Semi-parametric Bayes Conditional Graphical Models for Imaging Genetics

Article *in* Stat · May 2016

DOI: 10.1002/sta4.119

2 authors:

Suprateek Kundu
Emory University
12 PUBLICATIONS **21** CITATIONS
SEE PROFILE

Jian Kang
Emory University
58 PUBLICATIONS **291** CITATIONS
SEE PROFILE

Semi-parametric Bayes Conditional Graphical Models for Imaging Genetics

Suprateek Kundu and Jian Kang

Abstract: Motivated by the need for understanding neurological disorders, large-scale imaging genetic studies are being increasingly conducted. A salient objective in such studies is to characterize important neuroimaging biomarkers such as the brain functional connectivity, as well as genetic biomarkers which are predictive of disorders. However, typical approaches for estimating the group level brain functional connectivity do not account for potential variation resulting from demographic and genetic factors, while usual methods for discovering genetic biomarkers do not factor in the influence of the brain network on the imaging phenotype. Hence, it is of interest to develop methods to jointly estimate the brain network after accounting for heterogeneity, and infer significant genetic biomarkers. We propose a novel semi-parametric Bayesian conditional graphical model for joint covariate selection and graphical model estimation for this purpose. The proposed approach specifies novel priors on the regression coefficients and the graph space, which clusters brain regions having similar activation patterns depending on covariates, and encourages denser and sparse connections within and across clusters respectively. The method is straightforward to implement via a Markov chain Monte Carlo. We apply the approach to data obtained from the Alzheimer's Disease Neuroimaging Initiative, and demonstrate numerical advantages via simulations.

Keywords: brain functional modules; conditional graphical model; imaging genetics; semi-parametric Bayes; variable selection.

1. Introduction

During the last two decades, there have been tremendous advances in both neuroimaging and high-throughput genotyping technology, which has resulted in the development of an emergent interdisciplinary field known as imaging genetics, focusing on the genetic dissection of neuroimaging and clinical phenotypes. The goal of imaging genetics studies is to discover the brain-wide, genome-wide association patterns which drive complex neurological disorders (such as Alzheimer's disease, autism spectrum disorder, major depressive disorder, and so on). A key objective in these studies is to characterize important neuroimaging and genetic biomarkers which are predictive of psychological disorders.

One such important neuroimaging biomarker that has shown tremendous promise is the group level brain functional connectivity (Biswal et al., 1995; Smith et al., 2012; Huang et al., 2010, Kim et al., 2015), which characterizes the coherence of the neural activities among distinct brain regions for a collection of subjects. However typical approaches for estimating the group level brain network often fail to account for heterogeneity across subjects resulting from demographic, clinical and genetic variations, which may lead to spurious associations and erroneous inferences. In addition to functional connectivity, several genetic biomarkers have been shown to be predictive of neurological disorders. Significant genetic biomarkers are often inferred by modeling the association between gene products/ variants and the brain imaging phenotype (Stein et al., 2010; Zhu et al. 2014; Stingo, 2013), since some neuroimaging traits are potentially closer to the action of the gene compared to other clinical phenotypes (Mier et al., 2010; Munafo et al., 2008). However, existing approaches for detecting such associations usually do not take into account the underlying brain functional network influencing the imaging phenotype.

To our knowledge, the body of work for inferring genetic associations with the imaging phenotype to discover genetic biomarkers for neurological disorders, and the literature on estimating the brain functional connectivity, have developed in a largely independent manner. In fact there is a scarcity of approaches which can achieve the two goals simultaneously. We seek to bridge such a gap through this work, wherein we propose to jointly (a) estimate the group level brain functional network, after accounting for extrinsic sources of variation; (b) infer significant genetic and demographic associations with the imaging phenotype, leading to the discovery of important biomarkers; and (c) identify functional modules comprising brain regions having similar activation patterns influenced by covariates, and decipher the connectivity in each of these modules. Inferring functional modules is an appealing feature of the proposed approach, which is expected to provide deeper insights into the brain organization, as explained in the sequel.

A natural approach to fulfilling the stated goals is conditional graphical models, which structures the multivariate outcome as a sum of a linear term involving covariates and a Gaussian residual encapsulating the graphical structure. The estimated graph under a conditional graphical model provides a meaningful group level brain network comprising intrinsic connections after teasing out external sources of variation. Another advantage of the model is being able to compare brain networks across multiple groups, where it is imperative to account for variations due to genetic and demographic factors, in order to make the comparison across groups meaningful. This is particularly relevant for our real data application where we seek to compare the functional connectivity for subjects with Alzheimer's disease, subjects with mild cognitive disorder, and healthy individuals. In this work, we focus on sparse brain networks which is supported by findings showing that a brain

region usually interacts with only a few other regions in neurological processes (Stam et. al, 2007; Suprekar et. al, 2008).

To our knowledge, there is a limited literature for conditional graphical models, with the primary application area being essentially limited to genetic studies. Frequentist approaches such as Yin & Li (2011), Li et. al (2012), and Cai et. al (2013) mainly focus on graph estimation after adjusting for covariates. However, the performance of the variable selection by those methods is not well assessed. On the other hand, the Bayesian approach proposed by Bhadra & Mallick (2012) assumes the same inclusion status for each covariate across all nodes (imaging phenotypes in our case), which makes this approach clearly inadequate for imaging genetics applications, as evidenced in simulations.

A key challenge in conditional graphical models is achieving good variable selection and graphical model estimation simultaneously, with these two goals being closely intertwined. In particular, model mis-specification in terms of an overly sparse coefficient matrix can lead to spurious associations between nodes due to lack of adjustment for confounders (as noted in Yin & Li, 2011), while an artificially dense coefficient matrix will likely lead to over-fitting which may result in poor estimates of partial correlations (as evidenced in section 3). Ideally, a parsimonious Bayesian approach is desirable, which can provide a balance between the two goals, while providing uncertainty quantification to address the heterogeneity inherent in imaging genetics applications.

To achieve objectives (a)-(c), we propose a flexible Bayesian conditional graphical model for joint covariate selection and graphical model estimation. The proposed model clusters the columns of the regression coefficient matrix under an infinite mixture of Laplace prior, which results in dimension reduction and shrinkage. The brain functional connectivity is

estimated by a novel class of semi-parametric graphical priors depending on the unknown cluster allocations, which specify sparse associations across clusters, but allow for denser connectivity within a cluster. The approach leads to clusters of brain regions, called functional modules, characterized by similar activation patterns depending on covariates and distinct sub-networks, and is designed to mimic the brain organization suggested by several previous studies as explained below. The method is straightforward to implement via a Markov chain Monte Carlo (MCMC). Associations between the imaging phenotype and covariates are inferred via a post-MCMC approach involving multiplicity corrections, and it is also possible to determine the subset of covariates which influence a particular cluster.

The emphasis on discovering functional modules and corresponding sub-networks in goal (c) is motivated by recent studies suggesting that the brain functional connectivity pattern is highly complex such that a single cognitive function can recruit multiple distributed local clusters of neurons (Bullmore & Sporns, 2009), with different neuronal clusters potentially responsible for a large variety of brain states and functions. Moreover, widely used methods relying on independent component analysis or ICA (Guo, 2011; Guo & Tang, 2013), also yield clusters of regions of interest in the brain corresponding to different source signals, which can be likened to functional modules. Further evidence about the presence of functional modules stems from another recent work by Wang et al. (2016) demonstrating that the brain network (derived using partial correlations) can be divided into different modules, with very sparse connections across these modules but denser connections within each module. However, the above methods do not account for covariate information and hence are not strictly applicable to our problem of interest.

The paper is organized as follows. Section 2 proposes our semi-parametric conditional

graphical model and develops a posterior computation scheme, section 3 lays out our simulation study, and section 4 describes results of real data analysis using ADNI data.

2. Methodology

2.1 Semi-parametric Conditional Graphical Models

Let X and Z be the $n \times p$ and $n \times q$ dimensional outcome and covariate matrices respectively, with the i-th row of X and Z being denoted as $\mathbf{x}_i$ and $\mathbf{z}_i$, $i = 1, \ldots, n$. In our imaging genetics applications, $\mathbf{x}_i$ corresponds to the multivariate imaging phenotype, while $\mathbf{z}_i$ denotes the supplementary genetic and demographic information for the i-th individual. We assume that the rows of X have been centered, thus, it is not necessary to include an intercept term. Consider the following conditional graphical model

$$\begin{aligned} \mathbf{x}_i &= \mathbf{z}_i(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p) + \boldsymbol{\epsilon}_i,\ \boldsymbol{\epsilon}_i \sim N(0, \Sigma_G),\ \Sigma_G \sim \pi(\Sigma_G|G), \\ \boldsymbol{\beta}_k &= \sum_{l=1}^{\infty} w_l \delta_{\boldsymbol{\eta}_l}, \quad \boldsymbol{\eta}_l \sim \prod_{l=1}^{q} \mathrm{DE}(\eta_{jl}; \lambda_j), \end{aligned} \tag{1}$$

where δ_θ denotes a point mass at θ, $N(\cdot)$ and $DE(\cdot)$ denote Gaussian and double exponential/Laplace distributions respectively, $\boldsymbol{\epsilon}_i$ denotes the residual, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kq})^T$ is the vector of regression coefficients which characterizes the effect of covariates on the k-th outcome measurement ($k = 1, \ldots, p$), and $G \sim \pi(G)$ denotes the graph whose prior specification will be discussed below in (2). The prior of the covariance matrix is defined conditional on the graph G, and is also discussed in (2). We denote $B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ as the coefficient matrix, so that $\mathbf{x}_i = \mathbf{z}_i B + \boldsymbol{\epsilon}_i$ in (1).

The prior on the regression coefficients in (1) follows an infinite mixture of Laplace distributions, with the k-th component having a shrinkage parameter λ_k, and an associated

weight $w_k, , k = 1, \ldots, \infty$. The weights are structured as stick-breaking weights, so that $w_j = \nu_j \prod_{l<j}(1-\nu_l),\ \nu_l \sim Beta(1, M)$, with $\sum_{j=1}^{\infty} w_j = 1$. The mixture distribution enables dimension reduction by clustering the columns of B into distinct groups having varying degrees of shrinkage, with each cluster also translating to a group of measurements/nodes which are related to the covariates by similar magnitudes. In the special case when $\lambda_j = \lambda$ for all j, $\pi(\boldsymbol{\beta}_k), k = 1, \ldots, p$, reduces to a Dirichlet process mixture of Laplace distributions with a Laplace base measure having precision parameter M (Sethuraman, 1994). The parameter M controls the total number of clusters (H) which is random, with a high value of M implying more clusters. The ingenuity of our approach lies in proposing a novel class of semi-parametric graphical priors $\pi(G)$ in (2) which translates the parsimony implied by the clustering of the columns of B into sparsity in the precision matrix, by assuming sparse connections across distinct clusters.

Let the support of the graph space be restricted to the class of decomposable graphs $\mathcal{M}$. To construct the prior on $\mathcal{M}$, suppose there are H clusters induced under the mixture prior in (1), and denote the clusters as $S(\beta) = (S_1, \ldots, S_H)$, with S_h containing the indices of p_h nodes belonging to cluster h $(\sum_{h=1}^{H} p_h = p)$. Define the edge set E under the graph G as $E := \{e(k,l), k < l, k, l = 1, \ldots, p\}$, where $e(k,l)$ takes values 1 or 0 depending on whether the (k,l)-th edge is present in E or not. We formalize the semi-parametric graphical prior $\pi(G \mid S_1, \ldots, S_H)$, defined conditional on cluster allocations, as follows

$$
\begin{aligned}
e(k,l) &\sim \text{Ber}(\omega_1)1(\cup_{h=1}^{H}(k \in S_h, l \in S_h)) + \text{Ber}(\omega_0)1(\cup_{h \neq h'} k \in S_h, l \in S_{h'}, h \neq h'),\ k \neq l, \\
\omega_1 &\sim Be(a_{\omega,1}, b_{\omega,1}),\ \omega_2 \sim Be(a_{\omega,0}, b_{\omega,0}),\ \Sigma_G \mid G \sim HIW_G(b, D), \qquad (2)
\end{aligned}
$$

where $1(\cdot)$ is the indicator function, $HIW(b, D)$ refers to the hyper inverse-Wishart prior with scale matrix D and b degrees of freedom, and $Ber(\omega)$ denotes the Bernoulli distribution with inclusion probability ω. The scale matrix is assumed to be diagonal in our work, i.e. $D = \text{diag}(d_1, \ldots, d_p)$, with $d_j \sim \pi(d_j), j = 1, \ldots, p$. The hyper inverse-Wishart prior in (2) restricts the support of Σ_G^{-1} to a space of positive definite matrices having zero off-diagonal elements corresponding to absent edges. We refer to the prior on the covariance in (2) as the semi-parametric hyper inverse-Wishart prior or spHIW, due to dependence on the unknown cluster allocations.

Formulation (2) specifies the edge inclusion probabilities as ω_1 or ω_0, depending on whether the edge corresponds to two nodes belonging to the same cluster or different clusters. By choosing hyper-parameters $a_{\omega,0}, b_{\omega,0}$, to have a small prior mean, and $a_{\omega,1}, b_{\omega,1}$, to have a larger prior mean, one can encourage a higher density of edges within clusters, and sparse edges across clusters. The proposed approach thus results in clusters of nodes such that there are sparse connections between groups, but denser connections within each group. The membership of each group is influenced by covariates, and every group comprises similarly activated nodes which are connected by a distinct sub-network. Our approach naturally allows for detection of functional modules, which is expected to provide deeper insights into the organization of the human brain.

We can obtain an explicit form for the prior in (2) as

$$\begin{aligned}\pi(G \mid S_1, \ldots, S_H) &= K^{-1}\left(\omega_1^{a_{\omega,1}+t_{1G}-1}(1-\omega_1)^{b_{\omega,1}+\sum_{h=1}^{H} p_h(p_h-1)/2-t_{1G}-1}\right) \\ &\times \left(\omega_0^{a_{\omega,0}+t_{0G}-1}(1-\omega_0)^{b_{\omega,0}+p(p-1)/2-\sum_{h=1}^{H} p_h(p_h-1)/2-t_{0G}-1}\right),\end{aligned}$$

where $G \in \mathcal{M}$, K is the normalizing constant, and t_{1G}, t_{0G}, represent the number of edges within and across clusters respectively. We noted previously that, when $\lambda_j = \lambda$ for all $j = 1, \ldots, \infty$, the prior on the regression coefficients in (1) is a Dirichlet process mixture of Laplace distributions. In such a case, we can use results from Kyung et al. (2009) to obtain the following form of the prior on the graph space after marginalizing out the clustering parameters

$$\pi(G) = \frac{\Gamma(M)}{\Gamma(M+p)} \sum_{H=1}^{p} M^H \sum_{(S_1,\ldots,S_H):|S(\beta)|=H} \prod_{h=1}^{H} \Gamma(p_h)\pi(G \mid S_1, \ldots, S_H) 1(G \in \mathcal{M}),$$

where the set $\{(S_1, \ldots, S_H) : |S(\beta)| = H\}$ contains all possible clustering memberships given that the number of clusters is fixed at H, and $\Gamma(\cdot)$ denotes the Gamma function.

We note that given a decomposable graph G having cliques $C_1, \ldots, C_k$, and separators $Q_2, \ldots, Q_k$, the likelihood can be written as

$$\begin{aligned} L(X|Z,B,G) &= \frac{\prod_{k=1}^{K} L(X_{\bullet,C_k}|Z, B_{\bullet,C_k})}{\prod_{k=2}^{K} L(X_{\bullet,Q_k}|Z, B_{\bullet,Q_k})}, \\ L(X_{\bullet,C_k}|Z, B_{\bullet,C_k}) &= \frac{\pi^{-n/2}\Gamma_{|C_k|}((b+n+|C_k|-1)/2)[det\{D_{C_k}\}]^{-n/2}}{\Gamma_{|C_k|}((b+|C_k|-1)/2)[det\{I_n + \tilde{X}_{\bullet,C_k} D_{C_k}^{-1} \tilde{X}_{\bullet,C_k}^T\}]^{(b+|C_k|-1)/2}}, \end{aligned} \tag{3}$$

using results in section 5.3 in Lauritzen (1996) and equation (45) in Dawid & Lauritzen (1993). Here, $|C_k|$ denotes the cardinality of C_k, D_{C_k} is a diagonal matrix with diagonals $\{d_s, s \in C_k\}$, and $\tilde{X}_{\bullet,C_k} = X_{\bullet,C_k} - ZB_{\bullet,C_k}$, where $X_{\bullet,C_k}, B_{\bullet,C_k}$ represent sub-matrices of X, B, with columns corresponding to C_k. In the situation where there are two or more clusters with no edges between them, the above likelihood can be factorized even further leading to speed-ups in computation. The form of the likelihood in (3) is used when updating the graph using MCMC, and this update can be performed efficiently even for large p.

2.2 Variable Selection

We propose a post-MCMC variable selection mechanism which proceeds by constructing joint credible regions accounting for multiplicity corrections. The variable selection enables us to infer (a) covariate influences on individual imaging phenotypes, and (b) subsets of covariates influencing clusters of phenotypes, where each cluster is characterized by a distinct sub-network and where a particular covariate may affect more than one cluster.

We construct rectangular credible regions incorporating multiplicity corrections as $\mathcal{D} := \{\beta_{jk} : \beta_{jk}/std(_{\beta_{jk}}) > U_{\alpha^*}, j = 1, \ldots, q, k = 1, \ldots, p\}$, where $std(_{\beta_{jk}})$ is the standard deviation for β_{kl}, and α^* is the multiplicity adjusted width of the credible intervals. The above credible intervals enable us to test a set of local hypotheses $H_{0,jk} : |\beta_{jk}| \leq U^*_{jk}$ versus $|\beta_{jk}| > U^*_{jk}$ for $j = 1, \ldots, q, k = 1, \ldots, p$, where the threshold for each regression coefficient is adjusted according to it's standard deviation, and hence is different from "hard" thresholding approaches which choose a fixed threshold. The local hypothesis tests can be done using a t-test at a significance level $\alpha^* = \alpha/(pq)$ under a Bon-ferroni correction. Although it is straightforward to use more sophisticated alternatives such as the false discovery rate approach (Benjamini & Hochberg 1995), we use the Bon-feronni correction since it performs adequately for the examples we considered.

2.3 Posterior Computation

We propose an efficient approximate posterior computation scheme using a parameter expansion strategy. Under the original formulation (1), the computation of cluster membership probabilities different columns of B will require p matrix inversions of order $p-1$ each, which can be computationally restrictive. We devise a parameter expanded model which altogether

bypasses the need of inverting matrices when computing cluster memberships. We fit the modified model

$$\mathbf{x}_i = \mathbf{z}_i B + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(0, \delta I_p), \quad \boldsymbol{\alpha}_i \sim N(0, \Sigma_G),\ i = 1, \ldots, n, \tag{4}$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{ip})$ can be interpreted as the intercept term, and $\delta \sim Be(a_\delta, b_\delta)$ is the residual variance. The prior on the graph and the covariance matrix is defined similarly as in (2). Marginalizing out the intercept in (4) yields $\mathbf{x}_i \sim N(\mathbf{z}_i B, \Sigma_G + \delta I) \approx N(\mathbf{z}_i B, \Sigma_G)$, when $\delta \approx 0$, which essentially gives back our original formulation (1).

The computational advantage of (4) stems from the fact that all elements in the data matrix X are independent conditionally on $B, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n, \delta$. This allows the following form of the likelihood times prior, conditional on the clustering $(S_1, \ldots, S_H)$

$$L \propto \left(\prod_{i=1}^{n} N(\mathbf{x}_i; \mathbf{z}_i(\boldsymbol{\eta}_{s_1}, \ldots, \boldsymbol{\eta}_{s_p}) + \boldsymbol{\alpha}_i, \delta I_p) N(\boldsymbol{\alpha}_i; 0, \Sigma_G) \right) \left(\prod_{h=1}^{H} \prod_{l=1}^{q} \mathrm{DE}(\eta_{hl}; \lambda_h) \right) \pi(\Sigma_G \mid G) \pi(G | S_1, \ldots, S_H), \tag{5}$$

where $s_j \in \{1, \ldots, H\}, j = 1, \ldots, p$, denote the cluster memberships. Under the above likelihood, it is straightforward to compute cluster membership s_j for the j-th column of B independently of the other columns, in a computationally inexpensive manner which does not involve matrix inversions. In practice, the approximation under (4) is implemented by specifying a conjugate prior on δ with mode near zero and having a small variance such that it results in posterior samples of $\delta = O(10^{-3})$. In our experience, this choice works adequately for a variety of scenarios.

We use a MCMC for the posterior computation, which proceeds by (a) updating the cluster memberships and cluster atoms conditional on the intercepts and the residual variance

(δ); (b) updating the graph conditional on the cluster memberships, and the inverse covariance matrix conditional on the graph; and (c) updating the intercepts and residual variance conditional on the other parameters. We update the graph using a Metropolis-Hastings step in a manner similar to Bhadra & Mallick (2012), where the proposal distribution changes a non-zero element in the adjacency matrix to a zero element with probability $1 - a_G$, and the reverse proposal occurs with probability a_G. Barring the graph, all remaining parameters in (5) can be sampled via closed form posteriors. The MCMC steps are described in detail in the Supplementary Materials.

Inferring optimal clustering and point estimate for the graph: Our computation yields posterior samples of cluster membership allocations for each column of B. In order to estimate the optimal clustering over MCMC iterations, we use the least squares criteria in Dahl (2006). Letting $\mathcal{S}^{(m)}$ denote the vector of cluster allocations at the m-th MCMC iteration, the optimal cluster is selected as $\mathcal{S}^* = \arg\min_{\mathcal{S}^{(m)}, m=1,\ldots,T} \sum_{i=1}^{p} \sum_{j=1}^{p} (\Delta_{i,j}(\mathcal{S}^{(m)}) - \hat{\pi}_{i,j})^2$, where $\Delta_{i,j}(\mathcal{S}^{(m)})$=1 if (i,j) belong to the same cluster under $\mathcal{S}^{(m)}$, and 0 otherwise, $m = 1, \ldots, T$, and $\hat{\pi}$ is the estimated matrix of pair-wise probabilities of belonging to the same cluster, computed over all MCMC iterations. The final estimated graph structure is computed in a manner consistent with this optimal clustering, by computing the marginal inclusion probabilities of edges using MCMC samples corresponding to the clustering $\mathcal{S}^*$, and including edges with high probabilities.

3. Simulation Studies

3.1 Description

We consider three simulation settings (Cases I–III) with varying dimensions involving a true model of the form $\mathbf{x}_i \sim N(\mathbf{z}_i B_0, \Sigma_0)$ where Σ_0 is the true covariance matrix, and $B_0 = (\boldsymbol{\beta}_{01}, \ldots, \boldsymbol{\beta}_{0p})$ are the true regression coefficients. For Cases I and II the number of non-zero rows in the coefficient matrix (B_0) are 10 and 5 respectively, where the elements in these non-zero rows are randomly set to 2,3, or 0, and the proportion of zeros are high to ensure a sparse coefficient matrix. For both these cases, the inverse covariance matrix $\Sigma_0^{-1} = \Omega_0$ is generated as follows. First, we generate Σ^* having elements $\sigma^*(l, l') = 0.5\Big(||l - l'|+1|^{1.4} - 2|l-l'|^{1.4} + ||l-l'|-1|^{1.4}\Big)$, $l, l' = 1, \ldots, p$, which corresponds to a fractional Gaussian noise process with Hurst parameter as 0.7. We then invert Σ^* to obtain Ω^*, and subsequently fix all off-diagonal elements of Ω^* to be zero if the absolute value is less than 0.05, to obtain Ω_1^*. Finally we rescale the diagonal elements of Ω_1^* as $\omega_{1,kk}^* = 0.1 + \sum_{j \neq k, j=1}^{p} |\omega_{1,jk}^*|$ to obtain a diagonally dominant matrix which is positive definite, denoted as $\Omega_0 = \Sigma_0^{-1}$. This is the true precision matrix that is used to generate the data. The true graph G_0 is obtained by including all edges corresponding to an absolute partial correlation greater than 0. Note that the true model is a violation of the clustering as well as the block diagonal assumptions inherent in the proposed methodology. We consider dimensions $(n, p, q) = (100, 80, 100), (100, 80, 200)$ for Cases I and II.

For Case III, we fit our model (1)-(2) to the PET data for individuals with mild cognitive impairment (MCI) obtained from the ADNI dataset, and then use the fitted model to simulate data. The dataset in question contains PET measurements recorded from $p = 42$ regions of interest (ROIs) in the brain for each of the 121 samples (n), with additional information on $q = 546$ SNPs. The 42 regions reported in Table 1 were selected as in Huang et. al (2010). These regions are distributed in the frontal, parietal, occipital, and temporal

lobes, and are considered to be potentially related to Alzheimer's disease. We fit our model using dichotomized SNP data with value 1 if the minor allele frequency is 1 or 2, and 0 otherwise. This fitted model which is used to generate data corresponds to a high dimensional multivariate response regression model, having 546 covariates, and 140 edges.

We compare our approach (spHIW) to (a) the sparse seemingly unrelated regression (SSUR) method in Bhadra & Mallick (2012) for simultaneous graphical model estimation and variable selection; (b) a multivariate version of the Bayesian lasso (Park & Casella, 2008) denoted as BLASSO designed to perform variable selection; and (c) the frequentist graphical lasso (Friedman et. al, 2007) denoted as GLASSO for graphical model estimation without accounting for covariates; We wrote the MATLAB codes for implementing spHIW, and BLASSO, while the code for SSUR was obtained from the authors of that article. The GLASSO was implemented using the R package glasso.

For the Bayesian approaches, we ran 25,000 MCMC iterations with a burn in of 5,000. The initial adjacency matrix for the proposed approach and SSUR was chosen to be identity corresponding to a null graph, and the parameters in the hyper inverse-Wishart prior for these approaches was defined as $b = 3, D = dI_p$. We imposed a conjugate Gamma prior on d which seemed to work well in a variety of scenarios. In addition, we specify independent Gamma priors on $\lambda_l, l = 1, \ldots, q,$, as well as $M \sim Ga(1, 1)$, and $\delta^{-1} \sim Ga(1000, 1)$. All results are reported over 50 replicates.

3.2 Comparison Criteria

We looked at several metrics for comparison, including (i) out of sample prediction in terms of mean squared error or MSE; (ii) estimation of true regression coefficients in terms of L_2 error;

(iii) estimation of the precision matrix in terms of L_1 error; (iv) area under the ROC curve for variable selection; and (v) area under the ROC curve for graphical model estimation. The predicted test samples were obtained using posterior predictive distributions under Bayesian approaches, and this was used to compute MSE. However, it was not possible to report MSE under GLASSO since it does not incorporate covariate information. Estimation of the precision matrix and regression coefficients under our approach was based on MCMC samples corresponding to the optimal clustering as outlined in Section 2.3, while it was based on all MCMC samples for the other Bayesian approaches.

To compute the area under the curve (AUC) for variable selection under our approach and BLASSO, we looked at a series of regression models obtained by including all covariates for which $\hat{\beta}_{kl}/std(\hat{\beta}_{kl}) > t_{kl}$, and excluding remaining variables. Here t_{kl} is a threshold which controls the sparsity of the regression model, and $\hat{\beta}_{kl}, std(\hat{\beta}_{kl})$ are the estimated mean and standard errors for $\beta_{kl}, k = 1, \ldots, q, l = 1, \ldots, p$. For SSUR, the area under the curve was computed by looking at a series of regression models obtained by varying the threshold for the posterior inclusion probabilities. On the other hand, we computed the AUC for graphical model estimation under the spHIW, and SSUR by looking at a series of graphs obtained by varying the threshold for posterior inclusion probabilities for edges. Again, only the MCMC samples corresponding to the optimal clustering was used to compute the graph under our approach. The AUC for graph estimation under GLASSO was obtained by looking at a series of models corresponding to different values of the penalty parameter.

3.3 Results

The numerical results under all approaches are reported in Table 2. We observe that for

Cases I and II, the error for estimating the true regression coefficients under our approach was lower compared to BLASSO under Case I, but slightly higher under Case II; while it was lower compared to SSUR for all scenarios. The area under the ROC curve for variable selection is the highest under BLASSO, while it is close to 0.5 for SSUR. The poor area under SSUR is due to the inclusion of almost all covariates in the model, which indicates the inability of the approach to differentiate between important and unimportant covariates. The lower area under the curve for variable selection under the proposed approach compared to BLASSO is likely due to the presence of a sizable number of additional covariance parameters in the model. Although the proposed method has a lower area under the curve, it has does significantly better in terms of out of sample prediction compared to all approaches. This points to the advantage of incorporating the graph structure for predicting test sample observations, compared to BLASSO which assumes independence within the outcome measurements. SSUR has the largest out of sample MSE which is likely due to the inclusion of almost all covariates in the regression model.

For graphical model estimation, we note the the proposed approach, SSUR, and GLASSO all have a similar area under the curve. However the error for estimating the precision matrix is the lowest under the proposed approach, which points to a superior ability to accurately estimate partial correlations. We conjecture that a lower error in estimating the partial correlations is due to the removal of external sources of variation, which when unaccounted for, can potentially lead to erroneous estimates for strength of associations.

In summary, we conclude that under Cases I and II, the proposed conditional graphical model has superior out of sample prediction performance by incorporating the underlying graph structure, but suffers from a poorer variable selection performance due to the presence

of a sizable number of additional covariance parameters. In contrast, the competing SSUR approach does very poorly in terms of variable selection and out of sample prediction. Both SSUR, which includes almost all covariates in the regression model, and GLASSO, which does not include any covariate at all, have comparable graph estimation performance, but higher errors when estimating the strength of associations. This underlines the role of accurate variable selection as an important factor in the estimation of conditional associations.

Under Case III where the data resembles a real world application, the proposed approach has superior performance compared to all approaches. In particular, the method has comparable out of sample prediction, but a lower error for estimating true regression coefficients, and a significantly higher area under the curve for variable selection. The higher area under the curve compared to Cases I and II, points to the increased ability of the proposed method to differentiate between important and unimportant variables when the dimension of the multivariate outcome is moderate compared to the sample size, even when the number of candidate predictors is large. We also observe that the proposed approach has an improved graphical model estimation performance relative to competing approaches, as evident from significantly higher area under the curve, and a significantly lower error for estimating partial correlations.

4. Application to Imaging Genetics

4.1 Description of ADNI Data

The Alzheimer's Disease Neuroimaging Initiative (ADNI) collected a large amount of imaging, genetic and clinical data. The goal of the ADNI study is to determine whether different imaging biomarkers, along with genetic variants and clinical markers are strongly associ-

ated with the Alzheimer's disease (AD) and the progression of mild cognitive impairment (MCI). In this article, we primarily concentrate on identifying (i) important connections in the functional brain network after accounting for age, gender, handedness, weight, and genes; (ii) functional modules or collections of ROIs in the brain which work together to drive brain functions, and the corresponding sub-networks; and (iii) important genes influencing the imaging phenotype and the functional modules. The brain network is computed using PET measurements, however it is straightforward to apply the method to other imaging modalities such as MRI. We perform the analysis separately for the MCI, AD groups, and healthy controls (HC), and compare results across the three groups. We begin with a data description.

Imaging data: ADNI 1 collected the longitudinal PET scans at multiple time points across different imaging sites. To study the association between the imaging biomarkers and genetic variants, we used the PET scans at baseline for 49 AD patients, 121 MCI patients and 71 healthy subjects. The standard pre-processing steps including co-registration, normalization and spatial smoothing (8 mm FWHM) were applied to the PET dataset. We considered 90 brain regions that are defined according to the automated anatomical labeling (AAL) system. We computed the PET regional summaries using the first principal component scores over all voxels with each region, in a similar fashion as in Bowman et. al (2012). This 90×1 dimensional summary vector of PET scans is our outcome variable.

Genetics data: The SNPs in the ADNI study were genotyped using the Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA, USA). By following Zhu et al (2014) and Wang et al (2012), we only focused on SNPs that belong to the top 40 candidate genes reported in the AlzGene database (www.alzgene.org) as of June 10, 2010. Before the data analysis, we

performed standard pre-processing steps (see Wang et. al 2012) on the SNP data for quality control. We also removed the SNPs having (a) more than 1% missing values; (b) minor allele frequency less than 5% and (c) the Hardy-Weinberg Equilibrium p-value less than 10^{-6}. The final dataset includes 614 SNPs on 37 genes. Figure 1 shows the number of the SNPs in the analysis per gene. The total number of covariates is 618 including pre-selected 614 SNPs and four demographic variables including handedness, age, gender and weight.

4.2 Analysis Results

Brain Network Identification: Based on the MCMC samples, we computed the posterior edge inclusion probabilities of the brain network for each group. By thresholding the probability at 0.5, the group-specific brain networks are obtained in Figure 2. Specifically, the AD, MCI and HC networks have 79, 102 and 73 important edges, respectively. There are 14 edges shared by all the three groups. Some of the common edges are in the default mode network (Buckner et al, 2008). For example, the functional connections between left and right Precuneus (related to self-consciousness) appear in all the three networks, which implies that AD or MCI subjects have similar functional activities between the two Precuneus regions as the HC subjects. Also, all the three networks contain edges between the left and right Hippocampus, which indicates that the two regions are still functionally connected in the AD and MCI groups, although the damage in the Hippocampus has been confirmed to be related to AD.

As one of the defining features of the proposed method, it can identify functional modules or communities for each group specific network. In our analysis, there are three, three and two functional modules identified in the AD, MCI and HC groups, respectively, which are

shown in Figure 3. It can be seen that AD and MCI have two similar functional modules: communities 1 and 2, while the modules related to HC are quite different from the AD and MCI groups. Functional module 1 for AD and MCI groups collect many regions in parietal and temporal lobes, with the number of connections being 29 and 42 respectively. We observe that some functional connections between the two hemisphere are missing in the AD group compared to the MCI group. For example, the AD network does not have the functional connections between the right and left fusiform gyrus, the functionality of which is mainly related to face and body recognition (McCarthy et.al. 1997).

The functional module 2 in both AD and MCI networks include Hippocampus (HIP) and ParaHippocampal (PHG) in temporal lobe. The total number of connections between these two regions and all other regions are six in the MCI network and only two in the AD network. This implies that these regions become more isolated in AD group compared to the MCI group, which has been confirmed by previous findings (Supekar et al., 2008; Huang et al. 2010). The functional module 3 for the AD group mainly includes four regions: Postcentral gyrus (PoCG), Precentral gyrus (PreCG), Paracentral lobule (PCL) and Supplementary motor area (SMA), whose functions are mainly related to the motor skill and sense of touch. Since this is a separate module in the AD group, these regions have much fewer connections compared to the HC group which potentially implies reduced motor skills and sense of touch for AD subjects. Compared to the brain networks for AD and MCI groups, our analysis only detects two functional modules for the HC, which implies increased connectivity compared to the AD and MCI patients.

Important Genes for Brain Networks: Based on the MCMC samples, we identify important SNPs associated with each functional module for AD, MCI, and HC groups (Table 3). For

example, SNP "rs2018334" on NEDD9 is significantly associated with the sub-network community 2 in the AD group, which is supported by the findings in Wang et.al. (2012). GAB2 was also recognized as an important gene for both AD and MCI groups, but not the HC group; this corresponds to prior evidence implying that the gene modifies late onset AD risk in APOE $\epsilon 4$ carriers and influences Alzheimer's neuropathology (Reiman et al., 2007). Keeping in line with prior findings, CH25H was found to be significantly associated with AD risk but not MCI or HC (Wollmer, 2010). Further, genes which promote MCI disease risk but are not associated with HC individuals include ECE1 which is associated with cognitive ability in elderly individuals and disease risk (Hamilton et. al, 2012), as well as ADAM-10 which regulate neuronal plasticity affecting AD (Marcello et. al, 2013), and PICALM, which was one of the first AD loci identified by GWAS, and which has also been validated in independent samples. In addition, SORL1 which is known to be a potential tool for identifying MCI subjects at high risk of conversion to AD (Piscopo et. al, 2015), is found to be significant in the MCI and HC groups, but not with the AD group. In addition to the above genes, we found that age is a significant predictor for community 1 in the MCI group.

5. Discussion

We have developed a new Bayesian semi-parametric conditional graphical model for imaging genetics studies, and applied it for analyzing the ADNI dataset. Our approach can jointly estimate the brain network after accounting for external sources of variation, and infer important genetic and demographic factors associated with the imaging phenotype and the brain network. It can also simultaneously discover functional modules in the brain and infer the connectivity within each such module. To our knowledge, the proposed method is

among the first to jointly address the above aims and is expected to provide deeper insights in imaging genetic studies, compared to existing approaches.

References

[1] Benjamini, Y & Hochberg, Y (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society, Series B*, **57**, 289-300.

[2] Bhadra, A & Mallick, BK (2013), Joint High-Dimensional Bayesian Variable and Covariance Selection with an Application to eQTL Analysis, *Biometrics*, **69**, 447-457.

[3] Buckner, RL, Andrews–Hanna, JR & Schacter, DL (2008), The brain's default network, *Annals of the New York Academy of Sciences*, **1124**, 1-38.

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

[5] Cai, T, Li, H, Liu, W & Xie, J (2013), Covariate-adjusted precision matrix estimation with an application in genetical genomics, *Biometrika*, **100**, 139–156.

[6] Dawid, AP & Lauritzen, SL (1993), Hyper markov laws in the statistical analysis of decomposable graphical models, *Annals of Statistics*, **21**, 1272–1317.

[7] Friedman, J, Hastie, T, & Tibshirani, R (2008), Sparse inverse covariance estimation with the lasso, *Biostatistics*, **9**, 432-441.

[8] Green, PJ & Thomas, A (2013), Sampling decomposable graphs using a Markov chain on junction trees, *Biometrika*, **100**, 91-110.

[9] Guo, Y (2011), A general probabilistic model for group independent component analysis and its estimation methods, *Biometrics*, **67**, 1532-1542.

[10] Guo, Y & Tang, Li (2013), A hierarchical probabilistic model for group independent component analysis in fMRI studies, *Biometrics*, **69**, 970-81.

[11] Hamilton, G, Harris, SE, Davies, G, Liewald, DC, Tenesa, A & Payton, A (2012), The Role of ECE1 Variants in Cognitive Ability in Old Age and Alzheimer's Disease Risk, *American Journal of Medial Genetics, Part B*, **159 B**, 676 – 709.

[12] Huang S, Li, J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E & Alzheimer's Disease NeuroImaging Initiative, (2010), Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation, *Neuroimage*, **50**, 935-949.

[13] Kim, J, Wozniak, JR, Mueller, BA, & Pan, W, (2015), Testing Group Differences in Brain Functional Connectivity: Using Correlations or Partial Correlations?, *Brain Connectivity*, **5**, 214-231.

[14] Lauritzen, SL (1996), Graphical Models, *Oxford University Press*.

[15] Li, B, Chun, H & Zhao, H, (2012), Sparse Estimation of Conditional Graphical Models With Application to Gene Networks, *Journal of the American Statistical Association*, **497**, 152-167.

[16] McCarthy, G, Puce, A, Gore, JC & Allison, T, (1997). Face-specific processing in the human fusiform gyrus, *Journal of cognitive neuroscience*, **9**, 605-610.

[17] Mier, D, Kirsch, P & Meyer-Lindenberg, A, (2010), Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis, *Molecular Psychiatry*, **15**, 918927.

[18] Munafo, MR, Attwood, AS & Flint, J, (2008), Bias in genetic association studies: effects of research location and resources, *Psychological Medicine*, **38**, 12131214.

[19] Park, T, & Casella, G, (2008), The Bayesian Lasso, *Journal of the American Statistical Association*, **103**, 681–686.

[20] Piscopo, P, Tosto, G, Belli, C, Talarico, G, Galimberti, D, Gasparini, M, et al., (2015). SORL1 Gene is Associated with the Conversion from Mild Cognitive Impairment to Alzheimer's Disease, *Journal of Alzheimer's Disease*, **46**, 771–776.

[21] Reiman, EM, Webster, JA, Myers, AJ, Hardy, J, Dunckley, T, Zismann, VL, et al., (2007). GAB2 Alleles Modify Alzheimer's Risk in APOE $\epsilon 4$ Carriers, *Neuron*, **54**, 713-720.

[22] Sethuraman, J, (1994), A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639-650.

[23] Smith, SM, Beckmann, CF, Andersson, J, Auerbach, EJ, Bijsterbosch, J, Douaud, G, Duff, E, Feinberg, DA, Griffanti, L, Harms, MP, Kelly, M, Laumann, T, Miller, KL, Moeller, S, Petersen, S, Power, J, Salimi-Khorshidi, G, Snyder, AZ, Vu, AT, Woolrich, MW, Xu, J, Yacoub, E, Uurbil, K, Van Essen, DC & Glasser, MF, (2012), Resting-state fMRI in the Human Connectome Project, *Neuroimage*, **80**, 144168.

[24] Stam, CJ, Jones, BF, Nolte, G, Breakspear, M & Scheltens, P, (2007), Small-world networks and functional connectivity in Alzheimer's disease, *Cerebral Cortex*, **17**, 92-99.

[25] Stein, JL, Hua, X, Lee, S, Ho, AJ, Leow, AD, Toga, AW, et al., (2010), Voxelwise genome-wide association study (vGWAS). *Neuroimage*, **53**, 1160-1174.

[26] Stingo, FC, Guindani, M, Vannucci, M, and Calhoun, VD, (2013), An Integrative Bayesian Modeling Approach to Imaging Genetics, *Journal of the American Statistical Association*, **108**, 875–891.

[27] Supekar, K, Menon, V, Rubin, D, Musen, M & Greicius, MD, (2008), Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer's Disease, *PLoS Computional Biology*, **4**, 1-11.

[28] Wang, Y, Bi, L, Wang, H, Li, Y, Di, Q, Xu, W, & Qian, Y (2012), NEDD9 rs760678 polymorphism and the risk of Alzheimer's disease: A meta-analysis, *Neuroscience letters*, **527**, 121-125.

[29] Wollmer, MA, (2010), Cholesterol-related genes in Alzheimer's disease, *Biochim Biophysica Acta*, **1808**, 762-773.

[30] Yin, J & Li, H (2011), A Sparse Conditional Graphical Model for Analysis of Genetical Genomics Data, *Annals of Applied Statistics*, **5**, 2630–2650.

[31] Zhu, H, Khondker, Z, Lu, Z, Ibrahim, JG & Alzheimer's Disease Neuroimaging Initiative, (2014), Bayesian Generalized Low Rank Regression Models for Neuroimaging Phenotypes and Genetic Markers, *Journal of the American Statistical Association*, **109**, 977–990.

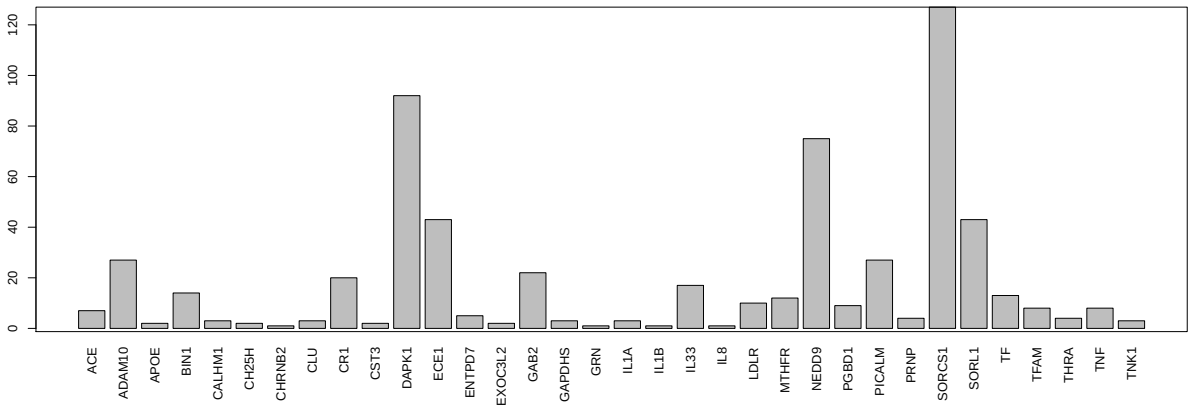

Figure 1: Top 37 genes in the analysis and the number of SNPs per gene. There are a total of 614 SNPs included in the analysis.

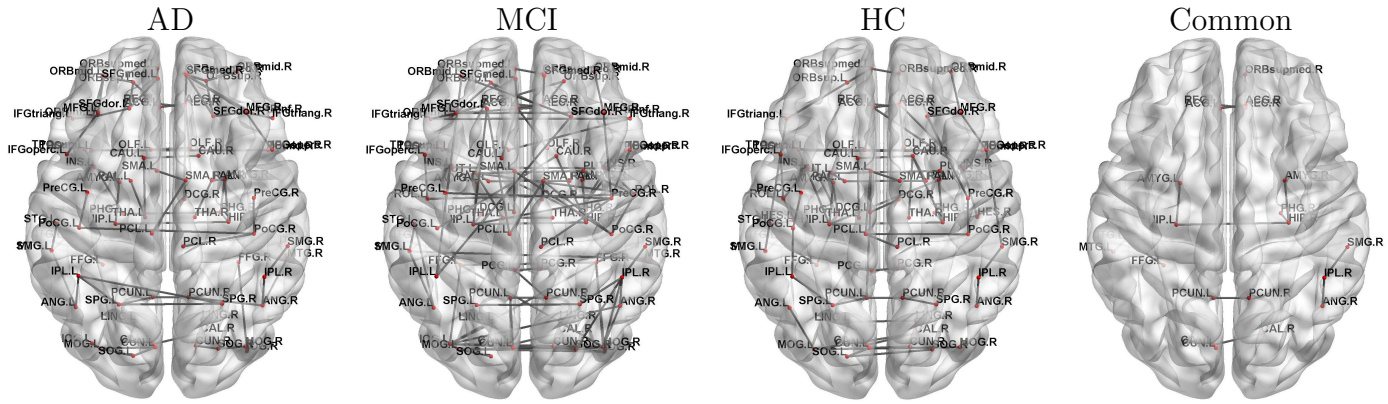

Figure 2: Functional brain network estimation for AD, MCI and HC groups and the common edges shared by the three networks

Table 1: List of 42 regions of interest for simulation Case III.

Frontal lobe	Parietal lobe	Occipital lobe	Temporal lobe
Frontal_Sup_L	Parietal_Sup_L	Occipital_Sup_L	Temporal_Sup_L
Frontal_Sup_R	Parietal_Sup_R	Occipital_Sup_R	Temporal_Sup_R
Frontal_Mid_L	Parietal_Inf_L	Occipital_Mid_L	Temporal_Pole_Sup_L
Frontal_Mid_R	Parietal_Inf_R	Occipital_Mid_R	Temporal_Pole_Sup_R
Frontal_Sup_Medial_L	Precuneus_L	O ccipital_Inf_L	Temporal_Mid_L
Frontal_Sup_Medial_R	Precuneus_R	Occipital_Inf_R	Temporal_Mid_R
Frontal_Mid_Orb_L	Cingulum_Post_L		Temporal_Pole_Mid_L
Frontal_Mid_Orb_R	Cingulum_Post_R		Temporal_Pole_Mid_R
Rectus_L			Temporal_Inf_L 8301
Rectus_R			Temporal_Inf_R 8302
Cingulum_Ant_L			Fusiform_L
Cingulum_Ant_R			Fusiform_R
			Hippocampus_L
			Hippocampus_R
			ParaHippocampal_L
			ParaHippocampal_R

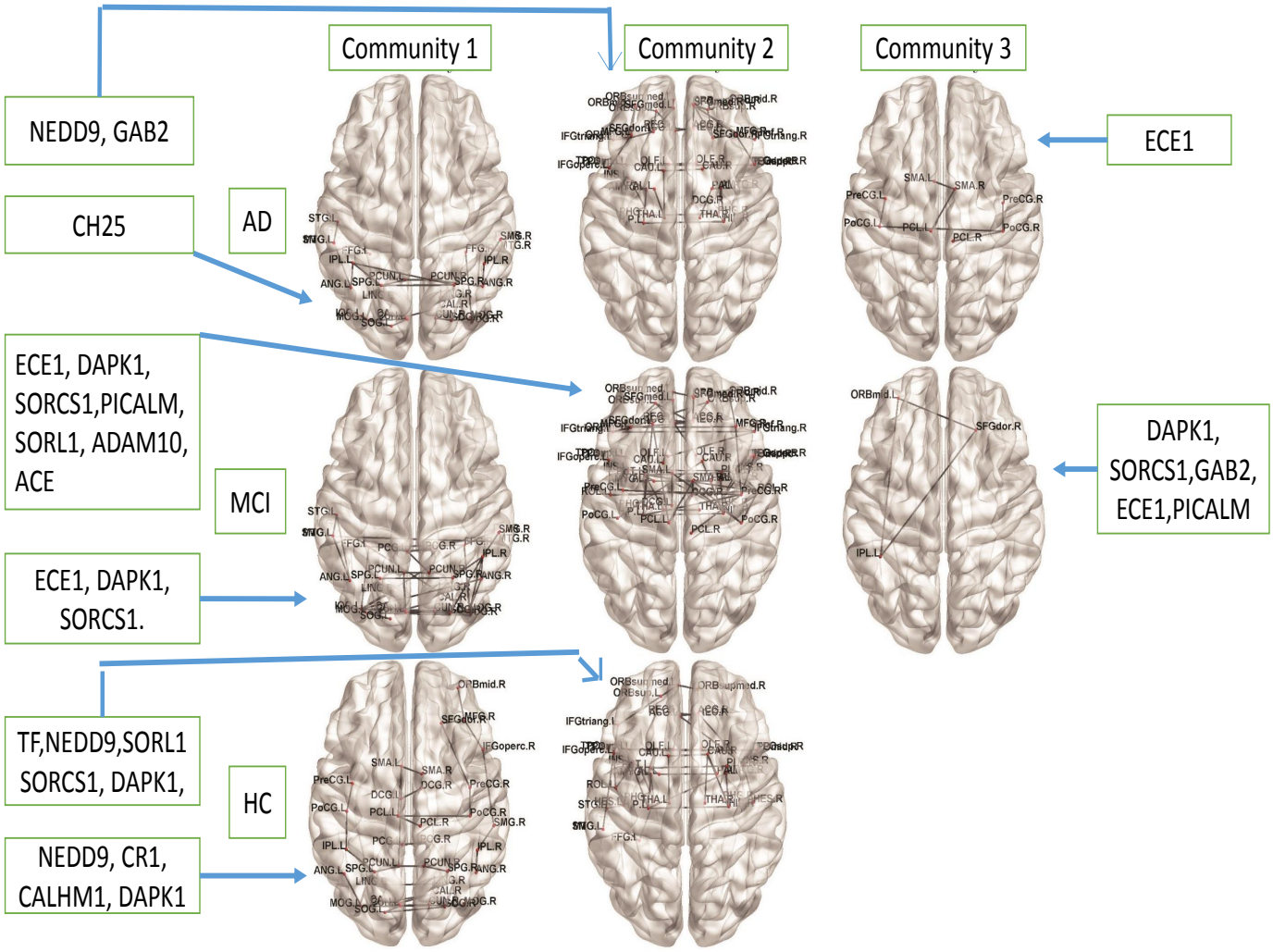

Figure 3: Functional modules or communities for AD, MCI and HC groups, along with important genes. The arrows relate the sub-networks to the significant genes influencing them. Each such gene can influence one or more connections in the functional modules, as well as one or more phenotypes in that module.

Table 2: Numerical comparison for different approaches under all Cases. MSE stands for out of sample mean squared error; $||\hat{\boldsymbol{\beta}}||^2_{L_2}$ implies squared L_2 error in estimating the regression coefficients; AUC(graph) and AUC(var) denote area under the curve for graphical model estimation and variable selection respectively; and $||\hat{\Omega}||_{L_1}$ denotes error for estimating the precision matrix.

Case (n,p,q)	Method	MSE	$\|\|\hat{\boldsymbol{\beta}}\|\|^2_{L_2}$	AUC(graph)	AUC(var)	$\|\|\hat{\Omega}\|\|_{L_1}$
I	spHIW	0.27	0.05	0.80	0.82	0.0138
(100,80,100)	BLASSO	0.55	0.07	NA	0.98	NA
	SSUR	0.62	0.12	0.78	0.50	0.0389
	GLASSO	NA	NA	0.79	NA	0.0447
I	spHIW	0.28	0.05	0.79	0.80	0.0141
(100,80,200)	BLASSO	0.56	0.09	NA	0.96	NA
	SSUR	0.60	0.10	0.76	0.52	0.0917
	GLASSO	NA	NA	0.78	NA	0.0401
II	spHIW	0.26	0.10	0.80	0.85	0.0112
(100,80,100)	BLASSO	0.54	0.08	NA	0.96	NA
	SSUR	0.61	0.13	0.77	0.50	0.0271
	GLASSO	NA	NA	0.79	NA	0.0432
II	spHIW	0.28	0.10	0.79	0.92	0.0127
(100,80,200)	BLASSO	0.57	0.08	NA	0.97	NA
	SSUR	0.65	0.11	0.78	0.52	0.12
	GLASSO	NA	NA	0.80	NA	0.0391
III	spHIW	0.01	0.0015	0.88	0.94	0.17
(121,42,546)	BLASSO	0.02	0.0028	NA	0.51	NA
	SSUR	0.02	0.0031	0.87	0.78	0.97
	GLASSO	NA	NA	0.89	NA	0.56

Table 3: Important SNPs (genes) that are significantly associated with the sub-network communities for each group

	AD	MCI	HC
Community 1	rs4933497 (CH25H)	rs11590928 (ECE1), rs3026886 (ECE1), rs1015477 (DAPK1), rs10868609 (DAPK1), rs1105384 (DAPK1), rs10509825 (SORCS1), rs10501608 (PICALM), rs1790213 (SORL1), rs12594742 (ADAM10), rs4309 (ACE)	rs4428180 (TF), rs7748486 (NEDD9), rs661319 (SORCS1), rs4713432 (NEDD9), rs10868644 (DAPK1), rs11601559 (SORL1)
Community 2	rs2018334 (NEDD9), rs11603112 (GAB2)	rs3026868 (ECE1), rs3026886 (ECE1), rs871495 (DAPK1), rs12248564 (SORCS1), rs821962 (SORCS1), rs1015477 (DAPK1)	rs16871157 (NEDD9), rs6691117 (CR1), rs729211 (CALHM1), rs7036781 (DAPK1)
Community 3	rs212518 (ECE1)	rs1015477 (DAPK1), rs1415020 (SORCS1), rs821962 (SORCS1), rs2450135 (GAB2), rs7941639 (GAB2), rs3026886 (ECE1) rs10509825 (SORCS1), rs34634755 (GAB2), rs666682 (PICALM)	