder (ADHD), a subject is diagnosed with ADHD if the subject meets the Diagnostic and Statistical Manual of Mental Disorders criteria. However, the test result, which is based on subjective criteria, can vary depending on the interview environment, the examiner and the examinee.

If abnormality of functional connectivity is a very early sign of a disorder or is a biomarker for prognosis, it will be promising to use the information as an objective diagnostic/prognostic tool. One more intriguing suggestion is that (3) abnormal functional connectivity will become prominent in neurogenetic studies [31]. Although genetic factors are emerging in research of psychiatric disorders, their effect sizes are very small and their working mechanisms still need to be elucidated. Because the brains organization and function are influenced by genetic factors [32, 33] and brain function has a relatively large impact on disease progression, functional connectivity can be a mediator to explain connections among genetic mutations and disorders.

References

- [1] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.
- [2] B. W. Bader, R. A. Harshman, and T. G. Kolda. Temporal analysis of semantic graphs using asalsan. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 33–42. IEEE, 2007.
- [3] L. Baldaçara, J. G. F. Borgio, W. A. d. S. Moraes, A. L. T. Lacerda, M. B. M. M. Montaño, S. Tufik, R. A. Bressan, L. R. Ramos, and A. P. Jackowski. Cerebellar volume in patients with dementia. *Revista Brasileira de Psiquiatria*, 33(2):122–129, 2011.
- [4] A. Bokde, P. Lopez-Bayo, T. Meindl, S. Pechler, C. Born, F. Faltraco, S. Teipel, H.-J. Möller, and H. Hampel. Functional connectivity of the fusiform gyrus during a face-matching task in subjects with mild cognitive impairment. *Brain*, 129(5):1113–1124, 2006.
- [5] R. Bro and H. A. Kiers. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics*, 17:274–286, 2003.
- [6] R. L. Buckner, W. Koutstaal, D. L. Schacter, and B. R. Rosen. Functional mri evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain*, 123(3):620–640, 2000.
- [7] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter. The brain’s default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008.
- [8] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10:186–98, Mar. 2009.
- [9] R. W. Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29:162–173, 1996.
- [10] M.-C. de LaCoste and C. L. White. The role of cortical connectivity in alzheimer’s disease pathogenesis: a review and model system. *Neurobiology of Aging*, 14:1–16, 1993.
- [11] X. Delbeuck, M. Van der Linden, and F. Collette. Alzheimer’s disease as a disconnection syndrome? *Neuropsychology Review*, 13:79–92, 2003.
- [12] D. Durante, D. Dunson, and J. T. Vogelstein. Nonparametric bayes modeling of populations of networks. *Journal of the American Statistical Association*, 2017.
- [13] H. Eavani, T. D. Satterthwaite, R. Filipovych, R. E. Gur, R. C. Gur, and C. Da-

- vatzikos. Identifying sparse connectivity patterns in the brain using resting-state fmri. *NeuroImage*, 105:286–299, 2015.
- [14] B. J. Everitt and T. W. Robbins. From the ventral to the dorsal striatum: devolving views of their roles in drug addiction. *Neuroscience & Biobehavioral Reviews*, 37(9):1946–1954, 2013.
 - [15] A. Fornito, A. Zalesky, and M. Breakspear. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage*, 80:426–444, 2013.
 - [16] M. D. Fox and M. Greicius. Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, 4:19, 2010.
 - [17] K. Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS biol*, 7(2):e1000033, 2009.
 - [18] W. Gao, H. Zhu, K. S. Giovanello, J. K. Smith, D. Shen, J. H. Gilmore, and W. Lin. Evidence on the emergence of the brain’s default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proceedings of the National Academy of Sciences*, 106(16):6790–6795, 2009.
 - [19] A. Gelman et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1:515–534, 2006.
 - [20] J. Geweke and G. Zhou. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9:557–587, 1996.
 - [21] J. Ghosh and D. B. Dunson. Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18:306–320, 2009.
 - [22] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.
 - [23] R. A. Harshman. Models for analysis of asymmetrical relationships among n objects or stimuli. In *First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, McMaster University, Hamilton, Ontario*, volume 5, 1978.
 - [24] R. A. Harshman and M. E. Lundy. Uniqueness proof for a family of models sharing features of tucker’s three-mode factor analysis and parafac/candecomp. *Psychometrika*, 61:133–154, 1996.
 - [25] P. D. Hoff. A hierarchical eigenmodel for pooled covariance estimation. *Journal of the Royal Statistical Society: Series B*, 71:971–992, 2009.
 - [26] S. A. Huettel, A. W. Song, and G. McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.
 - [27] H. A. Kiers. A three-step algorithm for candecomp/parafac analysis of large data sets

- with multicollinearity. *Journal of Chemometrics*, 12:155–171, 1998.
- [28] W. H. Kim, N. Adluru, M. K. Chung, O. C. Okonkwo, S. C. Johnson, B. B. Bendlin, and V. Singh. Multi-resolution statistical analysis of brain connectivity graphs in preclinical alzheimer’s disease. *NeuroImage*, 118:103–117, 2015.
 - [29] E. D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer Science, 2009.
 - [30] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51:455–500, 2009.
 - [31] K. Konrad and S. B. Eickhoff. Is the adhd brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human Brain Mapping*, 31:904–916, 2010.
 - [32] J. W. Koten, G. Wood, P. Hagoort, R. Goebel, P. Propping, K. Willmes, and D. I. Boomsma. Genetic contribution to variation in cognitive function: an fmri study in twins. *Science*, 323:1737–1740, 2009.
 - [33] R. K. Lenroot, J. E. Schmitt, S. J. Ordaz, G. L. Wallace, M. C. Neale, J. P. Lerch, K. S. Kendler, A. C. Evans, and J. N. Giedd. Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human Brain Mapping*, 30:163–174, 2009.
 - [34] C. Liu, D. B. Rubin, and Y. N. Wu. Parameter expansion to accelerate em: The px-em algorithm. *Biometrika*, 85:755–770, 1998.
 - [35] J. S. Liu and Y. N. Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94:1264–1274, 1999.
 - [36] M. Manto, J. M. Bower, A. B. Conforto, J. M. Delgado-García, S. N. F. da Guarda, M. Gerwig, C. Habas, N. Hagura, R. B. Ivry, P. Mariën, et al. Consensus paper: roles of the cerebellum in motor controlthe diversity of ideas on cerebellar involvement in movement. *The Cerebellum*, 11(2):457–487, 2012.
 - [37] E. K. Miller, D. J. Freedman, and J. D. Wallis. The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357:1123–1136, 2002.
 - [38] M. Mishkin, B. Malamut, and J. Bachevalier. Memories and habits: Two neural systems. *Neurobiology of learning and memory*, pages 65–77, 1984.
 - [39] M. G. Packard and B. J. Knowlton. Learning and memory functions of the basal ganglia. *Annual review of neuroscience*, 25(1):563–593, 2002.
 - [40] T. Park and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686, 2008.
 - [41] M. Rapoport, R. van Reekum, and H. Mayberg. The role of the cerebellum in cog-

- nition and behavior: a selective review. *The Journal of neuropsychiatry and clinical neurosciences*, 12(2):193–198, 2000.
- [42] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52:1059–1069, 2010.
- [43] S. L. Simpson, M. N. Moussa, and P. J. Laurienti. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks.”. *NeuroImage*, 60:1117–1126, 2012.
- [44] S. L. Simpson, F. Bowman, and P. J. Laurienti. Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Statistics Surveys*, 7:1–36, 2013.
- [45] C. J. Stam. Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15:683–695, 2014.
- [46] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.
- [47] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner. Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, 9:445–453, 2005.
- [48] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang. Altered functional connectivity in early alzheimer’s disease: A resting-state fmri study. *Human Brain Mapping*, 28(10):967–978, 2007.
- [49] L. Wang, Y. Zang, Y. He, M. Liang, X. Zhang, L. Tian, T. Wu, T. Jiang, and K. Li. Changes in hippocampal connectivity in the early stages of alzheimer’s disease: evidence from resting state fmri. *NeuroImage*, 31:496–504, 2006.
- [50] J. R. Wickens, J. N. Reynolds, and B. I. Hyland. Neural mechanisms of reward-related motor learning. *Current opinion in neurobiology*, 13(6):685–690, 2003.
- [51] M. Xia, J. Wang, and Y. He. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8:e68910, 2013.
- [52] Y. Yang and A. Raine. Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. *Psychiatry Research: Neuroimaging*, 174:81–88, 2009.
- [53] J. Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61:167–191, 2005.
- [54] A. Zalesky, A. Fornito, and E. Bullmore. Network-based statistic: identifying differences in brain networks. *NeuroImage*, 53:1197–1207, 2010.
- [55] Z. Zhang, Y. Liu, B. Zhou, J. Zheng, H. Yao, N. An, P. Wang, Y. Guo, H. Dai,

L. Wang, et al. Altered functional connectivity of the marginal division in alzheimers disease. *Current Alzheimer Research*, 11(2):145–155, 2014.

6 Appendix

CANDECOMP/PARAFAC (CP), proposed by Kiers [27], decomposes a tensor into a sum of component rank-one tensors such as

$$\mathbf{L} \approx \sum_{r=1}^R \mathbf{u}_r \circ \mathbf{s}_r \circ \mathbf{q}_r,$$

where \mathbf{L} is a $V \times V \times n$ tensor, and $\mathbf{q}_r \in \mathbb{R}^n$, $\mathbf{u}_r \in \mathbb{R}^V$, $\mathbf{s}_r \in \mathbb{R}^V$. Because L_i is symmetric, we set $\mathbf{u}_r = \mathbf{s}_r$. Here, we reformulate the CP model to be comparable with our model as follows:

$$L_i \approx \sum_{r=1}^R q_{ir} (\mathbf{u}_r \circ \mathbf{u}_r) \text{ for } i = 1, \dots, n.$$

Then \mathbf{u}_r plays the role of the common basis and q_{ir} can be considered as a subject-specific scalar coefficient. Three-way DEDICOM is an extended version of the DEDICOM model that can incorporate a third mode of the data [23] and decompose asymmetry matrices. Here, we only consider symmetric $L_i \in \mathbb{R}^{V \times V}$ for comparison. Then the model can be written as

$$L_i \approx \mathbf{A} \mathbf{D}_i \mathbf{Q} \mathbf{D}_i \mathbf{A}' \text{ for } i = 1, \dots, n,$$

where latent components $\mathbf{A} \in \mathbb{R}^{V \times R}$, interactions between different components $\mathbf{Q} \in \mathbb{R}^{R \times R}$, and a diagonal matrix $\mathbf{D}_i \in \mathbb{R}^{R \times R}$. The r -th diagonal element in \mathbf{D}_i represents the i -th subject-specific weight of the r -th latent component. Three-way DEDICOM is similar to our decomposition model in the sense that if Λ_i can be decomposed as $\mathbf{D}_i \mathbf{Q} \mathbf{D}_i$, then the center parts of the two decomposition models are equivalent. Thus, DEDICOM is a more constrained version of our decomposition model. Bader et al. [2] proposed an algorithm, alternating simultaneous approximation, least squares, and Newton (ASALSAN), to compute the three-way DEDICOM. We use a Python module "scikit-tensor" to use the ASALSAN algorithm available in <https://github.com/mnick/scikit-tensor>. To calculate the CP decomposition, we use the N-way Toolbox in MATLAB.

The frequentist LGRM estimates the parameter matrices in (2) by minimizing the Frobenius norm of $L_i - B\Lambda_i B^T$ using iterative optimization steps. This optimization problem is solved by adapting low-rank approximation techniques proposed by Ye [53]. In detail, we consider the following optimization problem

$$\min_{B, \Lambda_i} \sum_{i=1}^n \|L_i - B\Lambda_i B^T\|_F^2 \text{ such that } B'B = I_R. \quad (20)$$

Then,

$$\widehat{\Lambda}_i = \widehat{B} L_i \widehat{B}',$$

where the eigenmap B can be estimated from the following iterative optimization steps.

Under $\Lambda_i = BL_iB^T$, the above minimizing problem is equivalent to maximizing

$$\begin{aligned} \sum_{i=1}^n \|B'L_iB\|_F^2 &= \sum_{i=1}^n \text{tr}(B'L_iBB'L_iB) \\ &\approx \sum_{i=1}^n \text{tr}(B'L_iB_0B_0'L_iB), \end{aligned}$$

where B_0 is the B matrix from the previous iteration. Then B can be estimated by the following iterative steps:

1. Let B_0 be the B matrix from the previous iteration.
2. Calculate $\mathbf{Q} = \sum_{i=1}^n L_i B_0 B_0' L_i$.
3. Compute the R eigenvectors $\{\phi_i\}_{i=1}^R$ of \mathbf{Q} corresponding to the largest R eigenvalues.
4. Set $B = [\phi_1, \phi_2, \dots, \phi_R]$.
5. Repeat the above iterations until it converges.
6. Calculate $\widehat{\Lambda}_i = \widehat{B}L_i\widehat{B}'$, for $i = 1, \dots, n$.

Also, the variance component σ^2 can be estimated by maximum likelihood estimation:

$$\widehat{\sigma}^2 = \frac{2}{nV(V+1)} \text{tr} \left(\sum_{i=1}^n \left(L_i - \widehat{B}\widehat{\Lambda}_i\widehat{B}' \right)^2 \right).$$