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

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

March 2, 2017

Abstract

We propose a Bayesian low-rank graph regression modeling (BLGRM) framework for the regression analysis of matrix response data across subjects. This development is motivated by performing detailed comparisons of functional and structural connectivity data across subjects, groups, and time and relating 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

*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 supervison of Drs. Ibrahim and Zhu in 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, 27599. Hongtu Zhu is also Professor of Biostatistics, Department of Biostatistics, The University of Texas, MD Anderson Cancer Center, Houston, 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 upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. This work was partially supported by NIH grants MH086633, NSF grants SES-1357666 and DMS-1407655, and a grant from 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 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 ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: *http* : *//adni.loni.usc.edu/wp − content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*.

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 status, 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 graph regression model (PX-BLGRM) 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 its comparison with several competing approaches. We apply BLGRM to the rest functional magnetic resonance imaging (rfMRI) data set obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

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

1 Introduction

The Alzheimer's Disease Neuroimaging Initiative (ADNI) study ($http://www.adni-info.org/$) as 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 question of interest for ADNI is to quantify the clinical, cognitive, imaging, genetic and biochemical biomarker characteristics of the entire spectrum of AD as the pathology evolves from normal aging (NC), to mild cognitive impairment (MCI), to dementia or AD. This paper is motivated by the joint analysis of rest 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 questions:

- (Q1) the first one is to derive functional-connectivity based brain biomarkers for classifying AD, NC, and MCI groups;
- (Q2) the second one is to identify a common low-dimensional subspace that characterizes the major variations of brain functional network across subjects and groups.

Statistically, these questions of interest can be formulated as the use of a vector of predictors (e.g., diagnosis status), denoted as $\mathbf{x} = (x_1, \ldots, x_p)^T$, to predict a $V \times V$ matrix response, denoted as $\mathbf{L} = (L_{(g,g')})_{1 \le g,g' \le V}$, where g is a vertex and V is the total number of vertices. In this case, $\mathbf{L}$ is the rfMRI connectivity matrix and $\mathbf{x}$ may include age, gender, and diagnosis status (AD, NC, or MCI). The rfMRI data has 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 [8, 16, 13, 9, 3].

To answer questions (Q1) and (Q2), we develop a Bayesian low-rank graph regression modeling (BLGRM or Bayesian LGRM) framework to deal with three challenges arising

from the use of $\mathbf{x}$ to predict high-dimensional $\mathbf{L}$. 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 common to believe that organization of brain networks are governed by both short- and long-range connections among different brain regions. Moreover, in most neuroimaging studies, the dimension of functional and structural connectivity data (or $\mathbf{L}$) can be much larger than the number of subjects, which varies from several dozens to a few thousands. Specifically, for the rfMRI data set from ADNI, we have $n = 153$ and each connectivity matrix $\mathbf{L}$ contains $V(V-1)/2 = 6670$ unique elements for $V = 116$. There is a 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 graph theoretic methods. The graph theoretic methods consist of calculating graph 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. A key limitation of the graph theoretic methods is that it cannot reveal subtle differences at both 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. One fundamental issue of 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 the low statistical power of detecting from weak to moderate signals [17], even though there are some network analogues of cluster-based thresholding methods, such as the network-based statistic and spatial pairwise clustering methods. So, alternative connectivity analysis methods are crit-

ically needed for understanding the complex organization of brain network, while achieving dimension reduction.

There is a great interest in developing statistical models that explicitly delineate the conditional distribution of $\mathbf{L}$ given $\mathbf{x}$, that is, $p(\mathbf{L}|\mathbf{x})$. Ideally, such 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 graph models, stochastic block models, and latent space models, were primarily developed for binary networks, whereas their extensions to weighted networks remain in their infancy. Little has been done on 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 Bayesian low-rank graph regression models with matrix responses and clinical covariates. Our BLGRM can be regarded as a novel extension of the hierarchical eigenmodel for pooled covariance matrices [15], tensor decomposition [18], and standard linear model. The key idea of LGRM is to find an intrinsic low-dimensional subspace for all subjects, denoted as $\boldsymbol{B}$. Such $\boldsymbol{B}$ not only allows us to dramatically reduce dimension, but also characterizes the organization of brain connection maps across subjects at a system level. We further represent each graph response by using a $R \times R$ subject-specific coefficient matrix ($\boldsymbol{\Lambda}_i$), while preserving an individual network structure of the low-dimensional eigenspace. One more intriguing part is that we introduce a hierarchical structure of $\boldsymbol{\Lambda}_i$ in order to incorporate the effects of clinical/demographic covariates on graph responses. Due to weak identifiability and high posterior dependence among parameters of our decomposition model, BLGRM suffers from poor mixing and slow convergence of MCMC samplers. We adapt a parameter expansion approach to our graph regression model (PX-BLGRM) to address the issues. Based on this formulation, an efficient Markov chain Monte Carlo (MCMC) algorithm is used to perform posterior computation. We take a

Bayesian approach to estimate the parameters involved in $\boldsymbol{B}$ and the regression coefficients associated with covariates. We examine if effects of covariates of interest are important on explaining graph responses by utilizing highest posterior density (HPD) intervals.

The rest of this paper is organized as follows. In Section 2, we introduce BLGRM in details 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 dataset discussed above in order to address (Q1)-(Q2). Section 5 presents concluding remarks.

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 vertexes. Without loss of generality, it is assumed that $\mathbf{L}_i$ 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, including a common component model and a regression model. The *common component model* is given by

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

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

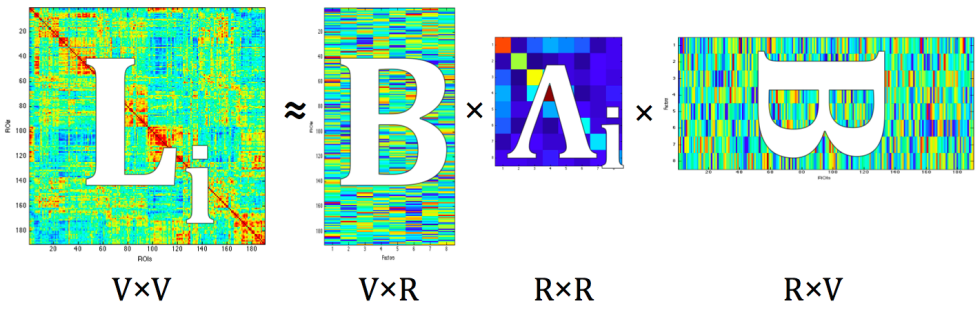
(1) can be written in a matrix form as follows:

$$\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i, \tag{2}$$

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

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

where $\delta_{i,(r,s)}$ are measurement errors and $f_{r,s}(\mathbf{x}_i)$ is an 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, say $\{\phi_k(\mathbf{x}_i)\}_{k\le K}$ 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 high-dimensionality of connectivity matrices. Second, the R eigenmaps $\boldsymbol{\beta}_r = \{\beta_r(g) : g \in \mathcal{G}\}$ can be considered as independent networks of vertices that characterize the latent organization of connectivity structures across subjects at a system level. Then, the subject-specific coefficient matrix $\boldsymbol{\Lambda_i}$ preserves an individual network structure in the low-dimensional space spanned by the eigenmaps. It enables us to construct an underlying relational structure among brain areas and to reduce heterogeneity of functional connectivity in the low-dimensional space. Third, $\boldsymbol{\eta}_{i,r} = \{\eta_{i,r}(g) = \sum_{s=1}^{R}\beta_s(g)\lambda_{i,(s,r)} : g \in \mathcal{G}\}$ delineate individual organization of connection maps between vertices across R different

latent networks. One more intriguing part is that we assume a hierarchical structure within a prior of Λ_i in order to estimate effects of clinical/demographic covariates. Our method facilitates estimation of the effects of covariates on the Λ_i matrices while decomposing the connectivity structure within MCMC iterations. By considering a diagnostic indicator as a covariate, e.g., Alzheimer's disease or cognitively normal, local differences in functional connectivity can be detected. Also, this hierarchical structure allows the decomposition to be supervised by average covariates of subjects. Thus, the estimated eigenmap $\boldsymbol{B}$ not only explains variation of the connectivity across subjects but also takes into account for interrelation between covariates and connectivity.

Details for a Bayesian approach will be followed in the next section.

2.2 Bayesian Approach with Standard Priors

In order to estimate parameters of interest, we take a Bayesian approach. The full posterior distribution is proportional to

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

We assume that the measurement errors $\epsilon_i(g, g')$ follow i.i.d. symmetric Normal distribution [28] with the mean 0 and the variance σ^2. For the pre-specified number of eigenvectors R, the likelihood of $\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n$ can be written by

$$p(\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n | \boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{nV(V+1)/2} \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma^2}\mathrm{tr}\left(\left(\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'\right)^2\right)\right].$$

To ensure identifiability of the decomposition model, we assume that the eigenmap $\boldsymbol{B}$ is a lower triangular matrix [11]. Also, the number of eigenvectors R should be chosen such that $VR - R(R-1)/2 + nR(R+1)/2 \leq nV(V+1)/2$. We assume the following priors

$$\begin{aligned} \beta_{ir} &\sim N\left(0, \tau_r^{-1}\phi_{ir}^{-1}\right), \ 1 \leq i \leq V, \ 1 \leq r \leq \min(i, R) \\ \phi_{ir} &\sim Gamma(1, a_0/2), \ a_0 \sim Gamma(a_1, a_2), \ \pi(\tau_r) = 1/\tau_r. \end{aligned}$$

$$\pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n | \boldsymbol{\Delta}_1, \cdots, \boldsymbol{\Delta}_n, \sigma_0^2) \propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma_0^2}\text{tr}\left((\boldsymbol{\Lambda}_i - \boldsymbol{\Delta}_i)^2\right)\right]$$

$$\pi(\boldsymbol{\Gamma}) \propto \exp\left[-\frac{1}{2\sigma_\gamma^2}\text{tr}\left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right)\right], \ \sigma^2 \sim \text{IG}(b_1, b_2), \ \sigma_\gamma^2 \sim \text{IG}(c_1, c_2),$$

where $vech(\boldsymbol{\Delta}_i) = \boldsymbol{\Gamma}'\boldsymbol{x}_i = \left[\boldsymbol{x}_i'\boldsymbol{\gamma}_1, \boldsymbol{x}_i'\boldsymbol{\gamma}_2, \cdots, \boldsymbol{x}_i'\boldsymbol{\gamma}_q\right]'$, $q = R(R+1)/2$. The regression covariates for the i-th subject are denoted by $\boldsymbol{x}_i = (1, x_1, \cdots, x_{p-1})'$ and the regression coefficients are given by $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_q] = [\boldsymbol{\gamma}_j]_{j=1}^q$, where $\boldsymbol{\gamma}_j = (\gamma_{0j}, \cdots, \gamma_{(p-1)j})'$. For lower triangular elements of $\boldsymbol{B}$, we adapted Bayesian lasso priors [27] for each column to induce sparse eigenmap loadings for each eigenvector.

2.3 Parameter-Expanded Model

Even with the identifiability constraints, our proposed model suffers from poor mixing and slow convergence of MCMC samplers due to weak identifiability and high posterior dependence of $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$'s. For better identification of parameters, we fix σ_0, the scale of $\boldsymbol{\Lambda}_i$, as 1. We adapt a parameter expansion approach to our graph regression model to diminish posterior dependence between $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$'s. Parameter expansion has been proposed by Liu et al. [23] for computational efficiency of expectation-maximization (EM) algorithms by introducing a parameter-expanded model with overparameterized parameters. The parameter expansion approach also improves convergence and mixing of MCMC samplers and reduces high posterior dependence among parameters [24, 10, 12]. Ghosh and Dunson [12] proposed parameter expansion in Bayesian factor analysis to weaken posterior dependence of parameters and to induce heavier-tailed priors for factor loadings. They showed improved mixing of parameters that were transformed back to the original model from the parameter-expanded model.

We rewrite the original model (2) as the following parameter-expanded model,

$$\begin{aligned} \boldsymbol{L}_i &= \boldsymbol{B}^*\boldsymbol{\Lambda}_i^*\boldsymbol{B}^{*\prime} + \boldsymbol{\epsilon}_i, \\ \boldsymbol{B} &= s(\boldsymbol{B}) \circ \boldsymbol{B}^*\boldsymbol{\Psi}^{-1/2}, \\ \boldsymbol{\Lambda}_i &= d(\boldsymbol{B})\boldsymbol{\Psi}^{1/2}\boldsymbol{\Lambda}_i^*\boldsymbol{\Psi}^{1/2}d(\boldsymbol{B}), \end{aligned} \tag{4}$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \cdots, \psi_R)$, $d(\boldsymbol{B}) = \text{diag}\left(\text{sign}(\beta_{11}), \cdots, \text{sign}(\beta_{RR})\right)$, and $s(\boldsymbol{B})$ is a $V \times R$ matrix whose every row is $[\text{sign}(\beta_{11}), \cdots, \text{sign}(\beta_{RR})]$. The sign of diagonal elements in $\boldsymbol{B}$ is accordingly multiplied for the indentifiability purpose. In a similar manner of the original setup, we assume the following priors for the model (4).

$$\begin{aligned} \beta_{ir}^* &\sim N\left(0, \tau_r^{-1}\phi_{ir}^{-1}\right), \ 1 \le i \le V, \ 1 \le r \le \min(i, R) \\ \phi_{ir} &\sim Gamma(1, a_{0r}/2), \ a_{0r} \sim Gamma(a_1, a_2), \ \pi(\tau_r) = 1/\tau_r \\ \pi(\boldsymbol{\Lambda}_1^*, \cdots, \boldsymbol{\Lambda}_n^* | \boldsymbol{\Delta}_1^*, \cdots, \boldsymbol{\Delta}_n^*, \boldsymbol{\Psi}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\text{tr}\left\{\left((\boldsymbol{\Lambda}_i^* - \boldsymbol{\Delta}_i^*)\boldsymbol{\Psi}\right)^2\right\}\right] \\ \pi(\boldsymbol{\Gamma}^*) &\propto \exp\left[-\frac{1}{2\sigma_\gamma{}^2}\text{tr}\left(\boldsymbol{\Gamma}^{*\prime}\boldsymbol{\Gamma}^*\right)\right], \ \sigma^2 \sim \text{IG}(b_1, b_2), \ \sigma_\gamma^2 \sim \text{IG}(c_1, c_2) \\ \psi_i &\sim Gamma(\nu_a, \nu_a), \ i = 1, \cdots, R, \end{aligned}$$

where $vech(\boldsymbol{\Delta}_i^*) = \boldsymbol{\Gamma}^{*\prime}\boldsymbol{x}_i = \left[\boldsymbol{x}_i'\boldsymbol{\gamma}_1^*, \boldsymbol{x}_i'\boldsymbol{\gamma}_2^*, \cdots, \boldsymbol{x}_i'\boldsymbol{\gamma}_q^*\right]'$, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^*(d(\boldsymbol{B})\boldsymbol{\Psi}^{1/2}) \otimes_s (d(\boldsymbol{B})\boldsymbol{\Psi}^{1/2})$, and $q = R(R+1)/2$.

Then, the posterior distribution for each parameter is given by

$$\pi(\boldsymbol{B}^*|\bullet) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\text{tr}\left(\left(\boldsymbol{L}_i - \boldsymbol{B}^*\boldsymbol{\Lambda}_i^*\boldsymbol{B}^{*\prime}\right)^2\right) - \frac{1}{2}\sum_{r=1}^{R}\sum_{i=r}^{V}(\beta_{ir}^*)^2\tau_r\phi_{ir}\right], \tag{5}$$

$$vech(\boldsymbol{\Lambda}_i^*|\bullet) \sim N\left(\boldsymbol{\mu}_i, \boldsymbol{S}\right), \tag{6}$$

where

$$
\begin{aligned}
\mu_i &= \boldsymbol{S}\boldsymbol{D}svec\left(\boldsymbol{P}_i\right), \quad \boldsymbol{S} = \left(\frac{\boldsymbol{D}(\boldsymbol{Q} \otimes_s \boldsymbol{Q})\boldsymbol{D}}{\sigma^2} + \boldsymbol{D}(\boldsymbol{\Psi} \otimes_s \boldsymbol{\Psi})\boldsymbol{D}\right)^{-1}, \\
\boldsymbol{Q} &= \boldsymbol{B}^{*\prime}\boldsymbol{B}^*, \ \boldsymbol{P}_i = \frac{1}{\sigma^2}\boldsymbol{B}^{*\prime}\boldsymbol{L}_i\boldsymbol{B}^* + \boldsymbol{\Psi}\Delta_i^*\boldsymbol{\Psi}, \\
\boldsymbol{D} &= \text{diag}(\underbrace{1, \sqrt{2}, \cdots, \sqrt{2}}_{R \text{ elements}}, \underbrace{1, \sqrt{2}, \cdots, \sqrt{2}}_{(R-1) \text{ elements}}, \cdots, \underbrace{1}_{1 \text{ element}}).
\end{aligned}
$$

The regression parameters are sampled from the following posterior:

$$
vec(\boldsymbol{\Gamma}^*|\bullet) \sim N\left(\boldsymbol{\mu}_\gamma, \boldsymbol{S}_\gamma\right), \tag{7}
$$

where

$$
\boldsymbol{\mu}_\gamma = \boldsymbol{S}_\gamma \sum_{i=1}^{n} \left(\boldsymbol{x}_i' vech(\boldsymbol{\Psi}\boldsymbol{\Lambda}_i^*\boldsymbol{\Psi})'\right), \quad \boldsymbol{S}_\gamma = \left(\sum_{i=1}^{n} \{(\boldsymbol{D}(\boldsymbol{\Psi} \otimes_s \boldsymbol{\Psi})\boldsymbol{D}) \otimes \boldsymbol{x}_i\boldsymbol{x}_i'\} + \frac{1}{\sigma_\gamma^2}\boldsymbol{I}\right)^{-1}.
$$

The hyper-parameters are sampled based on the following posterior distributions:

$$
\begin{aligned}
\phi_{ir} &\sim Inverse - Gaussian\left(\sqrt{\frac{a_{0r}}{\tau_r \beta_{ir}^{*\,2}}}, a_{0r}\right) && (8) \\
\tau_r &\sim Gamma\left(\frac{V-r+1}{2}, \frac{\sum_{l=r}^{V} \phi_{lr}\beta_{lr}^{*\,2}}{2}\right) && (9) \\
a_{0r} &\sim Gamma\left(a_1 + (V-r+1), a_2 + \frac{\sum_{l=r}^{V} 1/\phi_{lr}}{2}\right) && (10)
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &\sim IG\left(b_1 + \frac{nV(V+1)}{4}, \frac{1}{2}\text{tr}\left(\sum_{i=1}^{n}(\boldsymbol{L}_i - \boldsymbol{B}^*\boldsymbol{\Lambda}_i^*\boldsymbol{B}^{*\prime})^2\right) + b_2\right) && (11) \\
\sigma_\gamma^2 &\sim IG\left(c_0 + \frac{pR(R+1)}{4}, \frac{1}{2}\text{tr}\left(\boldsymbol{\Gamma}^{*\prime}\boldsymbol{\Gamma}^*\right) + c_1\right), && (12)
\end{aligned}
$$

and we employ slice sampling to get a posterior sample from

$$
(\psi_, \cdots, \psi_R) \propto \pi(\boldsymbol{\Psi})\pi(\boldsymbol{\Lambda}_1^* \cdots, \boldsymbol{\Lambda}_n^*|\boldsymbol{\Psi}). \tag{13}
$$

In summary, the posterior sampling proceeds as follows.

1. Employ slice sampling to get a posterior sample from $\pi(\boldsymbol{\beta}_g^*|\boldsymbol{B}_{(-g)}^*, \bullet)$ in (5) after a certain amont of burn-in iterations.

2. Update $\boldsymbol{\Lambda}_i^*$ generated from $N(\boldsymbol{\mu}_i, \boldsymbol{S})$ in (6).

3. Update $\boldsymbol{\Gamma}^*$ generated from $N\left(\boldsymbol{\mu}_\gamma, \boldsymbol{S}_\gamma\right)$ in (7).

4. Update hyperparameters from (8)-(13).

The number of eigenvectors R is chosen by a Bayesian information criterion (BIC) and the reconstruction error is measured by the ratio of Frobenius norms:

$$\text{error} = \frac{1}{n}\sum_{i=1}^{n}\frac{||\boldsymbol{L}_i - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Lambda}}_i\widehat{\boldsymbol{B}}'||_F}{||\boldsymbol{L}_i||_F}. \tag{14}$$

3 Simulation study

In this section, we conduct simulation studies to illustrate the performance of the parameter-expanded BLGRM. We assume that there are 100 subjects $i = 1, \cdots, 100$ and their response data are simulated from the underlying model $\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B} + \boldsymbol{\epsilon}_i$, where $\boldsymbol{L}_i$ denotes any symmetric matrix response. We applied our proposed method to estimate $\boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n$ and to recover $\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n$. We run 5,500 MCMC iterations with 500 burn-in. We repeated the simulation 50 times under four different scenarios.

In section 3.1, we examine if the true number of eigenvectors is correctly chosen by BIC. We investigate two different scenarios: there is a common basis $\boldsymbol{B}$ with rank 3 (scenario 1), and there are two different bases $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ for two subject groups, with rank 3 respectively (scenario 2). We don't include any covariates $\boldsymbol{x}_i$ in this section. In order to compare the performance of the proposed method with other competing methods, in scenario 1, we consider a frequentist version of LGRM, PARAFAC, and three-way DEDICOM, focusing on the reconstruction error defined in (14). PARAFAC and three-way DEDICOM

can be regarded to decompose a multiple (symmetric) matrices into a common basis and specific coefficients, while they impose more constraints than our method does. PARAFAC decomposes a tensor into a sum of component rank-one tensors such as

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

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

$$\boldsymbol{L}_i \approx \sum_{r=1}^{R} q_{ir}(\boldsymbol{u}_r \circ \boldsymbol{u}_r) \text{ for } i = 1, \cdots, n.$$

Then $\boldsymbol{u}_r$ plays a 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 so that it can incorporate a third mode of the data [14] and decompose asymmetry matrices. Here, we only consider symmetric $\boldsymbol{L}_i \in \mathbb{R}^{V \times V}$ for comparison. Then the model can be written as

$$\boldsymbol{L}_i \approx \boldsymbol{A}\boldsymbol{D}_i\boldsymbol{Q}\boldsymbol{D}_i\boldsymbol{A}' \text{ for } i = 1, \cdots, n,$$

where latent components $\boldsymbol{A} \in \mathbb{R}^{V \times R}$, interaction between different components $\boldsymbol{Q} \in \mathbb{R}^{R \times R}$, a diagonal matrix $\boldsymbol{D}_i \in \mathbb{R}^{R \times R}$. The r-th diagonal element in $\boldsymbol{D}_i$ 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 $\boldsymbol{\Lambda}_i$ can be decomposed as $\boldsymbol{D}_i\boldsymbol{Q}\boldsymbol{D}_i$, then the center parts of two decomposition models are equivalent. Thus, DEDICOM is a more constrained version of our decomposition model. We use a Python module "scikit-tensor" to calculate three-way DEDICOM available in $https://github.com/mnick/scikit-tensor$. For calculation of PARAFAC decomposition, we used the N-way Toolbox in MATLAB.

The frequentist LGRM estimates the parameter matrices in (2) by minimizing the Frobenius norm of $\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'$ using iterative optimization steps. This optimization prob-

lem is solved by adapting low rank approximation techniques proposed by Ye [34]. In detail, we consider the following optimization problem

$$\min_{\boldsymbol{B},\boldsymbol{\Lambda}_i} \sum_{i=1}^{n} ||\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'||_F^2 \text{ such that } \boldsymbol{B}'\boldsymbol{B} = I_R. \tag{15}$$

Then,

$$\widehat{\boldsymbol{\Lambda}}_i = \widehat{\boldsymbol{B}}\boldsymbol{L}_i\widehat{\boldsymbol{B}}',$$

where the eigenmap $\boldsymbol{B}$ can be estimated from the following iterative optimization steps. Under $\boldsymbol{\Lambda}_i = \boldsymbol{B}\boldsymbol{L}_i\boldsymbol{B}'$, the above minimizing problem is equivalent to maximizing

$$\begin{aligned}\sum_{i=1}^{n} ||\boldsymbol{B}'\boldsymbol{L}_i\boldsymbol{B}||_F^2 &= \sum_{i=1}^{n} tr(\boldsymbol{B}'\boldsymbol{L}_i\boldsymbol{B}\boldsymbol{B}'\boldsymbol{L}_i\boldsymbol{B}) \\ &\approx \sum_{i=1}^{n} tr(\boldsymbol{B}'\boldsymbol{L}_i\boldsymbol{B}_0\boldsymbol{B_0}'\boldsymbol{L}_i\boldsymbol{B}),\end{aligned}$$

where $\boldsymbol{B}_0$ is the $\boldsymbol{B}$ matrix from the previous iteration. Then $\boldsymbol{B}$ can be estimated by the following iterative steps:

1. Let $\boldsymbol{B}_0$ be the $\boldsymbol{B}$ matrix from the previous iteration.
2. Calculate $\boldsymbol{Q} = \sum_{i=1}^{n} \boldsymbol{L}_i\boldsymbol{B}_0\boldsymbol{B}_0{}'\boldsymbol{L}_i$.
3. Compute the R eigenvectors $\{\phi_i\}_{i=1}^{R}$ of $\boldsymbol{Q}$ corresponding to the largest R eigenvalues.
4. Set $\boldsymbol{B} = [\phi_1, \phi_2, \cdots, \phi_R]$.
5. Repeat the above iterations until it converges.
6. Calculate $\widehat{\boldsymbol{\Lambda}}_i = \widehat{\boldsymbol{B}}\boldsymbol{L}_i\widehat{\boldsymbol{B}}'$, for $i = 1, \cdots, n$.

Also, the variance component σ^2 can be estimated by the MLE:

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

Section 3.2 discusses the accuracy of estimation for regression coefficients in BLGRM. We consider binary (scenario 3) and continuous (scenario 4) covariates with an intercept. Reconstruction errors are calculated as well.

3.1 Simulation 1

A. Scenario 1

We generate all simulation data sets from $\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i$. There exist 3 true underlying common eigenmaps ($R = 3$). We assume that $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$ are 50×3 and 3×3 matrices, respectively. The eigenmaps, subject-specific coefficient matrices, and measurement errors are generated according to:

$$\begin{aligned} \beta_{lk}^{g} &\sim N(0,1),\ l = 1, \cdots, 50,\ k = 1, \cdots, R, \\ \pi(\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\epsilon}_i^2\right)\right],\ \pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) = \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}_i^2\right)\right]. \end{aligned} \tag{16}$$

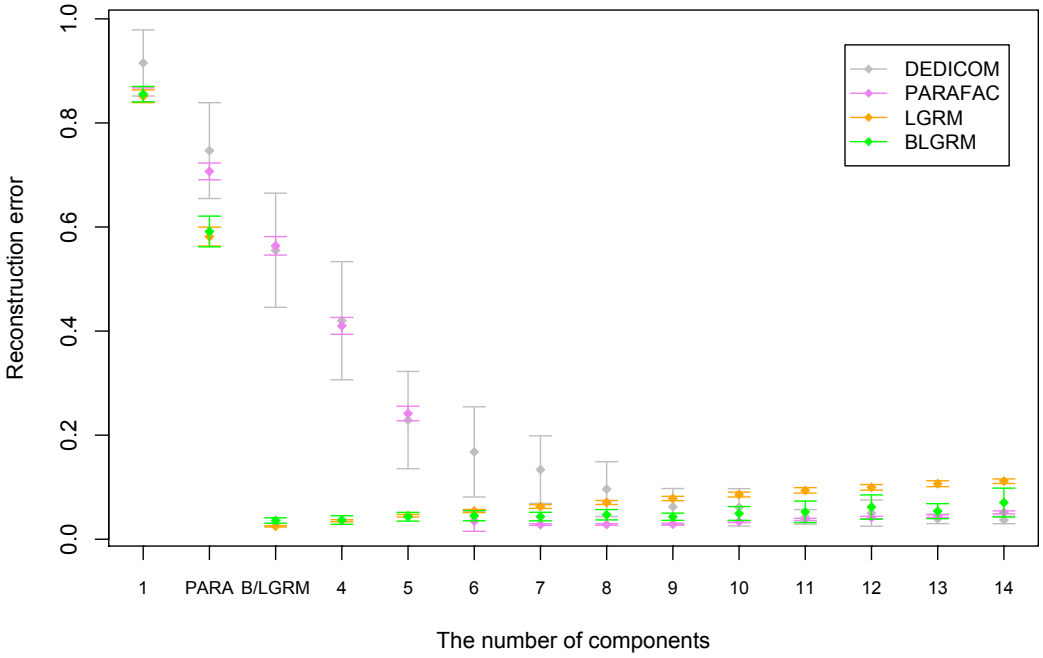
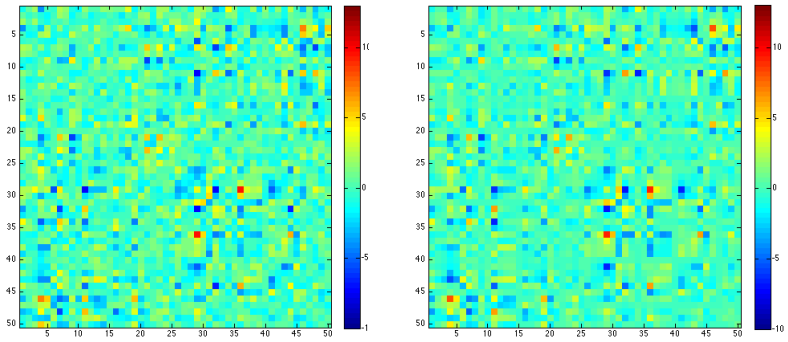


Figure 1: Panel (a) shows the first raw data matrix $\boldsymbol{L}_1$ in the first simulation data set, while panel (b) shows the approximated matrix by the proposed method.



(a) Raw structure (b) Approximated structure

Figure 2: Panel (a) shows the first raw data matrix $\boldsymbol{L}_1$ in the first simulation data set, while panel (b) shows the approximated matrix by the proposed method.

In order to select the number of common eigenvectors in BLGRM and LGRM, we used BIC. BIC perfectly selects the true number of eigenvectors, 3. Core consistency diagnostic has been used for choosing the proper number of components for the PARAFAC model [2]. For the three-way DEDICOM, there is no specific way to determine the number of components to examine approximation performances. We plot the reconstruction errors as increasing the number of eigenvectors (components). Figure 1 shows the simulation results based on 50 replications. The (Monte Carlo) error bars are depicted with the Monte Carlo mean of reconstruction errors. Core consistency diagnostic gives 2 as median of the adequate numbers of components for PARAFAC based on 50 repetitions, while 3 is always chosen by BIC for BLGRM and LGRM. One can see that the BLGRM and LGRM approximate the raw data matrices fairly well with the true number of eigenvectors, while the reconstruction errors tend to be slightly increasing as the number of eigenvectors increases. The PARAFAC model achieves satisfactory approximation performance when the number of components is 6, while its core consistency diagnostic recommends to use 2. On the other

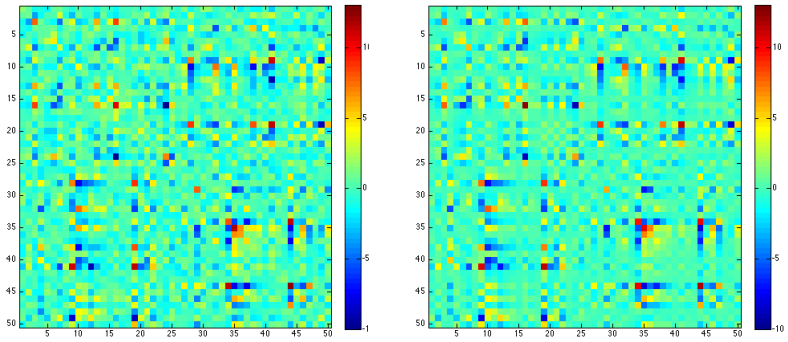
hand, the core consistency chosen model of PARAFAC gives about 0.7 reconstruction error in average, which is much larger than those BIC chosen BLGRM and LGRM give. And the reconstruction errors of three-way DEDICOM gradually decrease as the model complexity increases. Also, its variability is very high compared to the other competing methods as suggesting its unstable performance. BLGRM performs better as the decomposition model gets more complex than LGRM does. MCMC iterations and burn-in in slice sampling could provide BLGRM more opportunity for better approximation. Reconstruction error and Figure 2 demonstrate that our decomposition method can approximate the raw matrix well.

B. Scenario 2

We generate other simulation data sets from $\boldsymbol{L}_i^g = \boldsymbol{B}_g \boldsymbol{\Lambda}_i \boldsymbol{B_g}' + \boldsymbol{\epsilon}_i,\ g = 1, \cdots, G,\ i = 1, \cdots, n_g$, where $n = \sum_{g=1}^{G} n_g$. We consider $\sigma = 1$, $R = 3$ and two groups, i.e., $G = 2$. $\boldsymbol{B}_g = [\boldsymbol{\beta}_1^g, \cdots, \boldsymbol{\beta}_R^g]$ and $\boldsymbol{\Lambda}_i$ are generated from

$$\begin{aligned} \beta_{lk}^g &\sim N(0,1),\ l = 1, \cdots, 50,\ k = 1, \cdots, R, \\ \pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}_i^2\right)\right]. \end{aligned}$$

Because we have different (independent) eigenmaps for two groups, the total number of common eigenvectors across subjects must be 6.



(a) Raw structure (b) Approximated structure

Figure 3: Figure (a) shows the first raw data matrix $\boldsymbol{L}_i$ in the first simulation data set, while figure (b) shows the approximated matrix by the proposed method.

The true number of eigenvectors is perfectly recovered by BIC. The Monte Carlo mean and standard deviation of reconstruction errors are given by 0.059 and 0.021 respectively. Figure 3 demonstrate that our decomposition method approximates the raw matrix well, while it has a slightly larger reconstruction error (Monte Carlo mean=0.036, Monte Carlo s.d.=0.010) than the scenario 1.

3.2 Simulation 2

This simulation study aims to examine if the coefficient values are estimated well by the proposed method. We measure the estimation error for $\boldsymbol{\Gamma}_j$ $(j = 1, 2)$ in the original space, i.e., estimation error is defined by $\text{error}_{\gamma_j} = \frac{||\boldsymbol{B}\boldsymbol{\Gamma}_j\boldsymbol{B}' - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}_j\widehat{\boldsymbol{B}}'||_F}{||\boldsymbol{B}\boldsymbol{\Gamma}_j\boldsymbol{B}'||_F}$. Also we check convergence of MCMC outputs by using trace plots.

We generate all simulation data sets from $\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i$, where $R = 3$. Set $\sigma = 1$

Table 1: Estimation errors for $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ in the two different cases. They are mean of estimation errors from 50 replications and the corresponding Monte Carlo standard deviations are recorded in the parentheses.

	Scenario 3	Scenario 4
error_{γ_0}	0.074 (0.016)	0.038 (0.012)
error_{γ_1}	0.141 (0.036)	0.105 (0.040)

$\boldsymbol{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_R]$ and $\mathbf{\Lambda}_i$ are generated from

$$\begin{aligned} \beta_{lk}^g &\sim N(0,1),\ l = 1, \cdots, 50,\ k = 1, \cdots, R, \\ \pi(\mathbf{\Lambda}_1, \cdots, \mathbf{\Lambda}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\text{tr}\left((\mathbf{\Lambda}_i - \mathbf{\Delta}_i)^2\right)\right]. \end{aligned}$$

Set the regression coefficients as

$$\mathbf{\Gamma}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{\Gamma}_2 = \begin{bmatrix} 0 & 4 & 0 \\ 4 & 0 & 4 \\ 0 & 4 & 0 \end{bmatrix}.$$

A. Scenario 3

The scenario 3 considers a binary covariate $\boldsymbol{x}_i = [1,\ x_{1i}]$, where $x_{1i} \sim B\left(\frac{1}{2}\right)$ and calculate $vech(\mathbf{\Delta}_i) = \mathbf{\Gamma}'\boldsymbol{x}_i,\ \ \mathbf{\Gamma} = [vech(\mathbf{\Gamma}_1),\ vech(\mathbf{\Gamma}_2)]$,

B. Scenario 4

In this scenario we consider a continuous covariate, $\boldsymbol{x}_i = [1,\ x_{1i}]$, $x_{1i} \sim N(0.5, 1)$, where $vech(\mathbf{\Delta}_i) = \mathbf{\Gamma}'\boldsymbol{x}_i,\ \ \mathbf{\Gamma} = [vech(\mathbf{\Gamma}_1),\ vech(\mathbf{\Gamma}_2)]$.

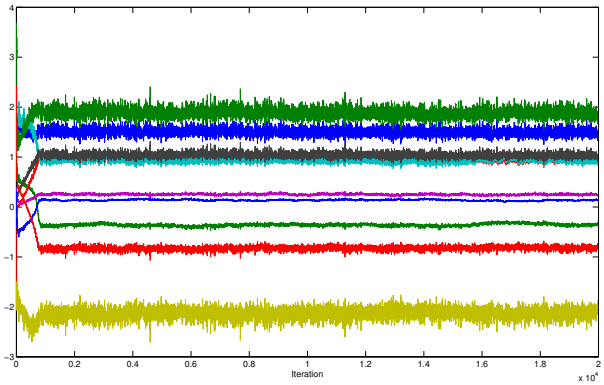


Figure 4: Trace plot for posterior samples of $\boldsymbol{B}$ elements. It shows rapid and nice mixing and convergence of MCMC chains.

Table 1 shows estimation errors for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ in the two different cases. They are mean estimation errors with 50 replications with the corresponding standard errors in the parentheses. One can see that the coefficients are estimated satisfactorily. Figure 4 shows rapid and nice mixing and convergence of MCMC chains for randomly chosen 10 elements in $\boldsymbol{B}$ matrix in the scenario 4.

4 Application to Alzheimer's Disease

Altered brain connectivity has been considered as a critical factor to explain cognitive decline in Alzheimer's disease (AD) [5, 6]. It has been reported that some regions in mild AD brains have abnormal functional connectivity with other brain regions including medial prefrontal cortex (MPFC), ventral anterior cingulate cortex (vACC), right inferotemporal cortex, right cuneus extending into precuneus, left cuneus, right superior and middle temporal gyrus and posterior cingulate cortex (PCG or PCC) through a seed based approach using Fisher's z-transformation and t-tests [31]. Wang et al. [30] showed that AD patients had decreased connectivity between prefrontal and parietal lobes and increased within-

lobe functional connectivity by whole brain ROI based t-tests. But they conducted 6670 (116*115/2) multiple testing at 0.01 significance level without any correction, their results therefore suffered from severe false positive problems.

To overcome limitations of the current methods, we applied our proposed method and tested group differences among normal control (NC), mild cognitive impairment (MCI), and AD patient groups. Instead of the raw correlation matrices, we used Fisher's z-transformed correlation matrices, which would make the data to be more suitable to our model assumptions. The transformed correlation coefficient ranges from negative infinitiy to positive infinity, while the original sign is preserved. We considered 4 covariates and the intercept: gender (γ_1), age(γ_2), MCI=1(γ_3), and AD=1(γ_4). By using HPD intervals of γ_3, γ_4, $\gamma_4 - \gamma_3$, we determine if there is any group difference in their connectivity patterns.

4.1 Data Acquisition and Pre-processing

For resting state fMRI, the imaging protocol is Field Strength=3.0 tesla; Flip Angle=80.0 degree; Manufacturer=Philips Medical Systems; Matrix X=64.0 pixels; Matrix Y=64.0 pixels; Mfg 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; TE=30.001 ms; TR=3000.0 ms; to obtain 140 volumes.

The fMRI data was pre-processed with the following steps: 1) discarding the first 10 time points, 2) slice timing, 3) head motion correction, 4) intensity scaling of each fMRI scan after motion correction to yield a whole-brain mean value of 10000, 5) temporally band-pass filtering (0.01 Hz-0.08 Hz), 6) regression out of a set of nuisance signals including signal averaged over the white matter, signal averaged over the cerebrospinal fluid, global signal averaged over the whole-brain, and six motion parameters, 7) nonlinear normalization into Montreal Neurological Institute (MNI) space with resolution 333mm^3 using SPM8; 8) spatially smoothing with a 6 mm full width at half maximum Gaussian kernel. The

Table 2: We summarized demographic information at the baseline of the 153 ADNI subjects. For an age variable, its mean and standard deviation (mean ± sd) were presented. For a binary variable, gender, the count and the percentage (in parentheses) of male subjects were 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

nonlinear normalization of fMRI data was implemented using DARTEL of SPM8 with the deformation fields of their co-registered T1-weighted images.

The data used in this study was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI study has aimed to detect and monitor the early stage of Alzheimer's disease (AD) by investigating serial magnetic resonance imaging (MRI), positron emission tomography (PET), genetic, biochemical biomarkers, and neuropsychological and clinical assessment. For up-to-date information, see www.adni-info.org.

We used 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 described that SMC subjects had a concern and exhibited slight forgetfulness, while their cognitive scores were within normal range. As their cognition was normal and forgetfulness was not consistent, SMC subjects were combined with CN subjects. We called this combined group "normal control (NC)" from now on. The demographic information at baseline was summarized in Table 2. In the whole data, there were 75 male and 78 female subjects. Their average age was 72.44 years with 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 6.06 years standard deviation. Among 75 MCI patients, 38

patients were male. The average age of the MCI subjects was 71.78 years with 6.81 years standard deviation. For AD patients, 12 of them were male patients, while their average age was 73.40 years with standard deviation of 7.27 years.

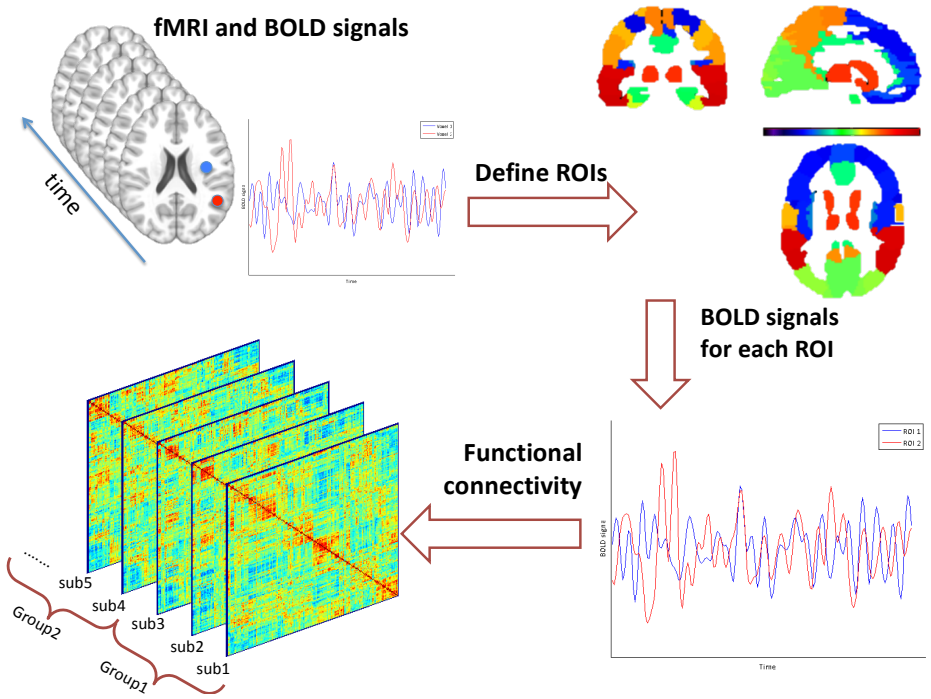
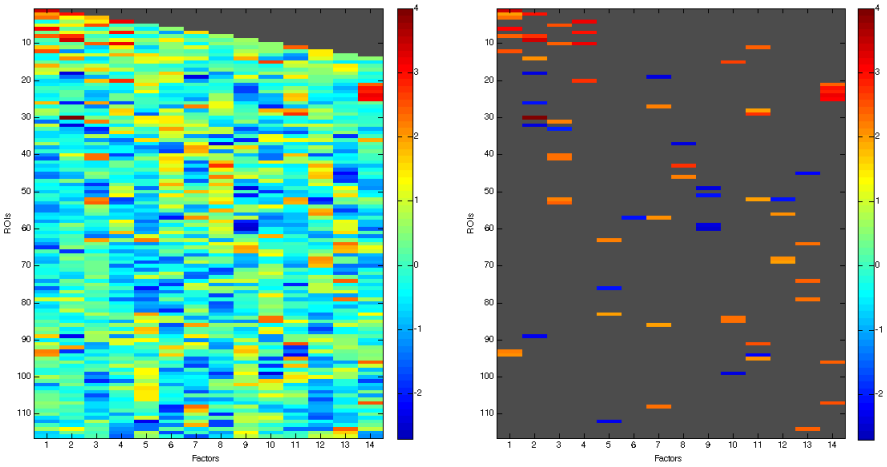


Figure 5: Process to estimate functional connectivity from resting-state fMRI data.

Figure 5 shows the overall procedure to calculate the resting-state functional connectivity from fMRI data. We used an Automated Anatomical Labeling (AAL) atlas, a widely used manual macroanatomical parcellation, and finally got 116 ROIs for a single subject. We used AFNI package of [4] to compute the average BOLD signal over a ROI of all voxel values.

4.2 Data Analysis Results



(a) Standardized (b) Thresholded

Figure 6: Figure (a) shows the estimated B matrix with standardized scale, while figure (b) shows the standardized B matrix after thresholding.

In order to find the proper number of eigenvectors, we use BIC and chose the number of eigenvectors whose corresponding BIC was the smallest. We set the number of eigenvectors, R to be 14. We run BLGRM with 5,500 MCMC iterations with 500 burn-in. We use the same setting for hyperparameters as in the simulation studies.

Figure 6 shows the estimated $\boldsymbol{B}$ matrix. For better presentation, we plotted the standardized $\boldsymbol{B}$ 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. The most of the main components (The superior parietal, supramarginal, postcentral, and inferior parietal gyrus, and the paracentral lobule) are all within the parietal lobe. The parieral lobe merges sensory

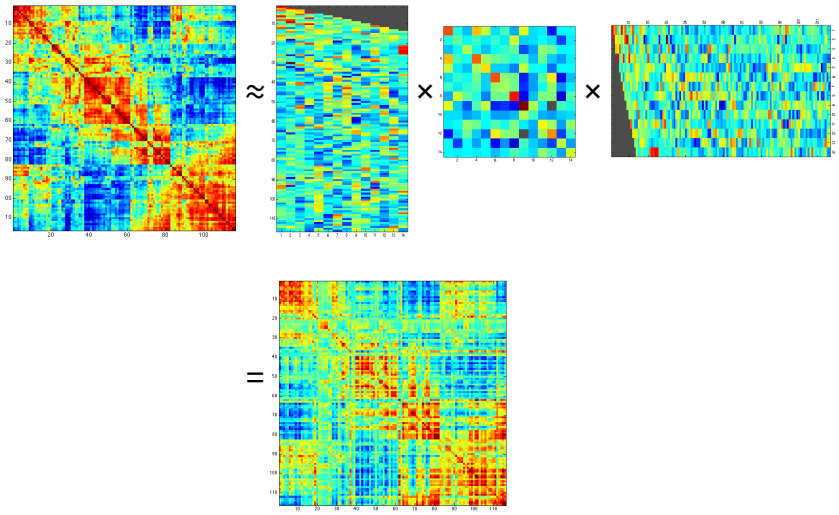


Figure 7: Shows how the raw correlation matrix of the first subject can be decomposed by the BGSC.

information among heterogeneous sources, such as proprioception, mechanoreception in the somatosensory cortex, and the dorsal stream of the visual system. A recent research finds that specific parietal regions including posterior parietal cortex contribute to retrieval tasks of episodic memory [29]. Thus, this eigenvector could represent brain functions of sensory information integration and episodic retrieval in a system level. The 3rd eigenvector has large weights on the left supplementary motor area, the left rolandic operculum, bilateral medial superior frontal, left superior frontal, left middle frontal, left triangulargyrus inferior frontal, and right superior occipital gyrus. It most consists of brain regions in pretemporal lobe. The prefrontal lobe plays a central role in cognitive control, "the ability to take charge of ones actions and direct them towards future, unseen goals" [26]. Also its impairment is replicated to be associated with antisocial behavior [33]. Thus, this eigenvector describes individual's cognitive control and behavior. The main components of the 13th eigenvector are the bilateral precuneus and the bilateral posterior cingulate gyrus. This eigenvector

represents default mode network, because PCG/PCUN and angular gyrus are functional hubs of default mode network, which is disrupted in people with AD, and autism spectrum disorder [1].

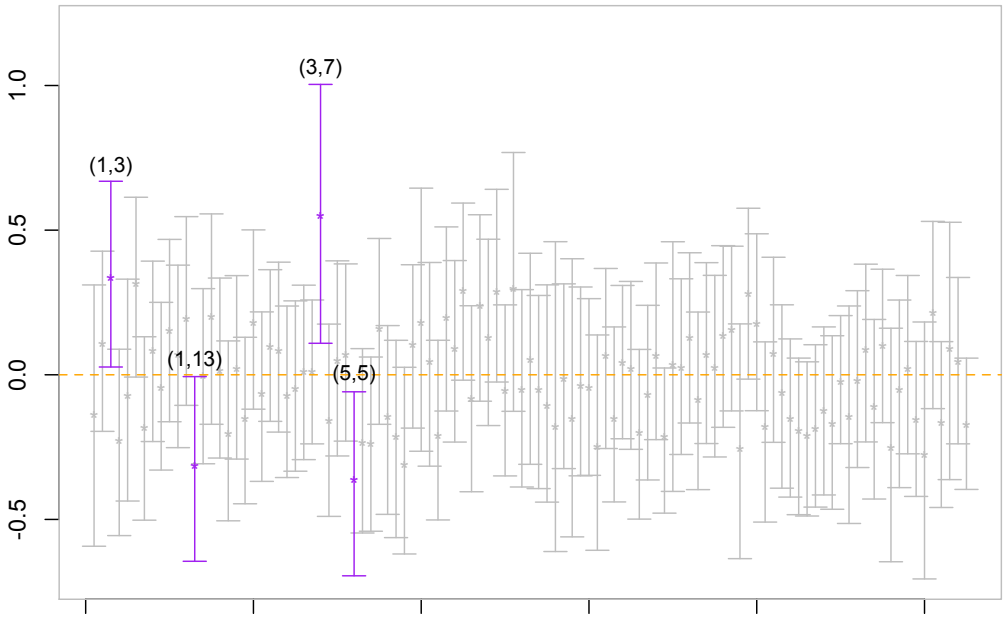


Figure 8: Shows the posterior means marked by "$*$" and the corresponding 95% HPD intervals (each vertical line) of $vech(\Gamma_4 - \Gamma_3)$. There are 4 pairs of eigenvectors that have important group differences, as their HPD intervals do not include 0. See the purple lines.

Figure 7 shows how well the raw correlation matrix of the first subject is approximated by the proposed method (with the total reconstruction error=0.67). We also examine the estimated regression coefficient $\Gamma_4 - \Gamma_3$ that represents the effect of a difference of MCI and AD patients on their functional connectivity maps. Table 8 shows the estimated regression coefficients by the posterior means and the correspondig 95% HPD intervals. We could identify important group differences of some connectivity in the eigenspace by examining the HPD intervals. There are 4 HPD intervals that do not include 0. It suggest us that the

4 pairs of eigenvectors have important group differences. Based on our interpretation of the estimated $\boldsymbol{B}$ matrix, for example of the pair of the 1st and 3rd eigenvectors, MCI and AD patients have different functional connectivty among brain regions of sensory information integration/episodic retrieval and cognitive control/behavior. For the pair of the 1st and 13th eigenvectors, there is a group difference in terms of connections among sensory information integration/episodic retrieval regions and default mode network.

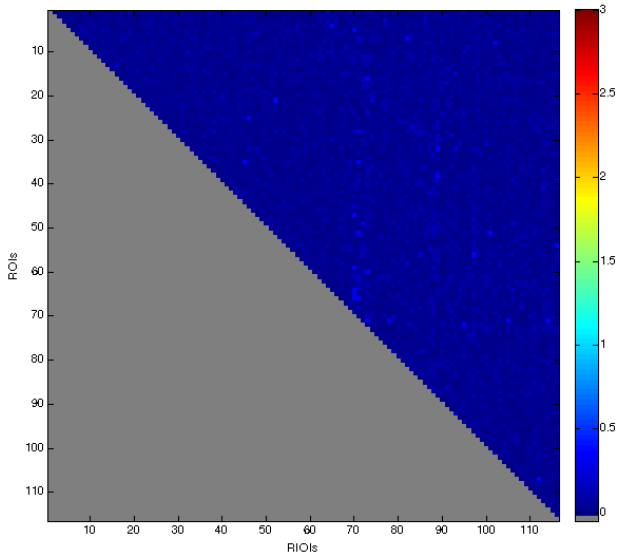


Figure 9: $-\log_{10}(\text{p})$ of univariate (ROI-wise) two-sample t-test for AD and MCI groups.

On the other hand, we could not find any significant group differences for AD and MCI patients by using ROI-wise univariate two-sample t-test for correlation coefficients after Fisher's z-transformation. Figure 9 presents $-\log_{10}(\text{p})$ of the two-sample t-tests for AD and MCI groups. After Bonferroni correction, $-\log_{10}(\text{p})$ should be greater than 5.13 at 0.05 significance level, or it should be greater than 1.30 without any multiple testing correction. But we could not identify any significance group difference with/without correction.

In order to see which ROIs have functional connectivity differences among MCI, and

AD, we map estimated coefficients $\widehat{\mathbf{\Gamma}}$ from the eigenspace to the original ROI space. We use $\widehat{\boldsymbol{B}}$ and "sparse" $\widehat{\mathbf{\Gamma}}$ to calculate the coefficient matrix by the following steps:

1. Estimate $\boldsymbol{B}$ and $\mathbf{\Gamma}$ by the proposed method.
2. Calculate 95% HPD interval of all the elements in $\mathbf{\Gamma}$ and find unimportant elements (pairs of brain regions) whose HPD intervals include 0.
3. Set the corresponding coefficient values to be 0. Denote the sparse $\widehat{\mathbf{\Gamma}}$ as $\widetilde{\mathbf{\Gamma}}$.
4. Use $\widehat{\boldsymbol{B}}\widehat{\mathbf{\Gamma}}\widehat{\boldsymbol{B}}'$ as estimated coefficients for the ROI space.

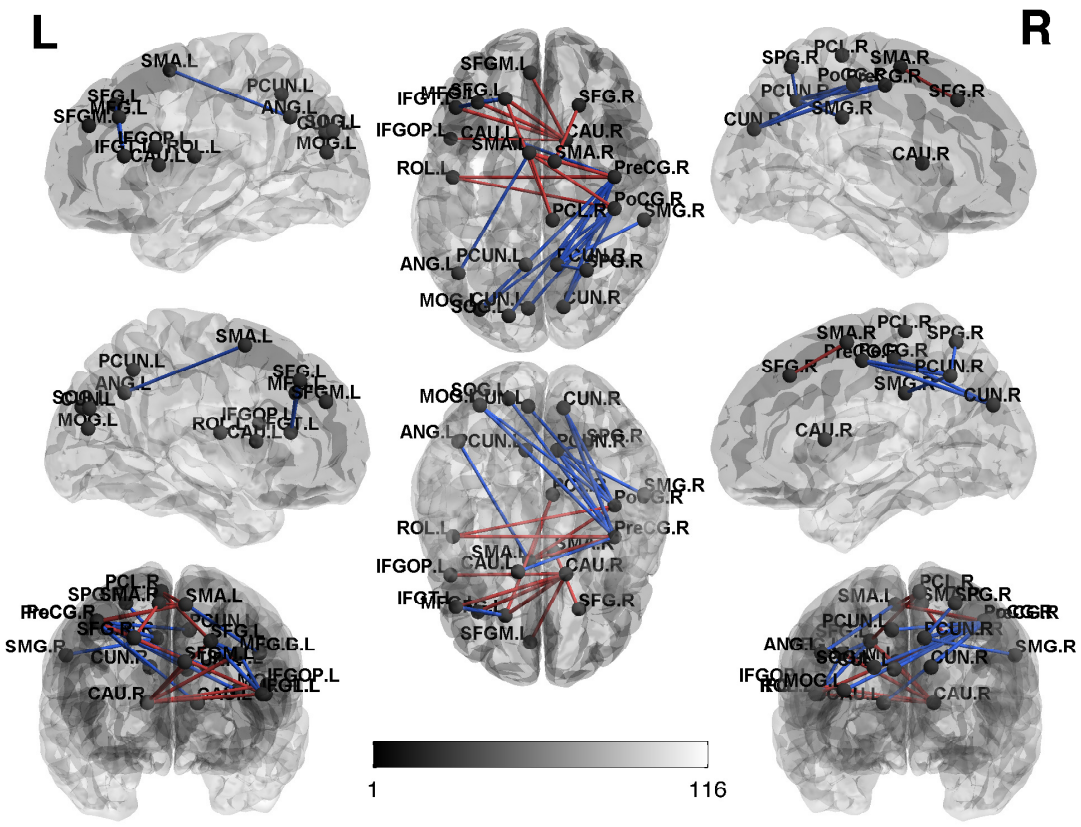


Figure 10: Shows which pairs of two ROIs have different connectivity between MCI and AD groups.

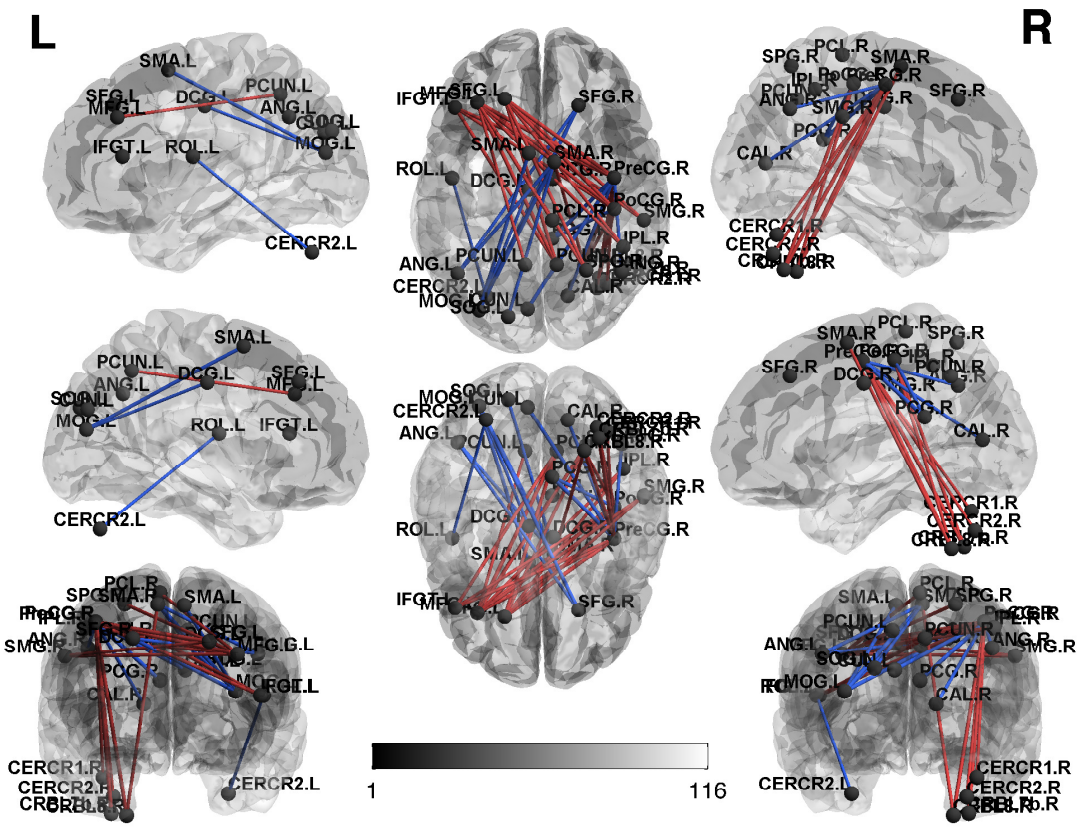


Figure 11: Shows which pairs of two ROIs have different connectivity between MCI and AD groups.

By using the above steps, we mapped the estimated coefficients from the eigenspace to the original ROI space. Then we plotted the estimated regression coefficients on the brain template using BrainNet Viewer of Xia et al. [32]. The first 5% largest effect sizes were selected to be shown in the figures. Figure 10 shows which brain regions have stronger (or weaker) positive connectivity for AD patients compared to MCI patients, while Figure 11 shows which brain regions have different negative connectivity pattern between MCI and AD groups. In Figure 10, the red line represents the estimated coefficient value $\gamma_4 - \gamma_3 > 0$, which implies that AD patients have a stronger positive connection than MCI subjects

between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_4 - \gamma_3 < 0$ indicating that AD patients have a weaker positive connection than MCI subjects between the corresponding two ROIs. For the negative connectivity, in Figure 11, the red line implies that AD patients have a stronger negative connection than MCI subjects between the corresponding two ROIs. The blue line indicates that AD patients have a weaker negative connection than MCI subjects between the corresponding two ROIs. It is observed that various brain regions have weaker or stronger connections for AD patients, which implies that AD patients would have altered functional network compared to MCI patients.

In order to examine functional connectivity differences between MCI and AD patients, we focus on Figures 10, 11, and Tables 4 and 5. Note that the regression coefficients have different interpretation depending on the sign of the average transformed correlation coefficient in MCI patients. Table 4 shows selected regression coefficient estimates where the average of transformed correlation coefficient in MCI patients is positive, while Table 5 presents selected coefficient estimates in the negative cases. We list some of regression coefficients whose sizes are the top 5%.

Figure 4 tells us MCI and AD patients have different positive functional connectivity among prefrontal, parietal, PCG/PCUN brain regions. The bilateral PCUN has weaker positive connections between the right precentral gyrus, the right postcentral gyrus, the right supramarginal gyrus, the right superior parietal gyrus, and the right paracentral lobe for AD patients than MCI patients. AD patients have lower positive connectivity among the right PCG and the right supramarginal gyrus gyrus, the right paracentral lobule, and the right supplementary motor area than MCI patients do. Furthermore, the bilateral PCUN has weaker negative connections between the right precentral gyrus, the right postcentral gyrus, the right supramarginal gyrus, the right superior parietal gyrus, and the left brain regions in prefrontal lobe for AD patients than MCI patients. It implies disrupted

connections for AD patients among default mode network, sensory information integration , episodic memory retrieval, and cognitive behavior.

From Figure 5, one can observe that AD patients have altered negative connections among prefrontal, parietal brain areas. We finds that AD patients have weaker negative connections among the left superior frontal/middle frontal gyrus and parietal lobe including postcentral, supramarginal, inferior parietal gyrus and paracentral lobule. It suggests that AD patients have weaker negative connectivity between sensory information integration/ episodic memory retrieval and cognitive behavior/control.

On the other hand, our study suggests that there is a connectivity difference of paracentral gyrus with other brain regions including superior, middle, inferior frontal gyri (see Tables 4 and 5). A fMRI study by Mason et al. [25] showed that, when stimuli from some senses are deliberately reduced or removed, the mind recruits some brain regions including the medial posterior cingulate, PCUN, PCL, the inferior frontal cortices, superior and middle frontal gyri, and a cluster spanning dorsal medial frontal regions. These distributed foci have temporal coherence and constitute a tightly coupled, organized neural network. It may indicate that brain network differences between MCI and AD patients can be explained by certain brain reaction to sensory deprivation.

5 Discussion

In this study, we proposed a BLGRM by taking a global approach to analyze brain functional connectivity. It 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 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 $\boldsymbol{\Lambda}_i$, so that we could automatically estimate the effects of covariates on the $\boldsymbol{\Lambda}_i$ matrices within MCMC iterations. Furthermore, the parameter-expansion approach was taken on our graph regression model to reduce posterior dependence between $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$'s.

The simulation studies demonstrated that our method efficiently approximated the raw symmetric matrix data with good MCMC mixing properties. We compared reconstruction errors of our method with those of LGRM, PARAFAC, and three-way DEDICOM. Our method efficiently approximated the raw data matrices by a fewer number of eigenvectors than the other competing methods. Also the regression parameters for $\boldsymbol{L}_i$ could be recovered satisfactorily.

The ADNI real data analysis revealed that the bilateral PCUN/PCG had weaker connections between parietal/pretemporal lobe for AD patients than MCI patients. It suggests that AD patients have disrupted connections among brain regions involving default mode network, sensory information integration, episodic memory retrieval, and cognitive behavior. Also, there was a connectivity difference of paracentral gyrus with other brain regions including superior, middle, inferior frontal gyri. This finding agreed with other functional connectivity studies by [3, 21, 25, 31].

This study could provide a guideline to elucidate hidden pathology of neurological disorders in a brain connectivity perspective. In clinical application, our method can be used to (1) examine if normal subjects and patients (or among disease subtypes) have a different functional connectivity structure and where the difference comes from. This exploratory analysis allows better understanding of the underlying mechanism of a disorder, which further may help to develop future treatments targeting some identified brain regions. Also, altered connectivity can provide (2) diagnostic and prognostic information [7]. 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 disorder. For example of ADHD, a subject is diagnosed as ADHD if the subject meets the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria. The test result can be varied depending on interview environment, examiners and examinees. And it is still subjective criteria. If abnormality of functional connectivity is a very early sign to develop a disorder or 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 [19]. While genetic factors are emerging in psychiatric disorder research, their effect size is very small and their working mechanism still needs to be elucidated. Because brain organization and function are influenced by genetic factors [20, 22] and brain information has relatively large impact on disease progression, functional connectivity can be a mediator to explain connection among genetic mutations and disorders.

References

[1] J. R. Andrews-Hanna, J. S. Reidler, J. Sepulcre, R. Poulin, and R. L. Buckner. Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4):550–562, 2010.

[2] R. Bro and H. A. Kiers. A new efficient method for determining the number of components in parafac models. *Journal of chemometrics*, 17(5):274–286, 2003.

[3] 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.

[4] R. W. Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.

[5] 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):1–16, 1993.

[6] X. Delbeuck, M. Van der Linden, and F. Collette. Alzheimer'disease as a disconnection syndrome? *Neuropsychology review*, 13(2):79–92, 2003.

[7] M. D. Fox and M. Greicius. Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, 4:19, 2010.

[8] K. Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS biol*, 7(2):e1000033, 2009.

[9] 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.

[10] A. Gelman et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534, 2006.

[11] J. Geweke and G. Zhou. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9(2):557–587, 1996.

[12] J. Ghosh and D. B. Dunson. Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320, 2009.

[13] 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.

[14] 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.

[15] P. D. Hoff. A hierarchical eigenmodel for pooled covariance estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):971–992, 2009.

[16] S. A. Huettel, A. W. Song, and G. McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[17] 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.

[18] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[19] 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(6):904–916, 2010.

[20] 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(5922):1737–1740, 2009.

[21] R. Leech and D. J. Sharp. The role of the posterior cingulate cortex in cognition and disease. *Brain*, 137(1):12–32, 2014.

[22] 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(1):163–174, 2009.

[23] C. Liu, D. B. Rubin, and Y. N. Wu. Parameter expansion to accelerate em: The px-em algorithm. *Biometrika*, 85(4):755–770, 1998.

[24] J. S. Liu and Y. N. Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

[25] M. F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae. Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810):393–395, 2007.

[26] 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(1424):1123–1136, 2002.

[27] T. Park and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[28] A. Schwartzman. *Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, 2006.

[29] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner. Parietal lobe contributions to episodic memory retrieval. *Trends in cognitive sciences*, 9(9):445–453, 2005.

[30] 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.

[31] 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(2):496–504, 2006.

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

[33] 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(2):81–88, 2009.

[34] J. Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61(1-3): 167–191, 2005.

Table 3: AAL parcellation of the entire brain and their 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
IFGOP	Inferior frontal gyrus, opercular	Prefrontal lobe
IFGT	Inferior frontal gyrus, triangular	Prefrontal lobe
PreCG	Precentral gyrus	Other frontal
SMA	Supplementary motor area	Other frontal
ROL	Rolandic operculum	Other frontal
ANG	Angular gyrus	Parietal lobe
IPL	Inferior parietal gyrus	Parietal lobe
PCG	Posterior cingulate gyrus	Parietal lobe
PCL	Paracentral lobule	Parietal lobe
PCUN	Precuneus	Parietal lobe
PoCG	Postcentral gyrus	Parietal lobe
SMG	Supramarginal gyrus	Parietal lobe
SPG	Superior parietal gyrus	Parietal lobe
CUN	Cuneous	Occipital lobe
SOG	Superior occipital gyrus	Occipital lobe
MOG	Middle occipital gyrus	Occipital lobe
CAU	Caudate nucleus	Corpus striatum
CRBL	Cerebellum	Cerebellum

Table 4: Shows the effect of a group difference between MCI and AD on the following pairs of brain regions (Regions 1 and 2). The average of transformed correlation coefficients in the MCI group is positive.

Region1	Region2	Coeff	Region1	Region2	Coeff
SFGM.R	PreCG.R	0.064	ANG.L	SMA.L	-0.075
SFGM.R	Cau.R	0.069	ANG.L	PCG.R	0.065
SFGM.L	SFG.L	-0.063	PCL.R	CUN.R	-0.068
SFGM.L	PreCG.R	0.063	PCL.R	MOG.L	-0.065
SFGM.L	Cau.R	0.078	PCL.R	CUN.L	-0.065
PCUN.R	PreCG.R	-0.089	PCL.R	PCUN.R	-0.063
PCUN.R	PoCG.R	-0.083	PCL.R	SOG.L	-0.060
PCUN.R	SMG.R	-0.074	PCL.R	CRBL8.L	0.064
PCUN.R	SPG.R	-0.070	PCL.R	CRBL7b.R	0.064
PCUN.R	SFGM.L	0.063	PCL.R	IFGT.L	0.064
PCUN.R	IPL.R	-0.065	PCL.R	ROL.L	0.067
PCUN.L	PreCG.R	-0.073	PCL.R	SMA.L	0.071
PCUN.L	SFG.L	0.059	PCL.R	SFG.L	0.073
PCG.R	SMG.R	-0.065	SMA.R	IPL.L	-0.059
PCG.R	SMA.R	-0.067	SMA.R	SPG.L	-0.056
PCG.R	PCL.R	-0.065	SMA.L	IPL.L	-0.060

Table 5: Shows the effect of a group difference between MCI and AD on the following pairs of brain regions (Regions 1 and 2). he average of transformed correlation coefficients in the MCI group is negative.

Region1	Region2	Coeff	Region1	Region2	Coeff
PCUN.R	IFGT.L	0.079	SFG.L	PoCG.R	0.079
PCUN.R	SFG.L	0.081	SFG.L	SMG.R	0.065
PCUN.R	MFG.L	0.089	SFG.L	IPL.R	0.068
PCUN.L	MFG.L	0.071	SFG.L	SPG.R	0.059
PCG.R	PreCG.R	-0.090	MFG.L	PoCG.R	0.081
PCG.R	PoCG.R	-0.081	MFG.L	SMG.R	0.065
ANG.R	PreCG.R	-0.075	MFG.L	IPL.R	0.071
ANG.L	SMA.R	-0.081	MFG.L	SPG.R	0.059
ANG.L	SFG.R	-0.073	MFG.L	PCL.R	0.077
PCL.R	MFG.L	0.076	ANG.R	SMA.R	-0.080