New Statistical Methods for Evaluating Brain Functional Connectivity

by

Emily L. Morris

A dissertation submitted in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2021

Doctoral Committee:

Professor Jian Kang, Chair
Associate Professor Kevin He
Professor Timothy Johnson
Professor Stephan Taylor

Emily L. Morris

emorrisl@umich.edu

ORCID iD: 0000-0001-7357-0483

ACKNOWLEDGEMENTS

I would like to sincerely thank all those who supported me throughout my graduate studies. This period of learning and growth would not have been possible without the support of my family and friends, and faculty, staff, and peers in the Department of Biostatistics.

First and foremost, I am extremely appreciative of my advisor, Jian Kang. Jian is a patient and encouraging mentor as well as a brilliant statistician. His support and guidance over the years was invaluable to my progress and education. I am inspired to work harder and persevere towards my goals because of his wonderful advising. This dissertation would not be possible without his continued encouragement and mentoring. I would also like to thank my dissertation committee for their support, input, and collaborations: Kevin He, Tim Johnson, and Steve Taylor.

I would like to acknowledge Matt Schipper, who supported my GSRA work for the past few years. He provided a great learning environment for working with collaborators in medicine. I want to thank Yi Li and his lab who supported me in my early years in the program and have continued to generously provide input on my work. Yi has always been supportive and encouraging, and I especially appreciate his support when I was first learning about collaborative research at KECC.

I would also like to acknowledge the faculty and students who participated in STATCOM with me, in particular Cathie Spino, Tom Braun, Steve Salerno, and the leadership team. I found this experience to be particularly enriching and rewarding in large part because of the work and dedication of this group.

I'd like to thank my professors at Miami University who were instrumental in my decision to go to graduate school, particularly John Bailer and Tom Fisher. Their commitment to teaching laid the foundation for my future studies, and I truly appreciate their continued support and mentorship.

Finally I'd like to thank my family and friends, who were the support I needed to make it through the ups and downs of pursuing a doctoral degree. Special thanks to the friends who studied by my side and pushed me to persevere through the years.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

ABCD Adolescent Brain Cognitive Development

ADMM Alternating direction method of multipliers

AUC Area under the curve

BOLD blood-oxygen-level-dependent

DNN Deep neural networks

PCA principal component analysis

PNC Philadelphia Neurodevelopmental Cohort

rfMRI Resting-state functional magnetic resonance imaging

SVM Support Vector Machine

CBCL Child Behavior Checklist

FM functional modules

ABSTRACT

The human brain functions through the coordination of a complex network of billions of neurons. This network, when defined by the functions it dictates, is known as functional brain connectivity. Associating brain networks with clinical symptoms and outcomes has great potential for shaping future work in neuroimaging and clinical practice. Resting-state functional magnetic resonance imaging (rfMRI) has commonly been used to establish the functional brain network; however, understanding the links to clinical characteristics is still an ongoing research question. Existing methods for analysis of functional brain networks, such as independent component analysis and canonical correlation analysis, have laid a good foundation for this research; yet most methods do not directly model the node-level association between connectivity and clinical characteristics, and thus provide limited ability for interpretation. To address those limitations, this dissertation research focuses on developing efficient methods that identify node-level associations to answer important research questions in brain imaging studies.

In the first project, we propose a joint modeling framework for estimating functional connectivity networks from rfMRI time series data and evaluating the predictability of individual's brain connectivity patterns using their clinical characteristics. Our goal is to understand the link between clinical presentations of psychiatric disorders and functional brain connectivity at different region pairs. Our modeling framework consists of two components: estimation of individual functional connectivity networks and identifying associations with clinical characteristics. We propose a model fitting procedure for jointly estimating these components via the alternating

direction method of multipliers (ADMM) algorithm. The key advantage of the proposed approach lies in its ability to directly identify the brain region pairs between which the functional connectivity is strongly associated with the clinical characteristics. Compared to existing methods, our framework has the flexibility to integrate machine learning methods to estimate the nonlinear predictive effects of clinical characteristics. Additionally, jointly modeling the precision matrix and the predictive model estimates provides a novel framework to accommodate the uncertainty in estimating functional connectivity.

In the second project, we focus on a scalar-on-network regression problem which utilizes brain functional connectivity networks to predict a single clinical outcome of interest, where the regression coefficient is edge-dependent. To improve estimation efficiency, we develop a two stage boosting algorithm to estimate the sparse edge-dependent regression coefficients by leveraging the knowledge of brain functional organization. Simulations have shown the proposed method has higher power to detect the true signals while controlling the false discovery rate better than existing approaches. We apply the proposed method to analysis of rfMRI data in the Adolescent Brain Cognitive Development (ABCD) study and identify the important functional connectivity sub-networks that are associated with general cognitive ability.

In the third project, we extend scalar-on-network regression via boosting in the second project by relaxing the homogeneity constraints within the prespecified functional connectivity networks. We adopt deep neural networks (DNN) to model the edge-dependent regression coefficients in light of the edge-level and node level features in the brain network, as well as the well-known brain functional organization. In addition, the proposed DNN-based scalar-on-network regression has the flexibility to incorporate the signal pattern from other imaging modalities into the model. We develop an efficient model fitting method based on ADMM. The proposed method is evaluated and compared with existing alternatives via simulations and analysis of

rfMRI and task fMRI data in the ABCD study.

CHAPTER I

Introduction

The human brain coordinates behavior through the integrated activity of billions of neurons acting in complex functional networks. Increasing interest in this field and imaging advances have propelled research with the goal to understand these functional networks and how they correspond to clinical presentations of disease. Resting-state functional magnetic resonance imaging (rfMRI) is noninvasive imaging that captures brain activity through variations in the blood-oxygen-level-dependent (BOLD) signal. The functional connectivity network is defined by the correlation of the BOLD signals across different regions of the brain.

Functional connectivity patterns across patients can help to identify those who are at high risk of developing a psychiatric disorder or serve as a disease marker. For most disorders this connection is still an open question, and therefore additional tools are needed to analyze functional connectivity. Long-term, large scale, imaging studies, such as the Adolescent Brain Cognitive Development (ABCD) study, are currently being performed increasing the availability and richness of brain imaging data.

Careful considerations must be made when working with any brain imaging study. For example the PNC cohort includes a smaller sample which may reduce the ability to differentiate signal from noise in the brain images; however all the images were collected at one location on one machine, therefore reducing outside sources of noise that

occur when these elements vary subject to subject. This characteristic of the study may also effect generalizability of results, which must also be carefully considered according to the ways in which the study and imaging was administered.

Another important consideration to make is the way an image is summarized for each individual to best capture the underlying activity and relate to other subjects. The brain may be viewed in different ways, for example, the image may be summarized by an even distribution of voxels throughout the brain without regard of the physical surfaces in the brain (commonly referred to as volumetric images). In contrast, the images could be summarized across location on the cortical surface of the brain, which does take into account distance of locations across the folded cortex. There is increasing evidence that accounting for cortical surface may improve results of functional connectivity analysis, though both image types are still commonly used [8, 2].

Some existing methods may be considered to address these questions. One initial challenge of functional connectivity analysis is defining the connectivity network for each individual. Broadly, this is done by quantifying the correlation between any two nodes' time series in the network. Using simple correlation may not adequately answer connectivity questions, and partial correlation captures the relationship between any two nodes independent of all other nodes' associations [81]. Other more complex methods may also be considered to do this, one approach uses Bayesian networks to estimate directed relationships between nodes. Bayesian networks uses directed acyclic graphs, which labels the direction of the connection between nodes [55]. Another common approach is using graphical models. *Smith et al.* [72] provides a more detailed overview of functional connectivity methods.

Graphical models, such as graphical lasso, have frequently been used to estimate the functional connectivity network[23, 95, 54]. There are many extensions to graphical lasso that could be applied to functional connectivity, such as smoothly clipped

absolute deviation (SCAD) which uses a different penalty and Bayesian graphical lasso which uses Bayesian methods [20, 80]. An alternative to graphical lasso is the semiparametric method for high-dimensional undirected graphs detailed in *Liu et al.* [49]. This method implements a semiparametric Gaussian copula to model the undirected graph, relaxing the normality assumption commonly needed in other graphical model estimation methods.

Current methods that connect functional connectivity to clinical characteristics include independent component analysis, canonical correlation analysis, and scalar-on-network regression. Some of these methods hinge upon dimension reductions methods to overcome the computational challenges of working with brain imaging data. Canonical correlation analysis is one example of this, in which the method links correlation structures of two sets of variables. Applying this to rfMRI data, it can link correlation structure among functional networks to correlation among groups of symptoms [83]. A focus on dimension reduction helps when the problem is computationally intractable but limits the extent with which brain regions can be identified.

Another broad approach is using scalar on network or image on scalar regression. One method proposed to do this is a classification method described in *Relión et al.* [61], this method performs variable selection on the nodes and edges in a network in order to classify the entire network. This is a similar idea to one of the methods proposed here, yet the implementation is different. Group selection methods, like group lasso, can be used to select functional networks that are associated with a symptom or disorder, yet they have limited power to detect such associations.

Machine learning methods may also be considered to address such questions. Support Vector Machine (SVM), random forest, and boosting are a few methods that we will consider for prediction. SVM has been shown to have good performance for classification in a wide variety of contexts, with some applications in neuroimaging [17, 57]. Random forest similarly has been used for classification and prediction in a

range of applications [7]. The Alternating direction method of multipliers (ADMM) algorithm was developed as a tool to solve distributed convex optimization problems, and specifically beneficial in high-dimensional settings where there is a very large number of features [5].

Boosting was developed to improve prediction by combining weak learners, with its beginning in the AdaBoost algorithm [53]. This method evolved into various other boosting methods including tree-based boosting, gradient boosting, and likelihood-based boosting. L_2 Boosting is one such form of boosting that uses the gradient descent algorithm with L_2 loss [9]; this form of boosting has been shown to perform well in high-dimensional settings such as in brain imaging or genomic applications[26].

Deep neural networks (DNN) have been shown to achieve very high predictive performance in applications as varied as speech recognition [56] to predicting crop yield [42]. Imaging, in particular, has benefited greatly from deep learning methods that efficiently utilize vast amounts of data for classification and regression problems [39, 89, 62]. In medical imaging, DNN has been used to aid in diagnostics or disease detection for several diseases like breast cancer [15], prostate cancer [48], and diabetic retinopathy [32]. They have also been used to assist with treatment in many contexts [47, 33]. DNN have been used for feature extraction with rfMRI images in previous work [38] or identifying functional networks using rfMRI images [92]. The applications are wide, but it is clear that DNN are a useful tool for medical imaging applications, including rfMRI.

Methods incorporating multiple imaging modalities are increasingly needed as imaging studies become more prevalent. To the best of our knowledge, there are very few existing methods to incorporate resting-state and task fMRI data, like we propose to do in this dissertation. Other approaches combine different imaging modalities, such as multiple tasks [11, 60] or multimodal PET scans [85]. These methods also have different approaches, like joint estimation procedures or simultaneous inference

across imaging modalities. Some assume the images are independent which may not be realistic in all settings, such as combining multiple tasks or task and resting-state.

The remainder of this dissertation aims to to develop state-of-the-art statistical methods to address some of the current questions in analysis of functional connectivity data. The proposed new methods incorporate and build upon some existing methods and algorithms, such as graphical models, SVM, ADMM, DNN, and boosting.

CHAPTER II

On Predictability of Individual Functional Connectivity Networks from Clinical Characteristics

2.1 Introduction

It is widely acknowledged that the integrated behavior of the approximately 100 billion neurons of the human brain in connected networks provides the substrate for complex behavior [68]. Growing interest in understanding the association between abnormal brain development and vulnerability to psychiatric disorders or symptoms has motivated recent research in functional brain connectivity, utilizing functional magnetic resonance imaging (fMRI) [64, 29, 70, 75]. Correlated activity of low frequency fluctuations of the blood-oxygenation level dependent (BOLD) signal provides a putative marker of large-scale networks, which may be used to find patterns among patients linked to specific clinical states. Identifying connectivity patterns associated with patient characteristics is clinically relevant for classifying high risk patients or identifying disease markers. Such fMRI scans, typically obtained while the brain is 'at rest,' i.e. not engaged in any particular task, capture dynamic activity across the brain with complex spatial and temporal covariance patterns. Extracting clinically meaningful information from these patterns represents important progress in

the analysis of fMRI images.

2.1.1 Motivating Data: The PNC Study

The motivating data set for this analysis is from the Philadelphia Neurodevelopmental Cohort (PNC) study [63]. The PNC is a community sample of 9,500 young persons from an urban hospital who presented for care for a wide range of physical and brain illnesses. Amongst this large sample, over 1,400 underwent neuroimaging, from which subjects for the current study were derived. The open source data set provides the opportunity to study the relationship between brain development and psychiatric symptoms.

The PNC study has been used to investigate abnormal brain development with an aim to identify youth at risk of developing psychiatric disorders [64, 41, 83]. Kessler et al [41] used independent component analysis to generate "growth charts" for functional brain network, linking this maturation to predict task outcomes. Xia et al. [83] identified functional connectivity patterns associated with four categories of psychopathology using canonical correlation analysis. These important findings assist in furthering knowledge of the link between brain development and psychopathology; in this work we aim to study this relationship at a more granular level of both connectivity and patient symptoms and characteristics. Using novel methods we aim to identify clinical characteristics that are predictive of functional connectivity.

2.1.2 Existing Methods

Several statistical methods have been proposed to estimate networks from these fMRI data, graphical models being one popular choice; see [72] for a recent overview of the functional connectivity methods. It has been shown that partial correlation, inverse covariance estimation, and Bayes net methods, are able to capture accurate connectivity estimation yet there are computational challenges in many cases. When

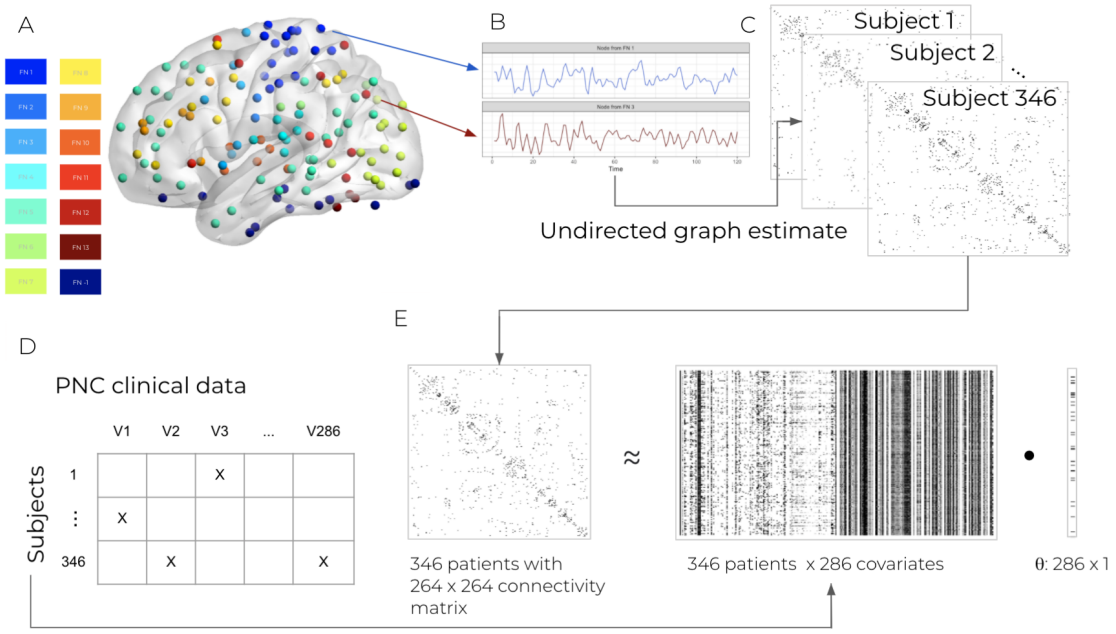


Figure 2.1: Depiction of use of imaging data and clinical characteristics to predict functional connectivity.

using undirected network estimation, there are often limitations regarding choice of link strength [73]. Though the drawbacks of various methods may be known, it is often difficult to compare approaches or decide on the best method given multiple options. Most approaches for comparison of networks involve basic, summary characteristics of a network. These summary metrics may not capture the true differences or similarities across metrics. Here the proposed method aims to utilize connectivity in a different way than existing methods to establish associations with clinical symptoms [73].

Understanding the predictive utility of functional connectivity is of critical interest to clinicians. Diseases like depression and Alzheimer's disease may be associated with unique connectivity patterns, but research to demonstrate true predictive power of these networks for clinical use is still in its infancy [70], [29]. On the other hand, predicting connectivity using clinical characteristics may be more accurate than the reverse; because the connectivity matrix is difficult to estimate, current methods that use that to model phenotype may be unreliable or infeasible. Clinicians are ulti-

mately interested in using connectivity to diagnosis or identify patients at greater risk of disease, and establishing symptoms that are predictive of connectivity is a reasonable step to advance this goal. Using clinical characteristics as predictors provides an approach to identifying associations between multiple phenotypic characteristics and functional brain connectivity, without relying on the accuracy of estimating a potentially noisy connectivity matrix.

2.1.3 Contribution

This work aims to provide a broad framework for constructing a reliable functional connectivity network and analyzing relationships with clinical characteristics using machine learning methods. Specifically, we propose a three-level model to specify the association between the voxel-level fMRI times series, the region-level brain networks and the clinical characteristics. We develop a fast computing method to estimate the model parameters and make prediction on brain functional connectivity by integrating different machine learning methods. A joint modeling framework is then used to improve connectivity estimation from the results of modeling connectivity with clinical characteristics. We illustrate the proposed method on evaluating the predictability of individual functional connectivity networks from the clinical characteristics in the PNC study, identifying the important functional sub-networks that are highly associated with psychiatry diseases such as Post-Traumatic Stress Disorder (PTSD) and Psychosis. We also perform a simulation study to evaluate the performance of the proposed framework in terms of the selection and prediction accuracy. In addition, we have developed an R package that implements the proposed method and provides a user-friendly software to study the association between functional brain networks and clinical characteristics, including the fMRI voxel-level time series prepossessing and the graphical presentation of the model fitting results. The R package will be freely available online after the paper is published and it is now

available upon request and on GitHub.

2.2 Methods

In this work, we propose a general modeling framework for analysis of brain functional connectivity and clinical characteristics. It consists of two major steps: 1) individual functional connectivity network construction, 2) covariate feature screening and network prediction. To improve estimation these two components are then iteratively jointly estimated using the ADMM algorithm.

Suppose we collect data from n subjects. We collect the resting-state fMRI signal with T scans, where the whole brain regions consist of V voxels and R regions. Let $i\,(i = 1, \ldots, n)$ index the subject, $v\,(v = 1, \ldots, V)$ index the voxels, $t\,(t = 1, \ldots, T)$ index the time scans. Let $r_v \in \{1, \ldots, R\}$ be the region index for voxel v. Let $y_{i,v,t}$ represent the observed resting state fMRI signals for subject i at voxel v and time t. For each subject, we also collect p covariates of clinical characteristics. Let $j\,(j = 1, \ldots, p)$ index the covariates and let $x_{i,j}$ denote the measurements of covariate j for subject i. Write $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top$.

2.2.1 A Generative Modeling Framework

We consider a generative modeling framework to specify the associations between the voxel-level fMRI time series, the region-level brain networks and the clinical characteristics. We consider a three-level hierarchical model.

At Level 1, we summarize the voxel level brain activity into region level signals:

$$y_{i,v,t} = \sum_{r=1}^{R} I(r_v = r)\tilde{\alpha}_{i,v,r}\tilde{y}_{i,r,t} + \epsilon_{i,v,t}, \tag{2.1}$$

where $\tilde{y}_{i,r,t}$ represents the summarized neural activity in region r at time t for subject i and $\alpha_{i,v,r}$'s are the weight coefficients that represent the contribution of voxel v to

region r for subject i. We assume the random error $\epsilon_{i,v,t}$ with mean zero and constant variance.

At Level 2, we model the region-level brain functional connectivity network. We consider a Gaussian copula graphical model. We transform each region specific signal $\tilde{y}_{i,r,t}$ into a latent variable $z_{i,r,t}$ according to the marginal distribution. We assume those latent variables follow a multivariate normal distribution. In particular, we have

$$\tilde{y}_{i,r,t} = F_{i,r}^{-1}\{\Phi(z_{i,r,t})\}, \quad \mathbf{z}_{i,t} = (z_{i,1,t}, \ldots, z_{i,R,t})^\top \sim \mathrm{N}(0, \mathbf{\Omega}_i^{-1}), \tag{2.2}$$

where $\mathbf{\Omega}_i = \{\omega_{i,r,r'}\}$ is an $R \times R$ precision matrix. The function $F_{i,r}(\cdot)$ is the cumulative distribution function of $\tilde{y}_{i,r,t}$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian distribution.

At Level 3, we impose sparsity on the precision matrix $\mathbf{\Omega}$. We introduce a latent selection indicator $a_{i,r,r'} \in \{0, 1\}$ for each region pair (r, r') to indicate whether the region r and r' are function connected for subject i; and for each region pair, we model the conditional distribution of $a_{i,r,r'}$ given the clinical characteristics $\mathbf{x}_i$ through a logistic regression model:

$$\omega_{i,r,r'} = \tilde{\omega}_{i,r,r'} a_{i,r,r'}, \quad \mathrm{logit}\{\mathrm{E}(a_{i,r,r'})\} = g_{r,r'}(\mathbf{x}_i), \tag{2.3}$$

where $g_{r,r'}(\cdot)$ is an unknown function representing the log odds of region r and r' being functional connected for subject i with clinical characteristics $\mathbf{x}_i$.

2.2.2 Individual functional connectivity network estimation

We consider two fast computing methods for Level 1 estimation: averaging voxel level signals within each region and using principal component analysis (PCA) to summarize region level. Specifically in the first case, taking a simple average of

voxels uses $\tilde{\alpha}_{i,v,r} = 1$. We will assume equal weight across all voxels and average within regions:

$$\tilde{y}_{i,r,t} = \frac{1}{\sum_v I(r_v = r)} \sum_{v:r_v=r} y_{i,v,t}, \tag{2.4}$$

An alternative estimation approach uses principal component analysis to assign weights, that procedure estimates $\tilde{\alpha}_{i,v,r}\tilde{y}_{i,v,t}$ with the first principal component.

Our network estimation for Level 2 implements the Meinshausen and Bühlmann method of estimating a sparse graphical model [91]. We have selected this method in part due to the flexibility to relax the normality assumption often imposed on observations in graphical models. Through fitting semi-parametric Gaussian copula models, this approach aims to better recover the true underlying undirected graph structure [49].

In particular, let $\Phi^{-1}(\cdot)$ be the quantile function of the standard Gaussian distribution. We have $z_{i,r,t} = \Phi^{-1}\left\{F_{i,r}\left(\tilde{y}_{i,r,t}\right)\right\}$ and the connectivity matrix is estimated as

$$\hat{\mathbf{\Omega}}_i = \arg\min_{\mathbf{\Omega}} \left[\mathrm{tr}\{\mathbf{\Omega} S_T(\mathbf{Z}_i)\} - \log|\mathbf{\Omega}| + \lambda\|\mathbf{\Omega}\|_1\right], \tag{2.5}$$

where $S_T(\mathbf{Z}_i)$ is the sample covariance of the transformed region-level connectivity signals $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,R})$, $|\mathbf{\Omega}|$ is the determinant of $\mathbf{\Omega} = \{\omega_{r,r'}\}$, λ is a regularization parameter and $\|\mathbf{\Omega}\|_1 = \sum_{r,r'} |\omega_{r,r'}|$ is the entry-wise L_1 norm. The solution of (2.5) enjoys the sparsity, thus we estimate the latent connectivity indicator to obtain the functional connectivity matrix for the entire brain $\mathbf{A}_i = (a_{i,r,r'})$ with

$$a_{i,r,r'} = I(\omega_{i,r,r'} \neq 0),$$

We use this network construction as the connectivity outcome when fitting the prediction models in the Level 3 estimation.

2.2.3 Prediction of Network Features

We can estimate the relationship between the covariates and connectivity of regions r and r', $g_{r,r'}(\cdot)$ using the initial estimate of the functional connectivity network. Although the framework presented in Section 2.2.1 is flexible and may incorporate more general models, we start from a linear model for simplicity:

$$\text{logit}\{\text{E}(a_{i,r,r'})\} = \sum_{j=1}^{p} \beta_{r,r',j} x_{i,j}, \tag{2.6}$$

where $x_{i,j}$ is the measured value for covariate j of subject i.

With a large number of region pairs and covariates to consider, it may be difficult to perform model estimation. A more efficient approach is to carry out a screening mechanism before fitting edgewise prediction models. We proposed to implement the elastic net regression to screen out clinical variables that are not associated with a given edge connection. We only consider predicted edges where at least five percent of subjects have a connection.

Machine learning methods are increasingly popular for predictive modeling. We consider two common machine learning methods to predict edgewise connectivity: SVM and random forests [17, 6].

When performing SVM we aim to minimize the following loss function for each region pair r, r':

$$\begin{aligned} &\min_{\boldsymbol{\theta}_{r,r'}} \left\{ \frac{1}{2} \mathbf{w}_{r,r'}^{\top} \mathbf{w}_{r,r'} + c_{r,r'} \sum_{i=1}^{n} \xi_{i,r,r'} \right\}, \\ &\text{such that} \quad (2a_{i,r,r'} - 1)\{\mathbf{w}_{r,r'}^{\top} \boldsymbol{\phi}(\mathbf{x}_i) + b\} \geq 1 - \xi_{i,r,r'}, \end{aligned} \tag{2.7}$$

where $\boldsymbol{\theta}_{r,r'} = \{\mathbf{w}_{r,r'}, b_{r,r'}, \{\xi_{i,r,r'}\}_{i=1}^{n}\}$ and $\boldsymbol{\phi}(\mathbf{x})$ is a vector of features in the transformed feature space derived from the kernel $k(\mathbf{x}, \mathbf{x}')$, such that $k(\mathbf{x}, \mathbf{x}') = \boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{x}')$. In this setup $\mathbf{w}_{r,r'}$ denotes the weight vector used to maximize the margin around the

hyperplane separating subjects with and without a connected edge between region pairs r and r' in the network. The penalty term $c_{r,r'} \sum_{i=1}^{n} \xi_{i,r,r'}$ is used to penalize for observations that are misclassified.

Random forest implements a series of decision trees, where individual trees form based on minimizing the residual sum of squares. In particular, for each region pair r, r', we aim to predict the functional connectivity $a_{i,r,r'}$ using clinical characteristics $\mathbf{x}_i$ using classification probability $m_{r,r'}(\mathbf{x}) = \Pr(a_{i,r,r'} = 1 \mid \mathbf{x}_i = \mathbf{x})$. We model $m_{r,r'}(\mathbf{x})$ as an ensemble of M randomized regression trees, i.e.

$$m_{r,r'}(\mathbf{x}) = \frac{1}{M} \sum_{j=1}^{M} m(\mathbf{x}, \mathcal{T}_{r,r',j}), \qquad m(\mathbf{x}, \mathcal{T}_{r,r',j}) = \sum_{s=1}^{S} p_{r,r',j,s} I(\mathbf{x} \in \mathcal{A}_{r,r',j,s}), \quad (2.8)$$

where $m(\mathbf{x}, \mathcal{T}_{r,r',j})$ is the classification probability given $\mathbf{x}_i = \mathbf{x}$ by the jth tree for region pairs r, r'. Each tree $\mathcal{T}_{r,r',j}$ consists of a tree-based partition $\{\mathcal{A}_{r,r',j,s}\}_{s=1}^{S}$ of the sample space $\mathcal{X}$ with $\mathcal{X} = \bigcup_{s=1}^{S} \mathcal{A}_{r,r',j,s}$ and $\mathcal{A}_{r,r',j,s} \cap \mathcal{A}_{r,r',j,s'} = \emptyset$ for $s \neq s'$ and the corresponding classification probability $p_{r,r',j,s}$ for partition $\mathcal{A}_{r,r,'j,s}$.

For the tuning parameters, we use the standard settings for SVM, using the normal kernel function and soft margin classification, and tune the number of trees and number of candidate variables at each split for random forest. Cross validated Area under the curve (AUC) is used to evaluate the predictive performance of edge classification.

The procedure of estimating the functional connectivity network and then fitting prediction models to each edge will be referred to as the 2-step update. These estimates will be used as the initial values in the iterative algorithm detailed in the next section.

2.2.4 ADMM Algorithm

Jointly estimating the individual functional connectivity network and the model estimates for prediction of network features is done using the ADMM algorithm. Let

$\mathbf{S}_i = S_T(\mathbf{Z}_i)$, $\mathbf{G}(\mathbf{x}_i;\boldsymbol{\theta}) = \{g_{r,r'}(\mathbf{x}_i;\boldsymbol{\theta})\}_{R\times R}$ and $\mathbf{L}(\mathbf{x}_i;\boldsymbol{\theta}) = \{l_{r,r'}(\mathbf{x}_i;\boldsymbol{\theta})\}$ be three $R \times R$ matrices, where $l_{r,r'}(\mathbf{x}_i;\boldsymbol{\theta}) = \log[1 + \exp\{g_{r,r'}(\mathbf{x}_i;\boldsymbol{\theta})\}]$. The algorithm minimizes the following objective function

$$\underset{\boldsymbol{\Theta}}{\text{minimize}} \sum_{i=1}^{n} [-\log\det \boldsymbol{\Omega}_i + \text{tr}(\boldsymbol{\Omega}_i\mathbf{S}_i) + \lambda\|\boldsymbol{\Omega}_i\|_1 - \gamma\text{tr}\{(2\mathbf{A}_i - 1)\boldsymbol{\Omega}_i\} \\ -\text{tr}\{\mathbf{A}_i\mathbf{G}(\mathbf{x}_i;\boldsymbol{\theta}_{r,r'})\} + \mathbf{L}(\mathbf{x}_i;\boldsymbol{\theta}_{r,r'})] \tag{2.9}$$

where $\boldsymbol{\Theta} = \{\{\boldsymbol{\Omega}_i\}_{i=1}^n, \{\mathbf{A}_i\}_{i=1}^n, \boldsymbol{\theta}\}$. $\boldsymbol{\Omega}_i$ is an $R \times R$ symmetric positive definite matrix, $\mathbf{A}_i \in \{0,1\}^{R\times R}$ is a $R \times R$ binary matrix, and $\boldsymbol{\theta}$ is the parameter in the logistic regression. λ and γ are tuning parameters. Note that the term $-\text{tr}\{\mathbf{A}_i\mathbf{G}(\mathbf{x}_i;\boldsymbol{\theta})\} + \mathbf{L}(\mathbf{x}_i;\boldsymbol{\theta})$ is equal to the negative log-likelihood of logistic regression for network edge predictions. To estimate $\boldsymbol{\Omega}$ and $\mathbf{A}$, in each iteration of the ADMM algorithm, we minimize the object function in three steps with respect to $\boldsymbol{\Omega}$, $\boldsymbol{\theta}$, and $\mathbf{A}$. Each is sequentially updated and then iterated until the algorithm converges. Suppose the initial values are $\boldsymbol{\Omega}_i^{(0)}$, $\mathbf{A}_i^{(0)}$ and $\boldsymbol{\theta}^{(0)}$.

2.2.4.1 Minimize with respect to Ω

In the kth iteration ($k = 1, 2, \ldots,$), we first update $\boldsymbol{\Omega}_i^{(k)}$ by minimizing the objective function with respect to $\boldsymbol{\Omega}_i$ and fixing $\mathbf{A}_i$ at the previous iteration, i.e. $\mathbf{A}_i^{(k-1)}$, for $i = 1, \ldots, n$,

$$\min_{\{\boldsymbol{\Omega}_i\}_{i=1}^n} \sum_{i=1}^{n} \Big[-\log(\det(\boldsymbol{\Omega}_i)) + \text{tr}(\boldsymbol{\Omega}_i\mathbf{S}_i) + \lambda\|\boldsymbol{\Omega}_i\|_1 - \gamma \cdot \text{tr}\{(2\mathbf{A}_i^{(k-1)} - 1\}\boldsymbol{\Omega}_i)\Big] \tag{2.10}$$

To implement the ADMM algorithm, we introduce $\mathbf{Y}_i^{(k)}$ and let $\mathbf{Z}_i^{(k)} = \boldsymbol{\Omega}_i^{(k)} - \mathbf{Y}_i^{(k)}$, we now minimize the following objective function with respect to $\boldsymbol{\Omega}_i$

$$\boldsymbol{\Omega}_i^k = \arg\min_{\boldsymbol{\Omega}_i} \left\{ -\log(\det(\boldsymbol{\Omega}_i)) + \frac{\mu}{2} \left\| \boldsymbol{\Omega}_i + \left(\mathbf{Z}_i^{(k-1)} - \mathbf{Y}_i^{(k-1)} + \frac{1}{\mu}\mathbf{S}_i - \frac{\gamma}{\mu}(2\mathbf{A}_i^{(k-1)} - 1) \right) \right\|^2 \right\}$$

Taking the derivative with respect to $\mathbf{\Omega}_i$, and defining $\mathbf{Y}_i - \mathbf{Z}_i^{(k-1)} - \frac{1}{\mu}\mathbf{S}_i + \frac{\gamma}{\mu}(2\mathbf{A}_i - 1) = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^\top$ with $\mathbf{\Lambda}_i = \text{diag}(\lambda_1, \ldots, \lambda_R)$ results in the following equations to solve $\mathbf{\Omega}_i^{(k)}$.

$$\begin{aligned}
\mathbf{0} &= -\mathbf{\Omega}_i^{-1} + \mu\mathbf{\Omega}_i - \mu\{\mathbf{Y}_i^{(k-1)} - \mathbf{Z}_i^{(k-1)} + \frac{1}{\mu}\mathbf{S}_i + \frac{\gamma}{\mu}(2\mathbf{A}_i - 1)\} \\
\mathbf{0} &= -F_\mu^{-1}(\mathbf{\Lambda}_i) + \mu F_\mu(\mathbf{\Lambda}_i) - \mu\mathbf{\Lambda}_i
\end{aligned}$$

The solution takes the form:

$$\mathbf{\Omega}_i^{(k)} = \mathbf{U}_i F_\mu(\mathbf{\Lambda}_i)\mathbf{U}_i^\top = \frac{1}{2}\mathbf{U}_i \left\{ \text{diag}\left(\lambda_{i1} + \sqrt{\lambda_{i1}^2 + \frac{4}{\mu}}, \ldots, \lambda_{iR} + \sqrt{\lambda_{iR}^2 + \frac{4}{\mu}}\right)\right\} \mathbf{U}_i^\top, \tag{2.11}$$

where $F_\mu(\mathbf{\Lambda}_i) = \text{diag}\{f_{i1}, \ldots, f_{iR}\}$ with $f_{ir} = \frac{1}{2}\left(\lambda_{ir} + \sqrt{\lambda_{ir}^2 + \frac{4}{\mu}}\right)$ for $r = 1, \ldots, R$.

2.2.4.2 Minimize with respect to θ

Next to update $\boldsymbol{\theta}^{(k)}$ we minimize the objective function (eq. 2.9) with respect to $\boldsymbol{\theta}$, given the current estimates $\mathbf{\Omega}_i^{(k)}$ and $\mathbf{A}_i^{(k-1)}$. This update is the same as solving for $\boldsymbol{\theta}$ in logistic regression using the negative log likelihood.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left[-\text{tr}\{\mathbf{A}_i^{(k-1)}\mathbf{G}(\mathbf{x}_i; \boldsymbol{\theta})\} + \mathbf{L}(\mathbf{x}_i; \boldsymbol{\theta}) \right] \tag{2.12}$$

For each node pair (r, r') this becomes:

$$\theta_{r,r'}^{(k)} = \arg\min_{\theta_{r,r'}} \sum_{i=1}^{n} \left[-a_{r,r',i}^{(k-1)}\, g_{r,r'}(\mathbf{x}_i; \boldsymbol{\theta}) + \log\left(1 + \exp\{g_{r,r'}(\mathbf{x}_i; \boldsymbol{\theta})\}\right) \right] \tag{2.13}$$

If we suppose $g_{r,r'}(\mathbf{x}_i; \boldsymbol{\theta}) = x_i\theta$, such as in the case of logistic regression, we can obtain the logistic regression estimates for $\boldsymbol{\theta}$. Similarly the log-odds can be estimated from SVM or random forest instead of logistic regression.

2.2.4.3 Minimize with respect to A_i

Next to update $\mathbf{A}_i^{(k)}$ we minimize the objective function (2.9) with respect to $\mathbf{A}_i$, fixing estimates of $\mathbf{\Omega}_i^{(k)}$ and $\boldsymbol{\theta}^{(k)}$.

$$\min_{\mathbf{A}_i} \sum_{i=1}^{n} \Big[\gamma \operatorname{tr}\{(2\mathbf{A}_i - 1)\mathbf{\Omega}_i^{(k)}\} - \operatorname{tr}\{\mathbf{A}_i \mathbf{G}(\mathbf{x}_i; \boldsymbol{\theta}^{(k)})\} + \mathbf{L}(\mathbf{x}_i; \boldsymbol{\theta}^{(k)}) \Big] \tag{2.14}$$

Note that $\mathbf{A}_i \in \{0,1\}^{R \times R}$. For each region pair (r, r'), we minimize this function by comparing the objective function for $a_{i,r,r'} = 0$ and $a_{i,r,r'} = 1$ given the values in $\mathbf{A}_i$ at all other pairs.

Equivalently, we can minimize the following for each subject i at each pair (r, r'):

$$a_{i,r,r'}^{(k)} = \underset{a_{i,r,r'} \in \{0,1\}}{\arg\min} \Big[a_{i,r,r'} \{2\gamma \omega_{i,r,r'} - g_{r,r'}(\mathbf{x}_i, \boldsymbol{\theta}^{(k)})\} - \omega_{i,r,r'} \Big] \tag{2.15}$$

Combining these steps the full algorithm is presented in Algorithm 1.

Data: $\{\mathbf{\Omega}_i^{(1)}, \mathbf{A}_i^{(1)}, \mathbf{x}_i\}_{i=1}^n$; Number of iterations K; Updating rate μ; Penalty term λ and tuning parameter γ.

Result: $\{\mathbf{\Omega}_i^{(K)}, \mathbf{A}_i^{(K)}\}_{i=1}^n$

begin

Initialize $\mathbf{\Omega}_i^{(1)}$ and $\mathbf{A}_i^{(1)}$

for $k = 1, \ldots, K$ **do**

for $i = 1, \ldots, n$ **do**

$\mathbf{Y}_i^{(k)} = UF_\mu(\Lambda_i)U^\top = \frac{1}{2}U\left(diag\left(\lambda_{ir} + \sqrt{\lambda_{ir}^2 + \frac{4}{\mu}}\right)\right)U^\top$

$$\mathbf{\Omega}_i^{(k)} = \begin{cases} Y_i^{(k)} + Z_i^{(k)} - \frac{\lambda}{\mu} & \text{if } Y_i^{(k)} + Z_i^{(k)} \geq \frac{\lambda}{\mu} \\ Y_i^{(k)} + Z_i^{(k)} + \frac{\lambda}{\mu} & \text{if } Y_i^{(k)} + Z_i^{(k)} \leq \frac{\lambda}{\mu} \\ 0 & \text{otherwise} \end{cases}$$

Update $Z_i^{(k)} = Z_i^{(k)} + \mu\,(\mathbf{\Omega}_i^{(k)} - Y_i^{(k)})$.

end

for *each pair* (r, r') **do**

Estimate $g_{r,r'}(\mathbf{x}; \boldsymbol{\theta}^{(k-1)})$ using logistic regression, SVM, or random forest.

end

for $i = 1, \ldots, n$ **do**

$F_k = \mathrm{tr}\{\gamma\,(2\mathbf{A}_i^{(k)} - 1)\mathbf{\Omega}_i^{(k)} - \mathbf{A}_i^{(k)}\mathbf{G}(\mathbf{x}_i; \boldsymbol{\theta}^{(k)})\}$

for *each* (r, r') **do**

if $a_{i,r,r'}^{(k)} = 0$ **then**

if $F_k + \gamma\,\omega_{(r,r')}^{(k)} - \mathbf{x}_i\boldsymbol{\theta}_{r,r'}^{(k)} < F_k$ **then**

$a_{i,r,r'} = 1$

end

else

if $F_k - \gamma\,\omega_{(r,r')}^{(k)} < F_k$ **then**

$a_{i,r,r'} = 0$

end

end

end

end

end

end

Iterate until the estimates converge. The choices of initial values for $\mathbf{\Omega}$ and $\mathbf{A}$ and $\boldsymbol{\theta}$ are described in Sections 2.2.2 and 2.2.3. Iteratively updating these estimates could

improve the predictive performance and power to detect true associations compared to the proposed framework without ADMM.

The choice of λ impacts the level of sparsity when estimating Ω. Individual level λ_i is used to control the sparsity of the initial estimates of $\mathbf{\Omega}_i$ in step 2.2.2. We have opted to use the same subject specific λ_i in the ADMM algorithm, though the results do not change significantly in simulation for a common population level λ. Similarly the ratio between λ and μ impacts the sparsity of the estimate for Ω, so μ is chosen to satisfy the desired level of sparsity. γ is used to control the similarity between $\mathbf{A}$ and $\mathbf{\Omega}$. For this reason we increase γ over iterations of k, presumably as the two estimates converge towards the same sparsity pattern this parameter enforces that relationship.

2.3 Simulation

2.3.1 Performance of the two-step update

The prediction procedure, detailed in Section 2.2, was evaluated using simulated data based on the real data application. True signals in the clinical variables were simulated by generating a β vector and using the observed clinical covariates to generate the corresponding time series data for 346 subjects. The ability to recover the true signals was evaluated by comparing the subset of variables selected by elastic net to those with true non-zero signals in β. This performance was summarized using sensitivity, specificity, and false discovery rate (FDR). FDR is defined as the proportion of true signals identified out of all selected variables. The performance of the prediction procedure was evaluated using AUC, comparing the ability to correctly classify edges as connected or not.

The data generation process was as follows: simulate β, use observed clinical variables and simulated β to assign connectivity to each edge for each subject, simulate

precision matrix Ω from a mixture of normal distributions with mean 3 or -3 and standard deviation 1 for connected edges, ensure Ω is positive definite, and finally simulate T time points for each node of each subject from a multivariate normal distribution with mean zero and covariance Ω.

Simulation was also used to understand how characteristics of the fMRI data impacts the Gaussian copula network estimation, implemented using the R package *huge* [91]. We considered how a differing number of time points in the fMRI and a different network size affected the ability to recover the connectivity matrix. Table 2.1 presents the results for each of the settings considered. We compared results using *huge* to other network construction methods, *clime* and *tiger*, and found that *huge* outperformed other existing methods for estimating sparse graphical models [45]. Due to space limitations, we did not report the detailed results here.

# of Nodes	# of time points	% of connected edges recovered	% of edges not connected recovered
50	120	47%	99%
	1000	54%	98%
264	120	12%	99%
	1000	51%	99%
	5000	73%	99%

Table 2.1: Results of simulation evaluating how well the connectivity estimation can recover the simulated connectivity network.

Though we do not achieve high power with this process, we consistently see good control in the false discovery rate; edges that are assigned to be connected are typically correctly labeled, and more frequently those incorrectly labeled are among edges that are connected in truth and not connected in the network estimate. This leads to the conclusion that signals identified from this procedure are likely to be true associations, though some true associations may be missed.

In Table 2.2 (a-c) we present the number of subjects that are correctly and in-

correctly identified as having a connection at a given edge, using SVM to predict connectivity. We compare the results to both the true simulated network and the network estimate. This is an example using one edge across 346 subjects.

		True Network	
		# connected	# not connected
Connectivity Matrix Estimate	# connected	15	0
	# not connected	10	321

(a) Comparison of network estimated through graphical modeling to the true simulated network.

		True Network	
		# connected	# not connected
Predicted Network	# connected	7	4
	# not connected	18	317

(b) Comparison of true simulated network to the predicted connectivity using SVM.

		Connectivity Matrix Estimate	
		# connected	# not connected
Predicted Network	# connected	6	5
	# not connected	9	326

(c) Comparison of network estimated through graphical modeling to the predicted connectivity using SVM.

Table 2.2: Simulation results for network estimation using the prediction procedure and *huge*.

Table 2.2 presents the accuracy of the network estimate compared to the true network connectivity in addition to the predicted network compared to the estimated and true network. Table 2.2a shows that the network estimate has a very low false discovery rate, with no edges incorrectly identified as connected. Using the network estimate to perform the prediction, we can evaluate the performance of the prediction results (Table 2.2c). This simulation provides evidence that we can be confident in the network estimate and prediction results compared to the true underlying network structure. Though some signals of connected edges are missed, those identified as connected most often are connected in the underlying network.

Simulation was also used to evaluate the ability to recover true signals among the clinical characteristics. When simulating 3,814 edges we observed an average FDR of about 11%. Average sensitivity is only 5% and specificity is 99%. The true signal is sparse, among 286 variables there are 20 true non-zero signals randomly selected for each edge. Again we conclude that there is low power in the procedure but good false discovery control. The variables selected are likely to be true associations, although many true signals will be missed. We expect this to be the case when using a relatively large network structure compared to a small number of subjects.

2.3.2 Performance of the ADMM algorithm

Potential improvement due to using a joint estimation procedure was quantified through simulation. We considered a toy example for demonstration purposes with a small network with 10 nodes and a sample of 100 subjects. Because the estimates from the two-step update are the initial values for the ADMM algorithm, the performance depends on how well the 2-step estimation performs, detailed in section 2.3.1.

In this simulation setting the estimation of both $\mathbf{A}$ and $\boldsymbol{\theta}$ are evaluated. Estimation of $\mathbf{A}$ is evaluated using sensitivity, specificity, and FDR in terms of correctly identifying the connected and not connected edges. Similarly estimation of $\boldsymbol{\theta}$ is evaluated in terms of sensitivity, specificity, and FDR, but defined in terms of the accuracy of clinical characteristics with non-zero effect estimates.

Table 2.3 compares three methods using the ADMM algorithm: logistic regression, SVM, and random forest. Each is also compared to the initial values of the algorithm, obtained via the 2-step update using graphical lasso and SVM. The results in Table 2.3 are the average of 100 iterations of the simulated setting. Increasing the effect size of non-zero signal in the simulation leads to higher sensitivity across all the methods, though it does not reduce FDR of $\mathbf{A}$ estimation in most cases. Increasing the sample size also seems to lead to improved sensitivity and lower FDR in some

Setting			**A** Estimation			$\boldsymbol{\theta}$ Estimation		
N	β	Method	Sens	Spec	FDR	Sens	Spec	FDR
100	3	Logistic	0.61	0.86	0.50	0.52	0.68	0.21
		SVM	0.57	0.80	0.60	0.23	0.92	0.12
		RF	0.33	0.97	0.37	0.18	0.93	0.12
		2-step	0.20	0.73	0.63	0.00	1.00	0.00
100	5	Logistic	0.73	0.86	0.50	0.64	0.63	0.20
		SVM	0.67	0.77	0.60	0.32	0.88	0.12
		RF	0.43	0.97	0.32	0.24	0.91	0.12
		2-step	0.20	0.74	0.63	0.00	1.00	0.00
300	5	Logistic	0.87	0.86	0.42	0.97	0.48	0.21
		SVM	0.77	0.74	0.60	0.48	0.77	0.19
		RF	0.49	0.98	0.28	0.41	0.87	0.11
		2-step	0.69	0.98	0.12	0.00	1.00	0.00

Table 2.3: Performance of ADMM algorithm using different methods to perform the prediction: logistic regression, SVM, and random forest (RF). Several simulation settings are presented with a different sample size (N) and average effect size of non-zero $\boldsymbol{\Omega}$ (β). Estimation of the adjacency matrix ($\mathbf{A}$) is evaluated as an average of the following metrics over 100 iterations: sensitivity (Sens), specificity (Spec), and FDR. Accuracy of the selected clinical characteristics ($\boldsymbol{\theta}$) is measured using sensitivity, specificity, and FDR.

cases (logistic regression and random forest). In terms of identifying covariates that are truly associated with connectivity patterns ($\boldsymbol{\theta}$ Estimation in Table 2.3), increasing sample size leads to much higher power with an increase in sensitivity from 0.52 with 100 subjects to 0.97 using 300 subjects when applying logistic regression in the ADMM algorithm. In terms of the false discovery rate among these covariates, it is not too large with a maximum of about 20% across all methods. SVM and random forest seem to perform better than logistic regression in terms of controlling FDR, which leads to slightly lower power in terms of lower sensitivity. With these simulation settings the 2-step update does not identify any variables to be included using elastic net, which demonstrates that incorporating the ADMM algorithm provides a more sensitive approach.

2.4 Data Application

2.4.1 PNC Data

Extensive assessment of behavior, life events, demographics and neuropsychological performance was obtained on all subjects in the PNC, in addition to performing resting state fMRIs. The sample has children with mental illnesses as well as healthy individuals, and all subjects underwent a structured neuropsychiatric interview to establish the presence, duration and effect of multiple psychiatric symptoms (if present) on functioning. The broad range of psychiatric disorders assessed included: depression, mania, simple phobia, social phobia, generalized anxiety, separation anxiety, social anxiety, panic disorder, obsessive-compulsive disorder, post traumatic stress disorder (PTSD), eating disorders, and psychosis.

MRI data for the resting state scans were obtained with BOLD-sensitive image acquisitions over 6 minutes, with 120 frames, each 2 seconds in duration. Voxel size was 3 by 3 in the transverse plane and 3 in the axial plane, yielding approximately 100,000 voxels in the brain [63]. Each voxel constituted a time series. The number of measurements in the time series and the temporal sampling rate are both fixed by the type of MRI performed. Data preprocessing was done to correct for timing differences in the acquisition, realign individual subject scans, and map the images to a common anatomical space so that the images could be combined and analyzed across subjects. This preprocessing occurred with the pipeline used for the 1000 Functional Connectomes Project.

Subjects with excessive movement (defined as subjects with greater than 0.25 mm volume-to-volume displacement) or poor scan quality were excluded from the analysis, which yielded 500 scans from the original 1,442 subjects scanned. Images were also spatially filtered to reduce residual anatomic variability, and for each time voxel, a time series was extracted and bandpass filtered (0.1 - 0.01 hz) to remove physiological

artifact from respirations and heart rate. In addition, regressors for white matter and cerebral spinal fluid were obtained, and variability from these additional sources of noise was removed. From a set of a priori nodes (see Figure 2.3), a time series from each node (10 mm sphere) was extracted, and a cross-correlation matrix of Pearson r-values was obtained and Z-transformed for each of the 264 nodes with every other node. The PNC data was obtained from the NCBI database of Genotypes and Phenotypes (dbGaP), a publicly available database. Informed consent was obtained for all subjects who participated in the PNC study; the original publication states "Participants had been previously enrolled in a genomics study at CAG and they and/or their parents had provided informed consent (assent) to be re-contacted for participation in additional studies such as this one. The institutional review boards of both the University of Pennsylvania and the Children's Hospital of Philadelphia approved all study procedures" [63].

2.4.2 Analysis Pipeline for PNC Data Analysis

Preprocessing of the PNC imaging data was done to reduce bias from motion and other known confounders. Additional steps were taken to reduce the sample and network size based on missingness and variability. The entire procedure using the PNC data is summarized in Figure 2.2.

Figure 2.2 shows the steps performed to manipulate the data into a workable form, as well as steps for data reduction. The rfMRI data was converted from time series for each node into the individual connectivity estimates (Panel A-B). Then the connectivity estimates were stacked across subjects (Panel C). We reduced the brain network to edges with sufficient variability for modeling, removing those connected in fewer than five percent of subjects (Panel D). For the clinical characteristics we reduced the subset of potential variables based on missingness; if a given variable was missing in more than five percent of subjects it was removed from the analysis (Panel

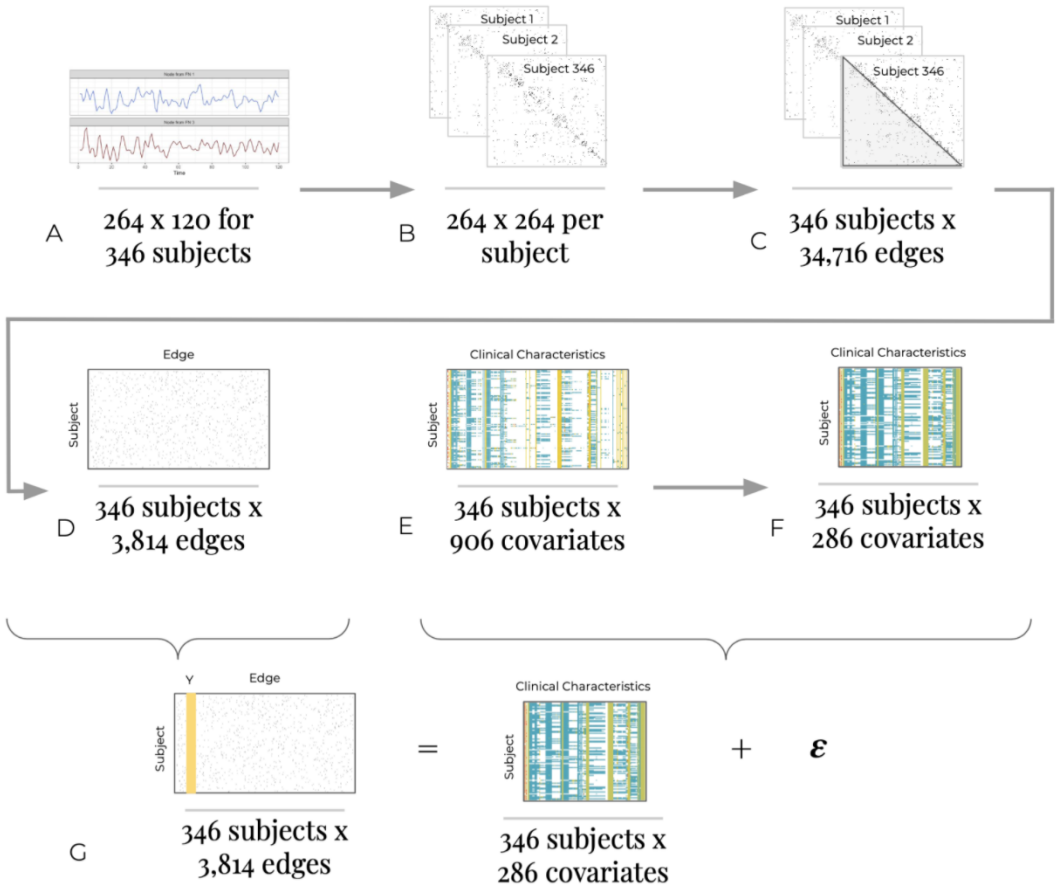


Figure 2.2: Procedure for PNC data analysis.

E-F). This conveniently reduced the number of variables (p) to fewer than the sample size (n), reducing to an $n > p$ problem. Lastly Panel G depicts the final model fitted associating the connectivity across subjects at a given edge and the clinical characteristics.

2.4.3 Brain Networks of Interest

For the analysis that follows, we used an a priori anatomic parcellation of the brain, which utilized 264 nodes organized into 13 different functional modules (FM), identified by Power et al. in the 2011 paper *Functional Network Organization of the Human Brain* [59].

The networks, made up of several regions, identified in this paper align well with

others proposed, such as the default mode network [28], dorsal and ventral attention [22], and fronto-parietal task control. The authors classified the remaining subnetworks by associated functions, including visual processing, memory, sensory and motor control, auditory, and somatosensory. Compared to voxel-based approaches to connectivity, these networks should minimize connectivity contributions from image smoothness, which causes adjacent voxels in an image to have very high correlation coefficients, irrespective of functional connections. Compared to anatomically defined nodes, such as automated anatomic labeling (AAL [79]) these units more likely reflect intrinsic functional organization in the brain and may be more meaningful probes of functional brain networks.

#	Function	#	Function
1	Sensory/somatomotor Hand	8	Fronto-parietal Task Control
2	Sensory/somatomotor Mouth	9	Salience
3	Cingulo-opercular Task Control	10	Subcortical
4	Auditory	11	Ventral attention
5	Default Mode	12	Dorsal attention
6	Memory Retrieval	13	Cerebellar
7	Visual	−1	Uncertain

Table 2.4: Power brain functional modules and associated brain functions [59].

Figure 2.3 shows the location of the nodes in the power parcellation, and colors indicate membership to the thirteen identified functional networks.

2.4.4 Predictability of Network

The ability to predict the entire functional brain network was evaluated in addition to the ability to predict connectivity within the subnetworks of interest. We found that the ability to capture the entire brain network was very limited using the machine learning methods tested, SVM and random forest. The average performance of predicting a connection was no better than choosing at random, with an average

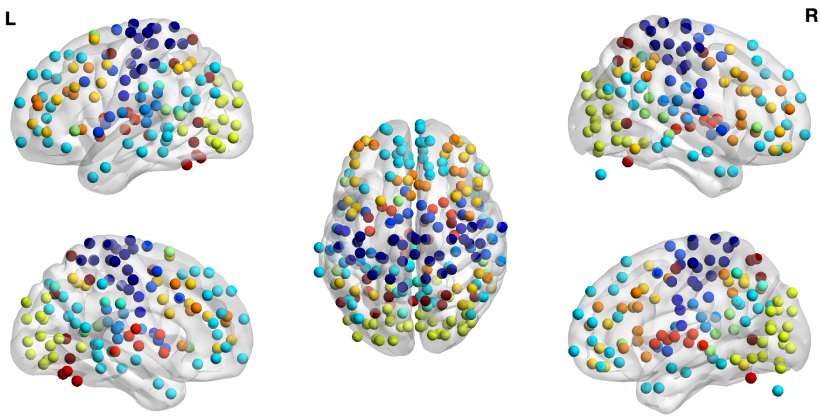


Figure 2.3: Power 264 node spatial parcellation. Each color represents one of the 13 functional brain networks of interest. Generated using BrainNet Viewer [84].

AUC of 0.5 when using 5-fold cross validation. However, when looking within specific regions we do see fairly high predictive performance for some. In the final iteration of the ADMM algorithm, only 480 edges had any clinical variables selected for inclusion in a model, so the following results reflect the performance within that subset of the network. Note that the 'truth' for computation of AUC was defined as the estimate of $\mathbf{A}$ from the ADMM algorithm before updating with random forest or SVM results, since we do not know the true underlying connectivity as we did in the simulations.

We observed the greatest ability to detect connected edges on average in the subcortical network (10) using random forest. Table 2.5 contains the average AUC, range of AUC from first to the third quartile, and the maximum AUC for a given edge within each network for both methods, random forest and SVM. Though on average the AUC within some networks was not high, we did have good performance (AUC > 0.95) for some edges in the following functional networks: somatomotor hand, auditory, default mode, visual, fronto-parietal task control, salience, and subcortical. These networks had the highest maximum AUC observed, but the most edges with good performance were in the somatomotor hand, default mode, and visual networks. The variables selected in the model with high predictive performance (highest AUC) for connectivity in the sensory and somatomotor hand include: race, indicator of liver

disease, indicator of infectious disease, four questions from the Structured Interview of Psychosis-risk Syndromes survey, an indicator of social anxiety, and questions from the Penn Age Differentiation Test, Penn Emotion Differentiation Test, and Visual Object Learning Test. Some of these variables are expected to be associated with this brain network, whereas the link between other variables is less obvious, as the sensorimotor network is not commonly associated with psychopathology. However, serious mental disorders, such as schizophrenia and bipolar disorder, show imbalances in this network [3, 52]. This finding provides an example where this technique might identify new associations between psychopathology and network dysfunction.

One model with AUC > 0.95 predicting connectivity for an edge from the visual network includes the following variables: indicator of vision problems and two components of the 1-Back trials. These variables are expected to be associated with activation in the visual processing network, and confirm these known associations.

Functional Brain Module	SVM		Random Forest	
	Mean (Q1- Q3)	Max	Mean (Q1- Q3)	Max
1 Sensory/somatomotor Hand	0.45 (0.45, 0.52)	0.59	0.67 (0.52, 0.89)	1.00
2 Sensory/somatomotor Mouth	0.34 (0.32, 0.40)	0.40	0.54 (0.38, 0.62)	0.86
3 Cingulo-opercular Task Control	0.44 (0.44, 0.48)	0.60	0.60 (0.55, 0.73)	0.76
4 Auditory	0.47 (0.46, 0.50)	0.69	0.69 (0.57, 0.80)	1.00
5 Default Mode	0.47 (0.45, 0.53)	0.59	0.63 (0.48, 0.77)	1.00
6 Memory retrieval	0.49 (0.49, 0.49)	0.49	0.45 (0.45, 0.45)	0.45
7 Visual	0.45 (0.41, 0.51)	0.57	0.64 (0.52, 0.70)	1.00
8 Fronto-parietal Task Control	0.49 (0.46, 0.51)	0.60	0.54 (0.43, 0.55)	1.00
9 Salience	0.41 (0.44, 0.50)	0.57	0.69 (0.49, 0.85)	1.00
10 Subcortical	0.57 (0.55, 0.59)	0.59	**0.83 (0.66, 1.00)**	1.00
11 Ventral Attention	0.40 (0.35, 0.46)	0.49	0.66 (0.55, 0.76)	0.79
12 Dorsal attention	0.49 (0.44, 0.54)	0.56	0.60 (0.41, 0.71)	0.83
13 Cerebellar	0.54 (0.54, 0.54)	0.54	0.44 (0.44, 0.44)	0.44

Table 2.5: Mean AUC for edges contained within each functional module from two methods, SVM and random forest.

The variables selected for two models with the best predictive performance are presented in Tables 2.6 and 2.7.

Variable	Description	Estimate
Race	Self-reported ethnicity of participant (EA, AI)	-0.63
MED807	Liver disease	0.33
MED809	Infectious disease	-4.25
SIP015	SIPS feeling odd things going on	-0.09
SIP016	SIPS feeling able to predict the future	-0.07
SIP018	SIPS Feeling different due to superstitions	-0.10
SOC001	Feeling afraid in social settings	-0.04
PADT	Penn Age Differentiation Test	-0.34
PADT	Number of correct responses with no age difference	-0.01
PEDT	Penn Emotion Differentiation Test	-1.30
VOLT	Visual Object Learning Test	0.18

Table 2.6: Variables selected for the model with AUC = 1.00 (using random forest) from network 1, associated with sensory somatomotor hand control. SIPS is the Structured Interview of Psychosis-risk Syndromess; PADT is the Penn Age Differentiation Test; PEDT is the Penn Emotion Discrimination Test; VOLT is the Visual Object Learning Test.

Variable	Description	Estimate
MED622	Vision problems	-0.75
PFMT	Penn Face Memory Test	-0.18
PEIT	Penn Emotion Identification	-1.09
LNB	Number of correct responses to 1-back trials	0.12
LNB	Number of incorrect responses to 1-back trials	-0.05

Table 2.7: Variables selected for the model with AUC = 1.00 (using random forest) from network 7, associated with visual processing. PFMT is the Penn Face Memory Test; PEIT is the Penn Emotion Identification Test; LNB is the Penn Letter N-Back test which tests working memory.

Several variables were frequently selected across models with good performance of classifying connected nodes, these include: indicator of ear/nose/throat problems, indicator of metabolic disease, separation anxiety, having thoughts of suicide, and results from the Penn Conditional Exclusion Test. The Penn Conditional Exclusion Test is designed to assess executive functioning ability.

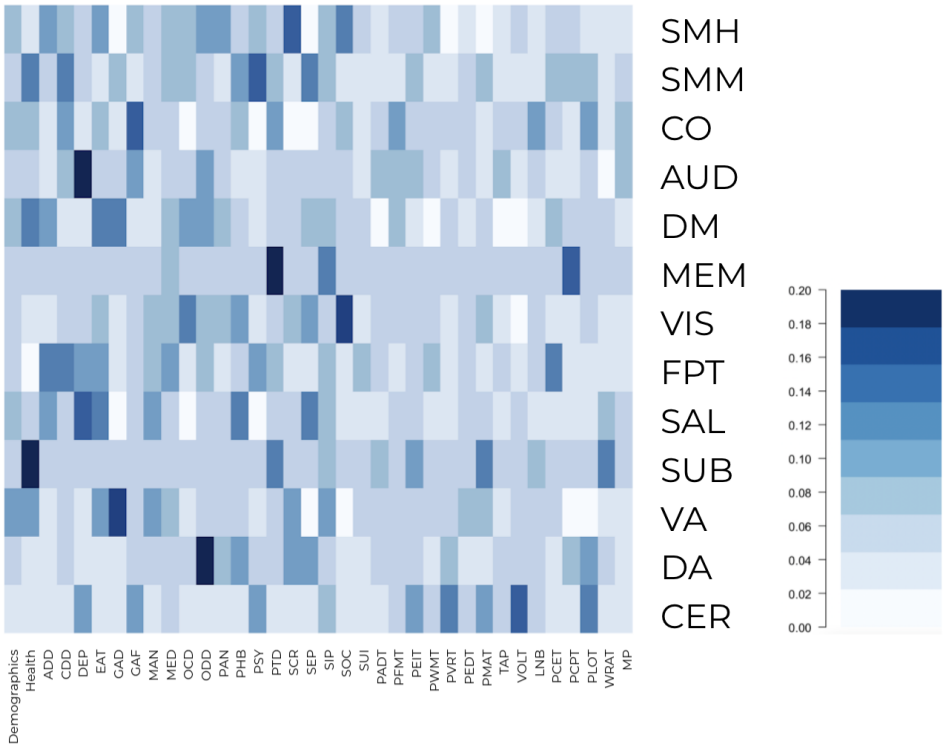
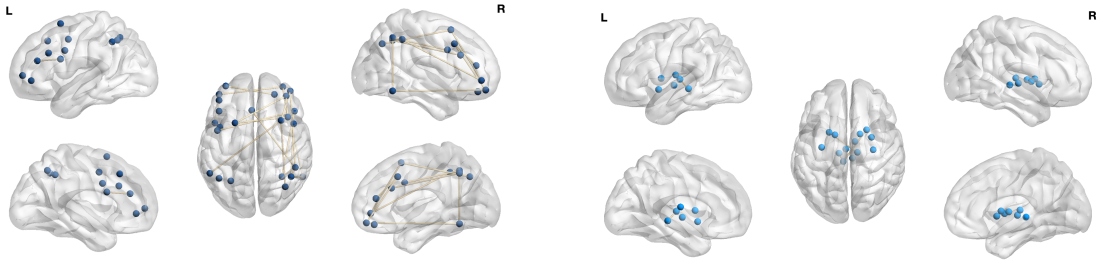


Figure 2.4: Proportion of models which a variable in the group was selected across edges in each functional module.

Figure 2.4 presents the proportion of models for each functional module that selected at least one of the clinical characteristics in the groups. The groups and abbreviations are as follows: demographics, overall medical metrics (health), Attention Deficit Disorder (ADD), Conduct Disorder (CDD), Depression (DEP), Eating Disorder (EAT), Generalized Anxiety Disorder (GAD), Children's Global Assessment Scale (GAF), Mania/Hypomania (MAN), other medical conditions (MED), Obsessive Compulsive Disorder (OCD), Oppositional Defiant Disorder (ODD), Panic Disorder (PAN), Specific Phobia (PHB), Psychosis (PSY), Post-Traumatic Stress (PTD), general probes about counseling and emotions (SCR), Separation Anxiety (SEP), Structured Interview for Prodromal Symptoms (SIP), Social Anxiety (SOC), Suicide (SUI), Penn Age Differentiation Test (PADT), Penn Facial Memory Test (PFMT), Penn Emotion Identification Test (PEIT), Penn Word Memory Test (PWMT), Penn Verbal Reasoning Test (PVRT), Penn Emotion Differentiation Test (PEDT), Penn Matrix Reasoning Test (PMAT), Tap hand trials (TAP), Visual Object Learning Test (VOLT), Letter N-Back test (LNB), Penn Conditional Exclusion Test (PCET), Penn Continuous Performance Test (PCPT), Penn Line Orientation Test (PLOT), Wide Range Assessment Test (WRAT), and Penn Motor Praxis Test (MP). These variable groupings were defined by the surveys and tools used to collect the data in the PNC study.

2.4.5 Association Analysis

Beyond the ability to predict connectivity in the network, it is often useful to understand characteristics or symptoms associated with levels of connectivity in the resting state brain. One of the main goals of the PNC study was to establish associations between brain development and psychiatric diseases. Another important way to interpret the results, is to identify variables that were frequently associated with higher connectivity within each of the thirteen functional networks.



(a) Fronto-parietal Task Control (b) Subcortical

Figure 2.5: Edges associated with SIPS or psychosis variables within fronto-parietal task control and subcortical functional networks.

We can build a brain network associated with one disease or symptom of interest. For example, the network built of edges associated with psychosis spans all thirteen functional networks; for specific diseases we may be interested in how connectivity of a specific network, rather than the entire brain, relates to the disorder of interest. Figure 2.5 shows the edges within the fronto-parietal task control network that are associated with psychosis variables (variables from the SIPS and psychosis surveys). This brain network has been identified in previous studies of schizophrenia [69]. We identified fronto-parietal task control and subcortical networks as most frequently associated with SIPS or psychosis variables. The subcortical networks, specifically the striatum, heavily enervated by dopaminergic projections from the midbrain, and thalamic nuclei, have been implicated in the pathophysiology of schizophrenia [36, 82, 16].

When we consider Post-traumatic Stress Disorder (PTSD), we see associations similar to those previously observed in other studies. Connections associated with PTSD span four behavioral networks. Previous research has shown that men with PTSD have differing rates of connectivity between the right amygdala and other subnetworks of the brain when compared to control patients [75]. Figure 2.6 presents these connections across the four functional networks associated with PTSD.

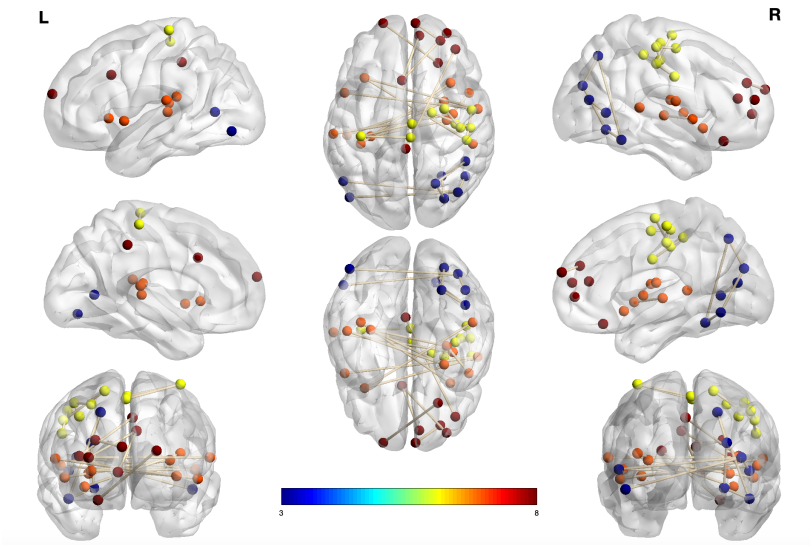


Figure 2.6: Network for indicator of Post-traumatic Stress Disorder. Nodes colored by corresponding functional networks.

2.5 Discussion

In this work, we present the framework for predicting functional connectivity with clinical characteristics, and demonstrate that it is feasible to predict some subnetworks in the brain. Many clinical characteristics identified in the PNC application are consistent with previous findings, as the simulation results suggest the ability to identify some true signals among clinical characteristics. Other findings suggest novel associations between behavioral measures and brain networks. While the field of psychiatry has traditionally relied on diagnostic categories, dating back to clinical observations made in the late 19th century, there has been increasing recognition for new methods of classification, such as theoretical constructs derived from neuroscience research [37]. The framework illustrated here is an atheoretical, purely data-driven approach that, lacking constraining assumptions, has the potential to provide new insights into brain and behavior correlations. Though it is difficult to detect associations with relatively few subjects compared to network size and number of clinical characteristics, this work provides a way to identify subnetworks that are predictive given clinical characteristics of interest.

Future extensions of this work may involve making valid statistical inferences on parameters and predictions. In addition, changing the graph estimation procedure or voxel level summary could influence the ability to predict connectivity. These extensions may provide further insight into the relationship between functional brain connectivity and behavioral phenotypes, relevant for psychopathology.

CHAPTER III

Scalar-on-network Regression via Boosting

3.1 Introduction

Brain function is dictated by the coordinated activity of many functional networks consisting of billions of neurons. These networks are captured through the resting-state functional magnetic resonance imaging (rfMRI), which measures neural activity via the blood oxygen level dependent (BOLD) signal when the subject is at rest, or when no particular task is performed. The synchronized activation of brain regions defines a functional module. Often in brain imaging, the Power 264 spatial parcellation is used to summarize the hundreds of thousands of voxels captured in images into regions of interest in the brain. This spatial parcellation is used to estimate resting state functional networks that are well defined and reproducible, as shown by *Power et al.* [59]. Many of these regions of interest have previously been associated with various functions of the brain, such as visual processing or task control, and the networks formed by their connectivity can provide further insight into brain functioning.

Functional brain connectivity has a wide range of clinical applications: patterns in functional connectivity have been associated with Alzheimer's Disease, development of psychosis, depression, and other psychiatric and neurological diseases [76, 66, 30, 21]. Current use of functional connectivity in clinical practice aims to compare patterns in groups of patients impacted or not impacted by a disorder, whereas

using connectivity patterns diagnostically or prognostically for individual patients is an ongoing research area for most diseases. With diverse clinical applications there are also diverse methodological approaches to consider. In some applications models have image responses and scalar predictors, commonly referred to as image-on-scalar regression [14, 93]. This approach may provide insight into spatial patterns and heterogeneity across individuals. For example, spatial varying coefficient models can be used to characterize the association between patient characteristics, like age, and neuroimaging networks [94, 46, 88]. In addition, image-on-scalar regression may reveal common patterns in connectivity across subjects or identify an underlying population-level pattern [90].

Other research focuses on optimizing estimation of functional networks from the time series resulting from imaging data, such as [81, 49, 58]. The first two examples focus on correlation-like metrics to establish the functional network, whereas the third paper proposes a Bayesian approach to generate a directed graph estimate of connectivity. Depending on the goals of the study directed graphs may be useful, although undirected estimates tend to be more believable especially when working with resting-state data. This paper will use a correlation-based metric to estimate the functional network, which we take to be well estimated for the context of this paper.

Another way to utilize functional connectivity is to use the connectivity patterns to predict clinical outcomes. Some existing literature using this framework relies on dimension reduction methods to make the statistical methods more tractable. Some examples include [19] which summarizes the use of independent component analysis (ICA) with rfMRI data and [83] which utilizes canonical correlation analysis to estimate the correlation between dimensions of psychopathology and brain structures. Alternatively, scalar-on-image may be employed to address similar questions. Several Bayesian methods have been developed to perform high dimensional variable selection

in the scalar-on-image context [44, 31, 40, 25]. The recent work by [31] leverages the properties of a spike-and-slab prior and transitivity of associations between nodes in the brain to perform variable selection.

This chapeter focuses on scalar-on-network regression; the main distinction from existing work is incorporation of the natural grouping across regions of interest in the brain employed in a machine learning algorithm. This has the benefit of providing interpretable associations that may be used to inform clinical practice. In this work, we will propose a two-stage modeling approach in which the first stage summarizes the image at the functional network level and the second stage selects individual edges that are associated with the outcome.

3.1.1 Motivating Data: ABCD Study

The ABCD Study is a longitudinal study aiming to understand how brain development is impacted by substance use. It is an ongoing multi-site study done across the U.S. following adolescents for about ten years, with current follow up of about three years. There are 10,000 children enrolled, and a subset of about 1,800 had rfMRI data released in conjunction with the first release of data. Additional information on these children, including their age (in months), gender, race, parents' education level, parents' marital status, and family income will also be utilized in addition to their rfMRI data. One of the outcomes of interest in this study is general cognitive ability, measured by a neurocognitive battery of twelve tests and summarized using Bayesian principal component analysis [77], and one motivating question for this research is how to relate a measure of cognition to changes in functional connectivity.

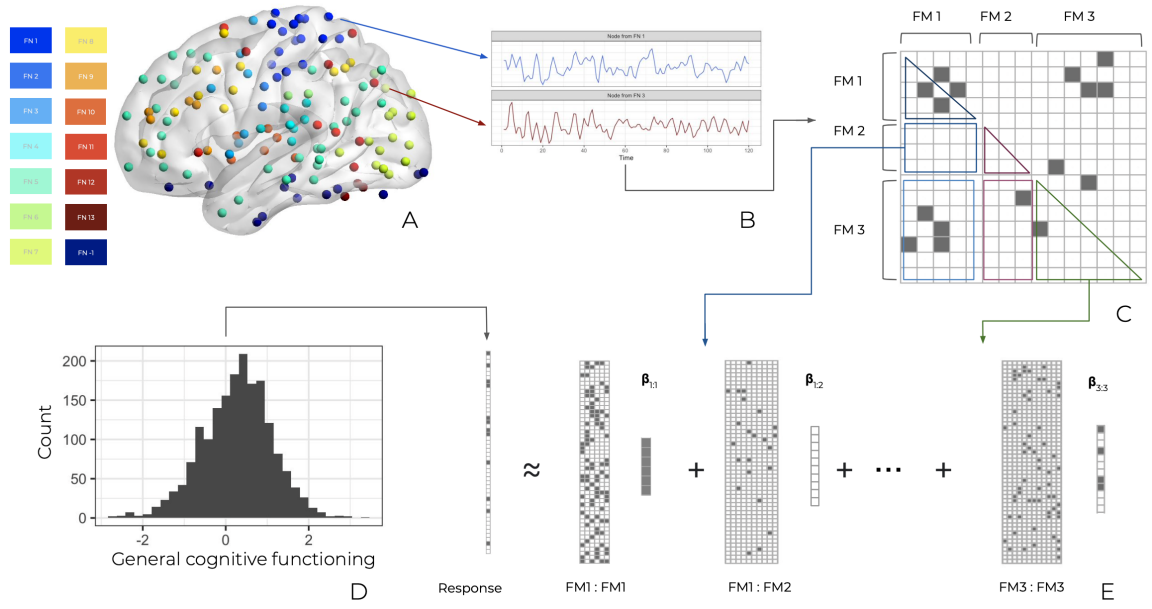


Figure 3.1: Illustration of the scalar-on-network regression framework and the process of converting time series data into network data from the rfMRI images.

Figure 3.1 illustrates the rfMRI data processing and general workflow to identify brain connectivity patterns associated with cognitive functioning. The rfMRI voxel-level data is summarized into 264 regions of interest (ROIs) as defined by the Power spatial parcellation [59], with a time series at each ROI (Figure 3.1 A-B). Here, the ROIs are grouped into thirteen known functional modules and one uncategorized group based on *Power et al.* [59] grouping, though alternative network partitions could be considered such as *Smith et al.* [71]. Node membership to functional modules like the default mode network (FM 5) is well accepted, whereas some other functional modules, such as the Power memory retrieval network (FM 6), are less understood and may differ more depending on the network partition. The correlation between any two ROIs is then used to construct a connectivity matrix for each subject, and each ROI is grouped into a functional module (Figure 3.1 C). The association between some clinical outcome or characteristic (Figure 3.1 D) and the connectivity patterns by functional modules are then analyzed (Figure 3.1 E). In the real data there are more functional modules, this is simplified to illustrate the method more clearly.

3.1.2 Existing Methods

In the first stage of the proposed method, we consider selecting groups of edges together rather than a single edge associated with the outcome. This can be viewed as a group selection method, similar to group lasso [87]. Another method that uses a similar scalar-on-network regression approach, but for classification, is proposed in *Relión et al.* [61]. This method uses a penalty similar to group lasso to select edges and nodes within the network to perform classification, working differently than the method proposed here which allows more general regression. Other work has aimed to investigate a similar question of identifying brain regions associated with a clinical outcome but utilizing dimension reduction to address the complexity of brain networks, such as the work done by [74] using principal component analysis. Most existing literature in neuroimaging that addresses a similar question, relies on dimension reduction to efficiently model and estimate associations; however, this leads to some trade-offs between loss of information and precision. In this work, we aim to identify some weak signals that may be missed by methods such as those that perform dimension reduction.

The proposed method utilizes a previously developed boosting algorithm, intended for variable selection in high dimensional settings [10]. Boosting was developed to improve prediction by combining weak learners, with its beginning in the AdaBoost algorithm [53]. This method evolved into various other boosting methods including tree-based boosting, gradient boosting, and likelihood-based boosting. L_2 Boosting is one such form of boosting that uses the gradient descent algorithm with squared error loss [9]; this form of boosting has been shown to perform well in high-dimensional settings such as in brain imaging or genomic applications [26].

In this work we develop a two-stage boosting algorithm to estimate the sparse edge-dependent regression coefficients by leveraging the knowledge of brain functional organization. This proposed method has higher power to detect the true signals while

controlling the false discovery rate better than existing approaches, when the edge-wise predictive effects within the sub-networks appear to be homogeneous. This method differs from existing approaches by not relying on dimension reduction to handle the large brain network structure, while still maintaining computational efficiency. Analyzing the ABCD rfMRI data identifies several edges that are consistently selected as associated with general cognitive functioning, most of these edges belong to four functional modules: default mode, cingulo-opercular task control, visual, and dorsal attention.

3.2 Methods

3.2.1 Scalar-on-network Regression

Suppose the data set consists of n subjects. Let i $(i = 1, \ldots, n)$ be the index of subjects. For each subject i, we observe a scalar outcome variable, denoted as $y_i \in \mathbb{R}$ and the subject-specific brain functional network consisting of q ROIs (nodes). Let (j, k) be the index of node pairs (edges) in the network for $1 \leq k < j \leq q$. Let $I = \{(j, k) : 1 \leq k < j \leq q\}$ be the whole edge set of interest in the network. Let $x_{ijk} \in \mathbb{R}$ represents the functional connectivity measure between nodes j and k for subject i. We refer to $\mathbf{X}_i = \{x_{ijk} : (j, k) \in I\}$ as a fully connected network for subject i. In addition, for each subject, we collect a set of p_A additional covariates that may affect the outcome variable. Let z_{il} denote the lth $(l = 1, \ldots, p_A)$ additional covariate. Write $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip_A})^\top$. In our motivating example, for each subject i, y_i is the general cognitive ability score and $\mathbf{X}_i$ represents the lower triangular part of the correlation matrix of the region-level rfMRI time series on 264 Power ROIs. The additional covariates $\mathbf{z}_i$ include age, gender, socioeconomic status, etc. We consider

a scalar-on-network regression model as follows:

$$y_i = \alpha_0 + \sum_{l=1}^{p_A} \alpha_l z_{il} + \sum_{(k,j)\in I} \beta_{jk} x_{ijk} + \epsilon_i, \tag{3.1}$$

where α_0 is the intercept; the coefficient β_{jk} captures the effect of edge (j,k) in the network on the outcome variable; and parameters $\{\alpha_l\}_{l=1}^{p_A}$ reflect the confounding effects of the additional covariates (non-network) predictors. The term ϵ_i represents the random noise. We assume that $E(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$.

When a natural group structure is present and known in the network, we may want to leverage this information to improve the estimation of β_{jk} and more accurately select non-zero β_{jk}. One way to utilize this information is to impose group structure on the nodes or edges of the network $\mathbf{X}_i$. Based on the group information, we can partition I into G disjoint groups, i.e., $I = \bigcup_{g=1}^{G} I_g$ and $I_g \cap I_{g'} = \emptyset$. We refer to $\{x_{ijk} : (j,k) \in I_g\}$ as the sub-network g for subject i. We may hypothesize that the effect of edges within each of these sub-networks on the outcome is similar.

Applying this to the setting of functional brain connectivity, we may expect that edges spanning within each functional module, e.g. the default mode network, have a similar effect on cognition or that none of the edges spanning another functional module have an effect on cognition. With this assumption, for group $g = 1, \ldots, G$, we decompose the effect of each edge in the group g as follows: for $(j,k) \in I_g$,

$$\beta_{jk} = \bar{\beta}_g + \omega_{jk}, \tag{3.2}$$

where $\bar{\beta}_g$ represents the average effect of the sub-network g and ω_{jk} represents the individual edge effect. To ensure the model identifiability, we impose a constraint on ω_{jk}, i.e.,

$$\sum_{(j,k)\in I_g} \omega_{jk} = 0, \tag{3.3}$$

which implies that $\bar{\beta}_g = m_g^{-1} \sum_{(j,k)\in I_g} \beta_{jk}$ with m_g being the number of edges in sub-network g. If β_{jk} is similar for all edges within sub-network g, then ω_{jk} will be close to zero for all pairs $(j,k) \in \mathcal{I}_g$. By combining (3.1), (3.2) and (3.3), we have the following scalar-on-network regression model with sub-network partition:

$$y_i = \alpha_0 + \sum_{l=1}^{p_A} \alpha_l z_{il} + \sum_{g=1}^{G} \lambda_g \bar{x}_{ig} + \sum_{(j,k)\in I} \omega_{jk} x_{ijk} + \epsilon_i, \tag{3.4}$$

where $\lambda_g = m_g \bar{\beta}_g$ represents the total effect of the sub-network g and $\bar{x}_{ig} = m_g^{-1} \sum_{(j,k)\in I_g} x_{ijk}$ the average of functional connectivity measurements of the edges in sub-network g.

Figure 3.2 illustrates an example of partitioning the edges and corresponding effects into groups, where 15 nodes are grouped by brain functional modules and generate six sub-networks (groups of edges). In this setup, the dark blue edges representing those contained within the first functional module will have the effect $\bar{\beta}_1 + \omega_{jk}$ on cognition.

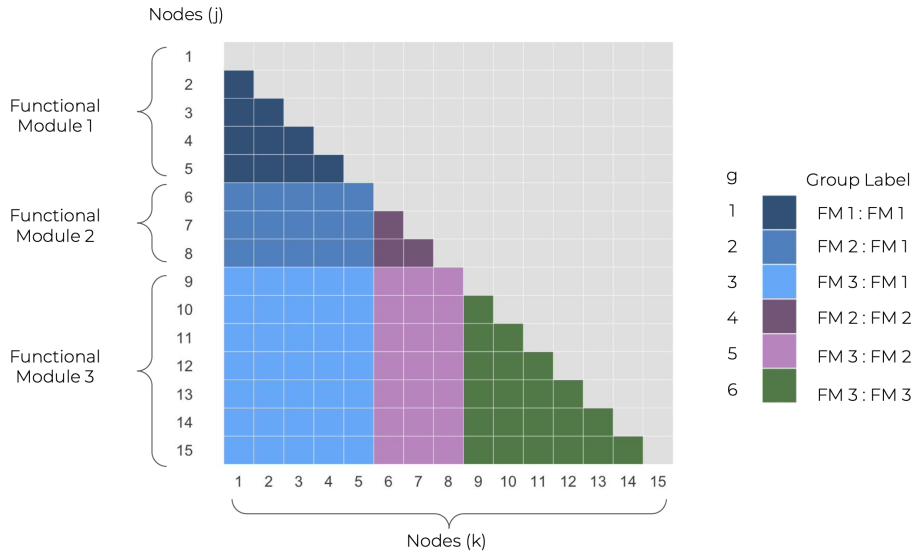


Figure 3.2: Example of network $\mathbf{X}_i$ with 15 nodes to illustrate how the edges in the network are partitioned by group ($G = 6$), using the functional modules to identify groups.

We will use the group effect λ_g to select groups which have a meaningful effect on

the outcome. As a secondary step we can estimate ω_{jk} to more accurately estimate the effect of each specific edge on the outcome.

3.2.2 Algorithm

The following two working models will be used to model the generative modeling framework described in Section 3.2.1. The final result will estimate the parameters in model (3.4), capturing the individual and group effect components. Let $z_{i0} = 1$ for $i = 1, \ldots, n$.

Working Model 1 (WM1):

$$y_i = \sum_{l=0}^{p_A} \alpha_l z_{il} + \sum_{g=1}^{G} \lambda_g \bar{x}_{ig} + \tilde{\epsilon}_i, \tag{3.5}$$

where $\tilde{\epsilon}_i = \sum_{(j,k)\in I} \omega_{jk} x_{ijk} + \epsilon_i$.

Working Model 2 (WM2):

$$\tilde{y}_i = \sum_{l=0}^{p_A} \alpha_l z_{il} + \sum_{(j,k)\in I} \omega_{jk} x_{ijk} + \epsilon_i, \tag{3.6}$$

where $\tilde{y}_i = y_i - \sum_{g=1}^{G} \lambda_g \bar{x}_{ig}$.

The first working model will be used to obtain estimates for λ_g; the second model will be conditional on the estimate of λ_g and ω_{jk} will be updated.

Instead of simply using an average of the edges within a group to define Z_i, we may consider a weighted average of the edges. This will allow us to weaken the assumption of homogeneity among the effects of edges within each group. One alternative formulation of Z_i we may consider is:

1. Identify strong signals within each group, using lasso to identify these.

2. Average the weak effects with the strong effects to upweight those weak effects to be more similar to the strong signals.

3. Average across weighted weak signals and the strong signal edges to define Z_i.

Alternately, it may be more efficient to consider using the first working model to screen out groups that do not have an effect on the outcome. Then the individual level updates of ω_{jk} may not be added to the group level effect λ_g. In this setting, we define the second working model as the following.

WM2*:

$$y_i = \sum_{l=0}^{p_A} \alpha_l z_{il} + \sum_{(j,k)\in S} \omega_{jk} x_{ijk} + \epsilon_i, \tag{3.7}$$

where $S = \bigcup_{g:\lambda_g \neq 0} I_g$ is the subset of edges where the total effects of sub-network g is nonzero, i.e. $\lambda_g \neq 0$, which can be estimated from the first stage. If there is more homogeneity among the effect of edges within groups, we expect this alternative approach to be more efficient.

One way to implement the group selection is by using the L_2 boosting algorithm [10]. The algorithm can be adapted to update a group of variables at the same time; in this case we minimize the loss function over a group of variables rather than just one at a time. For simplicity, we start from the case without confounding effects, i.e. $\alpha_l = 0$ for $l = 0, 1, \ldots, p_A$. See the following algorithm:

Data: $\{\mathbf{x}_i, y_i\}_{i=1}^n$; Number of iterations M and M'; Updating rate υ

Result: $\{\lambda_g\}$ and $\{\omega_{jk}\}$

begin

Initialize $\lambda_g = 0\ (g = 1, \ldots, G)$ and $\omega_{jk} = 0$ for $(j,k) \in I$.

for $m = 1, \ldots, M$ **do**

for $g = 1, \ldots, G$ **do**

Compute the first partial derivative with respect to λ_g:

$L_1(g) = \sum_{i=1}^n (y_i - \sum_{g'=1}^G \lambda_{g'} \bar{x}_{ig'}) \bar{x}_{ig}$.

end

Find $g^* = \text{argmin}_g\ |L_1(g)|$.

Calculate the second partial derivative with respect to g^*:

$L_2(g^*) = -\sum_{i=1}^n \bar{x}_{ig^*}^2$

Update $\lambda_{g^*} = \lambda_{g^*} - \upsilon\, m_{g^*}\, L_2(g^*)^{-1} L_1(g^*)$

end

Next estimate the ω_{jk} for edges $(j,k) \in I_g$ where $\lambda_g \neq 0$.

for $i = 1, \ldots, n$ **do**

$\tilde{y}_i = y_i - \sum_{g=1}^G \lambda_g \bar{x}_{ig}$.

end

Initialize $\omega_{jk} = 0,\ (j,k) \in S = \bigcup_{g:\lambda_g \neq 0} I_g$.

for $m = 1, \ldots, M'$ **do**

for $g = 1, \ldots, G$ **do**

Compute the first partial derivative with respect to ω:

$L_1(j,k) = \sum_{i=1}^n \{\tilde{y}_i - \sum_{(j',k') \in S} \omega_{j'k'} x_{ij'k'}\} x_{ijk}$.

end

Find $(j^*, k^*) = \text{argmin}_{(j^*,k^*)}\ |L_1(j,k)|$.

Calculate the second partial derivative with respect to $\omega_{j^*k^*}$:

$L_2(j^*, k^*) = -\sum_{i=1}^n x_{ij^*k^*}^2$

Update $\omega_{j^*k^*} = \omega_{j^*k^*} - \upsilon L_2(j^*, k^*)^{-1} L_1(j^*, k^*)$

end

end

The first loop over M iterations is used to estimate WM1 and the second loop for M' iterations estimates WM2* conditional on $\hat{\lambda}_g$ estimates from the first stage.

One of the difficulties observed in simulations with respect to tuning parameters is that the step size will vary depending on the sample size. One way to make step size sample size free is to use the average log likelihood instead of the sum of the log likelihood when updating λ and ω. The other important tuning parameter is the number of iterations to run the algorithm at each stage. In the first stage, we are less concerned about overfitting because we may be able to correct for this in the second stage; for that reason, we use AIC to determine the stopping point. In the second stage, we use BIC or EBIC to determine the stopping point, because these tend to stop sooner than AIC and lead to less overfitting [13].

This method has been implemented as an R package using Rcpp and is available on GitHub at https://github.com/EmilyLMorris/GroupBoosting.

3.2.3 Handling Adjustment Variables

Estimation of the adjustment variables can be handled in different ways; one approach is to treat them as nuisance parameters. When estimating these effects is not of interest, we can simply regress out the effect of the additional covariates. In this approach, the response (y_i) in the algorithm is replaced with the residuals (ϵ_i) obtained from the regression model

$$y_i = \sum_{l=0}^{p_A} \alpha_l z_{il} + \epsilon_i \tag{3.8}$$

This approach in the simulation studies will be referred to as 'Group Boosting with Residuals.'

In some settings, the association between these covariates and the outcome is also of interest in addition to the image associations. To address this, the covariates can

also be included in the first step of the boosting algorithm for selection. The only change in the algorithm is that at each step, instead of selecting the optimal g, the optimal g or l will be selected. Either the estimate for one group or one adjustment variable is selected and updated at each step. Then similar to the groups selected in the first step of the algorithm, any selected adjustment variables will be included for possible selection again in the second step. This approach will be referred to as 'Group Boosting with Adjustment Variables'.

3.2.4 Theoretical Properties

This section summarizes the main theoretical properties of this method. Broadly, we show that the first stage of the group boosting algorithm provides a close approximation to the true values when edges in the network are not grouped. The deviation from the true value is a function of the average value of β, the max deviation within a group from β_g, and the expected value of the squared summation of all the elements in group g. We also argue that we can use Buhlmann's proof of consistency for boosting in high dimensional settings [10]. The properties will be investigated separately for two cases: homogeneity of the effects within groups and heterogeneity of the effects within each group.

We will denote the summary of edges within a given group as Z_{ig} and the individual covariates as X_{ijk}, such that $Z_{ig} = \frac{\sum_{jk \in g} X_{ijk}}{m_g}$. The networks are defined by g and m_g is the number of edges in group g.

The properties of this method vary based on the assumptions made on the group effects. In some cases, one may expect the effect of all edges in a given region to be the same, whereas in other cases more heterogeneity may be assumed. Consider the following cases: one homogeneous group, one heterogeneous group, multiple homogeneous groups, and multiple heterogeneous groups.

Suppose for the simplest case that all the edges belong to one group and each

edge has the same effect on the outcome. The only difference between this approach and standard L_2Boosting without group selection is that we force all the variables to have the same effect on the outcome - homogeneity within a single group. First we will assume this to the be the true underlying model, so we are estimating with a correctly specified model, where we assume $\beta_1 = \beta_2 = \ldots = \beta_{m_g}$. In the case that the true model is $\beta_1 = \ldots = \beta_{m_g}$, this approach may be more efficient because we force the estimates to be equal within the group.

In this case, we can show that the expected difference between the true Z and the estimated $\tilde{Z}$ is zero, $\tilde{Z}\ E\{\|Z - \tilde{Z}\|_2^2\} = 0$. This result expands to the case where we have multiple groups and the effect within each group is homogeneous. In this case we also see that the expected squared error of $\|Z - \tilde{Z}\|$ is zero.

If we don't assume that the effect is homogeneous within every group, the expected difference of $\|Z - \tilde{Z}\|$ is bounded. Define the heterogeneity among a group of edges as the difference from the average effect, then we can define $\beta_{jk} = \bar{\beta}_g + \delta_{jk}$, where $\bar{\beta}_g$ is the average effect of an edge in group g on the outcome. δ_{jk} represents the deviation of the effect of edge (j, k) on the outcome from the average of the group's effect. If we assume

$$E\left\{\frac{\sum_{(j,k)\in g}\delta_{jk}x_{ijk}}{\sum x_{ijk}}\right\} = 0 \tag{3.9}$$

and that $|\delta_{jk}| < \epsilon$ for all $(j, k) \in g$, then $\|Z - \tilde{Z}\|$ is bounded by a function of the average value of β within groups, the max deviation any β has from the average (ϵ), and the expected value of the square of the summation of all elements in group g. See Appendix A.1 for more detailed derivations of these properties.

$$E\left\{\sum_{i=1}^{n}(Z_i - \tilde{Z}_i)^2\right\} = \frac{n}{m_g}(\bar{\beta}_g\epsilon^2 + 2\epsilon^3 + 3\bar{\beta}_g^{-1}\epsilon^4 + 2\bar{\beta}_g^{-2}\epsilon^5 + \bar{\beta}_g^{-3}\epsilon^6)E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\} \tag{3.10}$$

Our working model framework can be written as two linear regression models. *Bühlmann et al.* [10] proves consistency for L_2 boosting with linear models in the high dimensional setting. In the simple case when the model is not misspecified, the effect size within groups is entirely homogeneous, we can apply the theory and corresponding proof established by Buehlmann to show consistency. The conditions assumed for that theorem to hold are satisfied in this setting.

A1. $p_n = O(exp(Cn^{1-\xi})$, for sample size n, number of possible predictors p_n, fixed constant C, and $0 < \xi < 1$.

A2. $sup_{n\in\mathbb{N}} \sum_{j=1}^{p_n} |\beta_{j,n}| < \infty$.

A3. $sup_{1\leq j\leq p_n, n\in\mathbb{N}} \|X^{(j)}\|_\infty < \infty$

A4. $\mathbb{E}|\epsilon|^s < \infty$, for some $s > 4/\xi$

A1 will hold for a fixed network size that is not too large in relation to sample size n. Assumption A2 will hold for any setting where the number of nodes in the network is known and finite, if the number of j such that $\beta_{j,n} \neq 0$ is independent of n itself and is a finite number. We believe the networks to be relatively sparse and the effect sizes to be bounded so assumptions A3 and A4 are reasonable to assume.

With these assumptions the main theorem from *Bühlmann et al.* [10] for consistency is

Theorem III.1.

$$\mathbb{E}_X|\hat{F}_n^{(m_n)}(X) - f_n(X)|^2 = o_p(1)$$

as $n \to \infty$, *where* $\hat{F}_n^{(m_n)}(X)$ *is the boosting estimate for* $f_n(X)$, $\hat{F}^{(m)}(X) = \hat{F}^{(m-1)}(X) + \nu\hat{\beta}_{\hat{s}_m} x^{\hat{s}_m}$, $\hat{F}^{(1)} = \nu\hat{g}(X)$, *and* $\hat{s} = argmin_{1\leq j\leq p} \sum_{i=1}^n (U_i - \hat{\beta}_j X_i^{(j)})^2$, m *denotes the step, and* ν *denotes the step size.*

We can apply this theorem to our setting, showing that the group level boosting algorithm can accurately estimate the functional form of f_n in the case that the group selection model is not misspecified, ie $\lambda_g = m_g\beta_{jk}$ for any $(j,k) \in g$.

3.3 Simulation Studies

Our simulation procedure involves: simulating n random graphs as the adjacency matrix for each subject, defining true β with signal among dense groups and sparse groups, and obtaining true $y = \beta X + \epsilon$. We will repeat this 100 times for each simulation setting considered to evaluate the average performance under those conditions. The performance will be evaluated with selection criteria (sensitivity, specificity, and false discovery rate) and estimation criteria (mean squared error). The false discovery rate (FDR) is defined as the ratio of false positives to the total number of selected edges.

Existing methods considered in simulation include: lasso, elastic net, and group lasso. All the existing methods were compared using the following R packages: glmnet [24] and gglasso [86]. Cross validation was used for each of the existing methods to select the tuning parameters.

First, we consider three relatively small simulation settings that demonstrate how the method will perform in three general settings: (1) small sample size compared to network size with a homogeneous effect within groups, (2) the same network size and assumptions as the first setting but with a larger sample size, and (3) same network size as previous two settings and with the larger sample size of the second setting but without assuming a homogeneous effect within groups. In the first two settings if a group has non-zero signal, then the effect size is the same for all edges in that group; the last setting allows some heterogeneity of the effect within groups, meaning about 80% of the edges within a group have the same effect on the outcome. All three of these settings, presented in Table 3.1, are for a network with 20 nodes, resulting

in 190 edges for selection. The number of groups with non-zero signal differs across these three settings, in the first setting three of six groups have non-zero signal. In the second setting three of fifteen groups have non-zero signal, and finally in the third setting 80% of edges in five of fifteen groups have non-zero signal. The results summarized from 100 iterations of simulated datasets are presented in Table 3.1.

Simulation Setting			Method	Sensitivity	Specificity	FDR	MSE
n	q	Homogeneous					
			Group Boosting R	0.42 (0.04)	0.99 (0.03)	0.04 (0.11)	6.20 (0.64)
			Group Boosting AV	0.31 (0.16)	1.00 (0.01)	0.01 (0.04)	27.49 (8.27)
100	20	Yes	Lasso	0.10 (0.07)	0.99 (0.04)	0.09 (0.16)	26.89 (7.62)
			Elastic Net	0.32 (0.06)	0.98 (0.04)	0.07 (0.12)	8.66 (0.66)
			Group Lasso	0.07 (0.03)	1.00 (0.01)	0.06 (0.12)	26.82 (7.55)
			Group Boosting R	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.09 (0.02)
			Group Boosting AV	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.04 (0.01)
500	20	Yes	Lasso	1.00 (0.00)	0.93 (0.04)	0.38 (0.13)	0.05 (0.01)
			Elastic net	1.00 (0.00)	0.89 (0.05)	0.49 (0.10)	0.06 (0.02)
			Group lasso	0.52 (0.04)	0.89 (0.03)	0.66 (0.06)	2.10 (0.05)
			Group Boosting R	0.99 (0.01)	1.00 (0.00)	0.02 (0.04)	0.19 (0.05)
			Group Boosting AV	1.00 (0.00)	0.98 (0.01)	0.13 (0.05)	0.07 (0.02)
500	20	No	Lasso	1.00 (0.00)	0.92 (0.04)	0.39 (0.11)	0.14 (0.03)
			Elastic net	1.00 (0.00)	0.88 (0.05)	0.48 (0.11)	0.15 (0.03)
			Group lasso	0.62 (0.08)	0.89 (0.03)	0.60 (0.08)	1.87 (0.32)

Table 3.1: Results from 100 simulated datasets. This table presents the average (sd) in each column for selection sensitivity, specificity, and false discovery rate (FDR) defined as proportion of false positives over the number of false and true positive edges selected, and mean squared error of the effect estimate (MSE). 'Group Boosting R' denotes the method that regresses out the adjustment variables using the residuals and 'Group Boosting AV' allows the adjustment variables to be selected in the boosting algorithm. 'n' denotes the number of subjects and 'q' denotes the number of nodes in the network.

The first simulation setting demonstrates that our proposed method has greater sensitivity while still maintaining good specificity and well controlled FDR. This case with $n < p$ is challenging for many traditional methods, reflected in the results

from lasso. The two group boosting methods differ only in the way they treat the adjustment variables. The first row shows the results, when covariates are regressed out before implementing the two-stage boosting algorithm. 'Group boosting AV' instead treats each adjustment variable as its own group in the first stage and allows them to each be selected individually and included in the model. In this case, the approach regressing out the effect performs better; this is not surprising given the simulated adjustment variables are correlated with some edges in the network.

Figure 3.3 shows how the estimated β_{jk} compare to the simulated values. Panels (A) and (C) show the setting where we assume a homogeneous effect within groups, like in the first two settings presented in Table 3.1. Panels (B) and (D) compare the simulated and estimated values in a setting where only 80% of edges within a group have an effect on the outcome, such as in the third setting in Table 3.1.

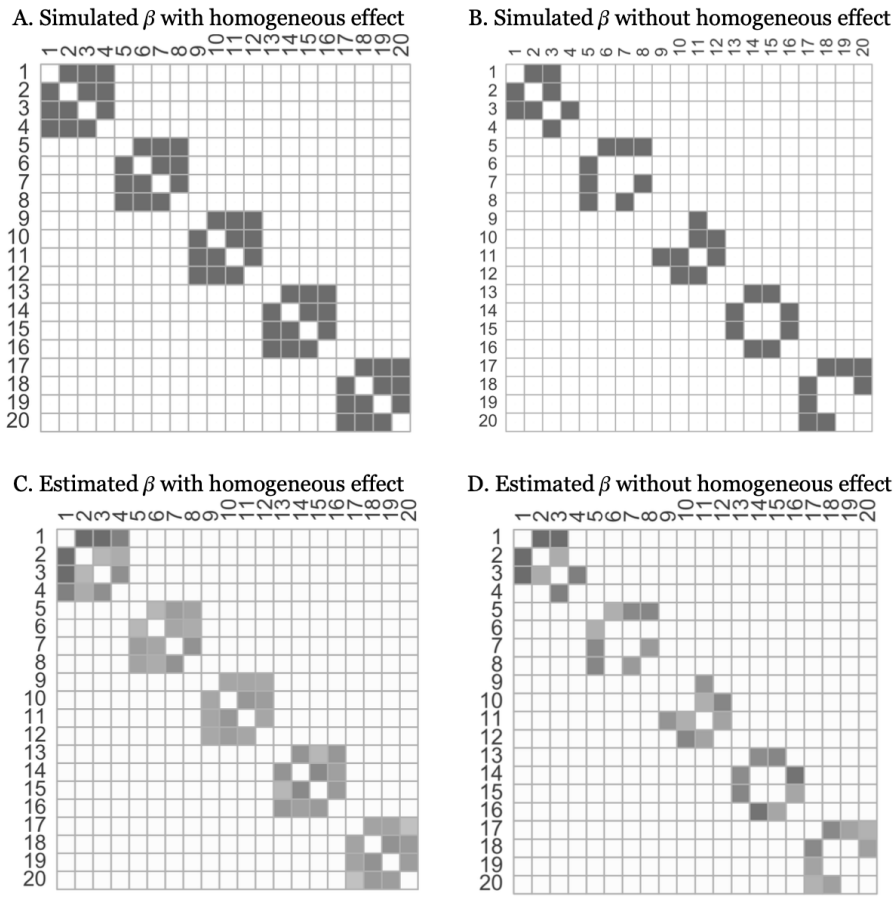


Figure 3.3: Simulated and estimated β values for two settings: assuming a homogeneous effect within groups and allowing some heterogeneity of the effect within groups. The shading represents the effect size.

The number of adjustment variables, and the correlation of these variables with the image, also varies some across these settings. In the first setting ($n = 100$) there are 5 adjustment variables in addition to the image, and they are correlated with some edges in the brain network, with an average correlation of 0.45. In the two other settings ($n = 500$) there are only three adjustment variables with an average correlation of 0.10 with the image.

All three of these settings reveal that the group boosting method can perform well in terms of sensitivity and specificity, while also controlling FDR. Though some of the existing methods are able to achieve high sensitivity and high specificity, FDR is also quite high among these methods. Interestingly, lasso and elastic net are outperforming

group lasso in these scenarios.

In order to understand the expected performance of our results in the real data application, we will also simulate a setting that has a similar ratio of network size to sample size. This setting has 1,000 subjects, 150 nodes with 15 regions. This results in a total of 11,180 edges belonging to 120 groups. We will assume homogeneity among the effect size in each group, similar to the first simulation setting. The signal to noise ratio is 0.08, and there are 5 adjustment variables with a true effect on the outcome. Results from this setting are shown in Table 3.2.

Simulation Setting			Method	Sensitivity	Specificity	FDR	MSE
n	q	Homogeneous					
1000	100	Yes	**Group Boosting R**	0.78 (0.07)	0.99 (0.01)	0.46 (0.15)	0.55 (0.08)
			Lasso	0.17 (0.08)	0.99 (0.00)	0.51 (0.18)	0.49 (0.02)
			Elastic net	0.16 (0.09)	0.99 (0.00)	0.46 (0.23)	0.49 (0.02)
			Group lasso	0.01 (0.01)	1.00 (0.00)	0.70 (0.40)	0.52 (0.01)
1810	264	No	**Group Boosting R**	0.56 (0.10)	1.00 (0.00)	0.10 (0.20)	< 0.01 (0.00)
			Lasso	0.38 (0.05)	1.00 (0.00)	0.74 (0.07)	< 0.01 (0.00)
			Elastic net	0.42 (0.04)	1.00 (0.00)	0.77 (0.05)	< 0.01 (0.00)
			Group lasso	< 0.01 (0.01)	1.00 (0.00)	1.00 (0.00)	< 0.01 (0.00)

Table 3.2: Results from 100 simulated datasets. This table presents the average (sd) in each column for selection sensitivity, specificity, and false discovery rate (FDR) defined as proportion of false positives over the number of false and true positive edges selected, and mean squared error of the effect estimate (MSE). 'Group Boosting R' denotes the method that regresses out the adjustment variables using the residuals. 'n' denotes the number of subjects and 'q' denotes the number of nodes in the network.

This simulation setting shows that it is difficult to achieve high sensitivity and specificity while still controlling FDR when the dimension of the network is large in relation to the sample size. The group boosting approach still outperforms the other methods by having higher sensitivity and similar specificity, FDR, and MSE. The run time for Group Boosting is longer compared to the existing methods, but is still

feasible for obtaining results and runs in roughly 10 minutes in the setting with 1,000 subjects and 100 nodes.

The final simulation setting was designed to resemble the ABCD data used for the real data application. The observed precision matrices are used to simulate the outcome, with a similar signal to noise ratio as observed in the ABCD data with general intelligence factor as the response. The network has 264 nodes, and the sample size is 1,810. This setting again demonstrates that the group boosting method outperforms existing methods in terms of sensitivity and FDR (see Table 3.2). In this case the FDR is well controlled and the sensitivity is higher than the other methods, so we are quite confident in the associations identified through the selected edges. We can also observe that it takes approximately an hour to run the proposed method on data the same size as our real data application, slower than the existing methods but still computationally feasible.

3.4 Real Data Application

Data from the first release of the ABCD study was used to evaluate the performance of this method on real world data. This study is a longitudinal study designed to follow adolescents for ten years to understand how substance use impacts brain development. The first release contains cross sectional data of the baseline characteristics of the children, all aged nine to ten, as well as rfMRI imaging. The outcome of interest is general cognitive ability.

This data contains 1,810 subjects with some descriptive information such as study site, age, race/ethnicity, parents' education level, parents' marital status, family income, and general psychopathology score. The brain network has been summarized into 264 regions, which we will use as the network nodes. *Power et al.* [59] categorized those 264 regions into 13 functional modules. We will use those as the grouping structure for the network edges (see Table 3.3).

#	Function	#	Function
1	Sensory/somatomotor Hand	8	Fronto-parietal Task Control
2	Sensory/somatomotor Mouth	9	Salience
3	Cingulo-opercular Task Control	10	Subcortical
4	Auditory	11	Ventral attention
5	Default Mode	12	Dorsal attention
6	Memory Retrieval	13	Cerebellar
7	Visual	−1	Uncertain

Table 3.3: Functional brain networks and associated brain functions as identified using resting state and task oriented fMRI by *Power et al.* [59].

General cognitive functioning was measured from the neurocognitive battery administered as part of the study, which consisted of seven measures based on the NIH toolbox and five additional tasks. Bayesian probabilistic principal components analysis (BPPCA) was used to summarize these measures into three broad categories of cognition: general ability, executive function, and learning and memory [77]. For this analysis we will use the general cognitive ability component as the outcome.

Since many factors, separate from brain imaging, are expected to impact general intelligence we will include the following adjustment variables: age, gender, race/ethnicity (white, Asian, Black, Hispanic, or other), parents' education level (high school diploma or GED, some college, Bachelor's degree, or post-graduate degree), family income (less than 50K, between 50K and 100K, or more than 100K), and parents' marital status. We will use the approach of regressing out the effect of these variables before then using the boosting algorithm to select edges associated with the general cognitive ability.

In order to define groups for the stage 1 selection of associated edges, we observed

that the assumption of homogeneity may be too strong when strictly using the thirteen previously identified functional modules. To overcome this challenge, we combined these groups with grouping edges based on the coefficient estimate from Lasso. There were 376 edges selected by Lasso when using the entire dataset, and these edges were grouped into 57 groups, enforcing homogeneity in the estimated effect of these groups. Combining with the 105 groups defined by the Power functional modules, we have a total of 162 groups. To assess how sensitive this method is to the choice of group partition of nodes in the network, we considered other group partition strategies as well. For these we used the coordinates to spatially group the nodes within each functional module, testing a range of the number of spatial clusters in each functional module.

The proposed method will be compared to using Lasso directly to perform selection and to fitting a linear regression including only the confounding variables. Some of the confounding variables, such as the psychopathology factor, are expected to highly influence cognitive ability so considering these variables only provides a benchmark for how much additional information the brain image provides.

The ABCD data will be split 100 times into 80% training and 20% testing sets in order to compare the performance of these methods on average and to evaluate reproducibility of selected edges. We will report the average predicted R^2 and predicted MSE (and standard deviation), defined as the difference between the predicted cognition score and true cognition score among the testing data, across these iterations. Additionally we will assess the stability of the selected subset of edges over different splits, to see whether it is sensitive to the subset of patients included in the training set.

3.4.1 Results

Using this method, we identified 20 edges that are associated with general cognitive ability. These edges span 10 functional modules from 15 of the 162 groups. The most frequently identified functional modules are the default mode (involved in 8 edges), visual (5 edges), cingulo-opercular task control network (5 edges), and dorsal attention (4 edges). Figure 3.4 presents the selected edges spanning these four networks, the default mode network seems especially important. Interestingly, the work done by [74] also found that interactions between the default mode and task controls networks were implicated in cognitive ability, similar to our findings.

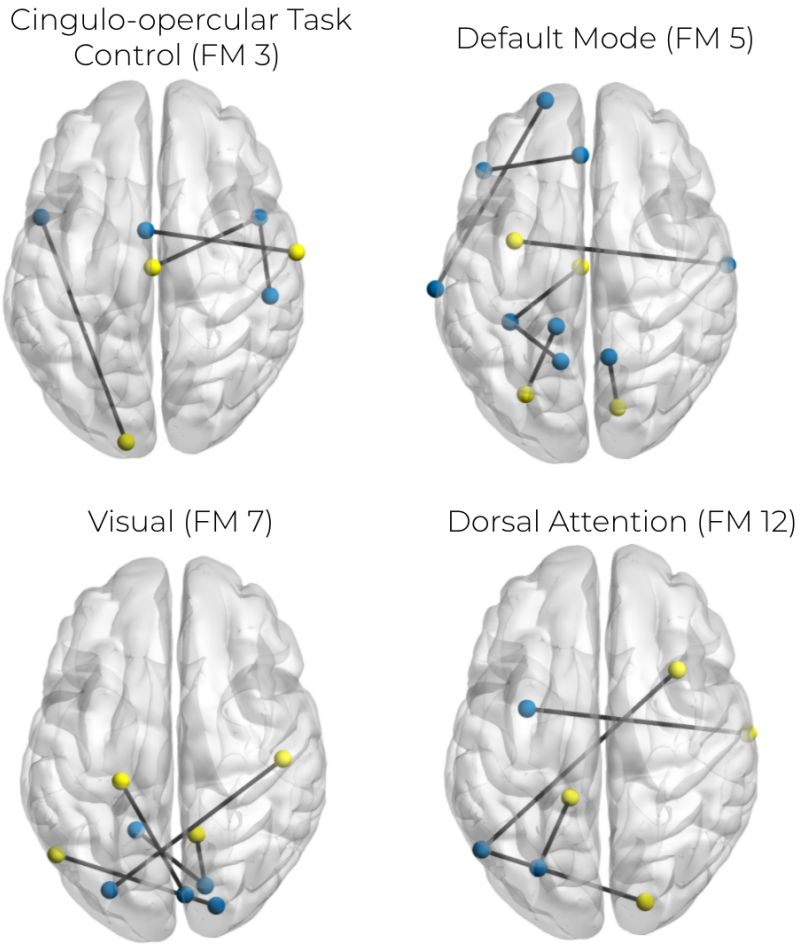


Figure 3.4: Nodes and edges identified as associated with cognition, split by most frequently identified functional modules.

When evaluating the 100 splits into testing and training data, we found that the group boosting algorithm was much more stable than lasso. In boosting, over 90% (31 out of 34) of the edges selected across all 100 iterations were selected in more than one split of the data; in Lasso, only 63% (997 of 1590) of the total edges identified were selected more than once, see Figure 3.5. Boosting identified 6 edges selected across all 100 splits and 18 selected in at least half of the splits. These edges span the following functional modules: sensory/somatomotor hand (1), sensory/somatomotor Mouth (2), cingulo-opercular task control (3), default mode (5), visual (7), salience (9), subcortical (10), and dorsal attention (12). The functional modules that have the most edges selected spanning them include: cingulo-opercular task control (3), default mode (5), visual (7), and dorsal attention (12). All 18 of these edges are present in the 20 edges identified using the entire dataset as described previously.

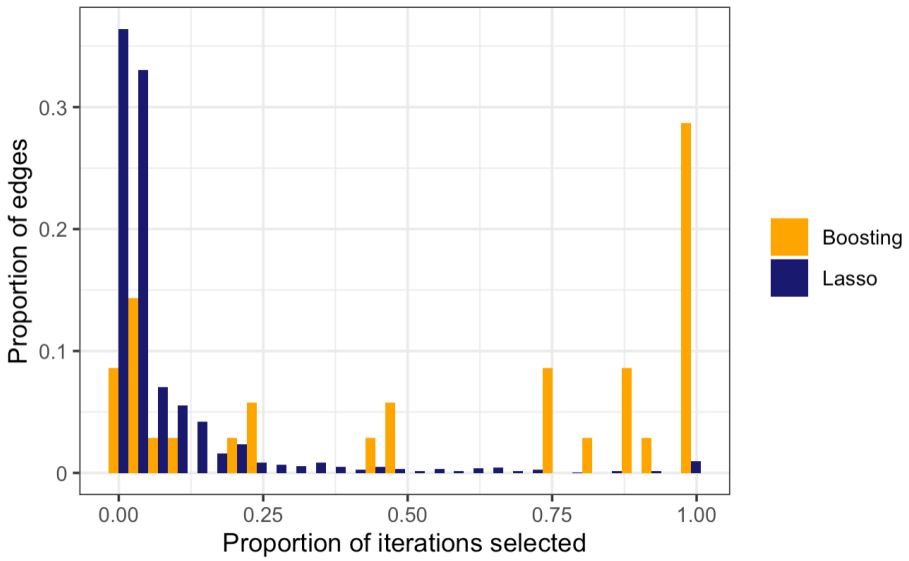


Figure 3.5: Stability of the edge selection for lasso and boosting. Histogram of the proportion of iterations each edge was selected scaled by the total number of edges selected by the corresponding method.

Other metrics we compared were the predicted R^2 and MSE. Group boosting performed similarly to lasso, with the average R^2 for group boosting $= 0.302$ (sd $= 0.033$) and average R^2 for lasso $= 0.293$ (sd $= 0.032$). Of the 100 splits, group boosting had

a better R^2 in 75 of the iterations. *Sripada et al.* [74] found an $R^2 = 0.10$ in their analysis predicting cognition using a dimension reduction approach. This suggests that considering the network directly, rather than components from a dimension reduction approach, improves performance. MSE of the coefficient estimates results were similar with group boosting slightly outperforming lasso, the average MSE for group boosting = 0.504 (sd = 0.026) and the lasso average MSE = 0.508 (sd = 0.026). Though the predictive performance is similar, the stability of selected edges gives us more confidence in the association of the selected edges identified, compared to that of lasso. Not only do we want to see good prediction performance, we also aim to understand the functional networks associated with cognitive functioning.

We may consider the proportion of variation explained by each functional module to understand how much each of these networks contributes. Table 3.4 presents these results, showing that the edges contained in or spanning the default mode network seems to contribute the most in terms of variation explained.

Functional Module	Number of Edges/Variables	% Variation Explained	AAL Regions
Default Mode	7	8.12	Superior frontal gyrus, Middle frontal gyrus, Inferior frontal gyrus, Superior frontal gyrus, Lingual gyrus, Superior occipital gyrus, Fusiform gyrus, Precuneus, Middle temporal gyrus
Visual	5	3.80	Precental gyrus, Calcarine fissure and surrounding cortex, Lingual gyrus, Superior occipital gyrus, Middle occipital gyrus, Postcentral gyrus, Precuneus, Middle temporal gyrus
Cingulo-opercular task control	4	3.75	Supplementary motor area, Insula, Median cingulate and paracingulate gyri, Postcentral gyrus, Supramarginal gyrus, Temporal pole: superior temporal gyrus
Dorsal Attention	4	2.77	Middle frontal gyrus, Insula, Lingual gyrus, Superior occipital gyrus, Middle temporal gyrus
Adjusted Variables	15	14.70	

Table 3.4: Summary of contribution of edges from the most frequently identified functional modules, in terms of proportion of variation explained and associated AAL116 regions.

We compared the results of two additional group partitions, determined using the spatial coordinates of the nodes within each functional module. The fitted R^2 was similar to that of the primary analysis. The specific edges identified were different, although many of the functional modules they belonged to were similar. For example, when we split the functional modules into three clusters each, we identified several edges spanning the default mode, visual, and dorsal attention networks. The edges identified using the full data, regardless of the group partition we chose, seemed to share many of the same functions.

3.5 Discussion

This work introduces a two-stage boosting algorithm to perform scalar-on-network regression, utilizing structure of subnetworks. The utility of this method has been

demonstrated through simulations, with higher sensitivity and better controlled false discovery rate compared to commonly used existing regularization and group variable selection methods. rfMRI data from the ABCD study was used to identify associations between cognition and ROIs in the default mode, cingulo-opercular task control, visual, and dorsal attention networks. The analysis reveals stable and reproducible associations among edges spanning these functional modules.

Future work will involve relaxing the homogeneity assumption. Assuming the effect size is very homogeneous within groups may be too strong for some real data brain imaging applications, in that case relaxing this assumption may improve the ability to detect true associations. Other extensions to this method could be considered as well, such as implementing a tree-based boosting method rather than the L_2 boosting algorithm implemented in this work.

CHAPTER IV

Scalar-on-network Regression via Deep Neural Networks

4.1 Introduction

Functional Magnetic Resonance Imaging (fMRI) has increasingly been used to study how patterns in brain activation relate to psychiatric diseases and disease progression. The complex nature of brain functioning necessitates the development of new methods to understand these associations. As one of the most popular imaging technologies, fMRI measures the blood oxygen level dependent (BOLD) signal at hundreds of thousands of voxels in the brain over a period of time; fMRI may capture the brain activity at rest, also known as resting-state fMRI (rfMRI), or while a specific task is being performed, known as task-oriented fMRI. Both of these imaging modalities are important for assessing brain functions and specifically for psychiatric disease progression.

The statistical approaches for analyzing the task fMRI data and the rfMRI data are quite different due to the nature of imaging acquisition and the goal of analysis. The task fMRI focuses on identifying the brain activation regions that are related to certain brain functions. The task fMRI time series across individuals can be aligned in time based on the task stimuli. From a typical statistical analysis for the task

fMRI data, e.g., statistical parametric mapping (SPM), the individual-level contrast maps or t-maps can be constructed for examining differences in brain activity under different task conditions. The rfMRI aims to study brain activity at rest, i.e., when a explicit task is not being performed. Due to this design, the rfMRI time series across individuals are not directly comparable, while it has been of interest to study the regional interactions of brain activity for single individuals or a group of individuals. In particular, the rfMRI data can be summarized as the functional connectivity measures or functional brain networks between regions, e.g., the correlation matrix of averaged BOLD time series at region level.

In this chapter, we perform a novel statistical association analysis between the scalar outcome variable (e.g., neurocognitive functions) and functional connectivity brain networks (e.g., correlation matrices) from rfMRI data. We aim to perform the network feature selection, i.e., to identify the important region pairs in which the functional connectivity measure is strongly associated with the outcome variable, while integrating multiple sources of information including well-known brain functional networks and task fMRI brain activation patterns. We are interested in exploring how much the task fMRI brain activation patterns may facilitate to improve the network feature selection.

4.1.1 Existing Methods

Previous research has taken different approaches to perform association analysis between scalar outcome and brain networks. One such study, used a dimension reduction method to understand the association between general cognitive ability and 75 components of the brain network, identified via brain basis set modeling [74]. This research provides a good foundation for understanding broadly which functional networks of the brain are associated with cognition, yet the interpretation of the components is limited.

A more flexible modeling strategy that can take into account all the information from the brain network, rather than just the more important components, may be able to improve upon these findings. The deep neural networks (DNN) provide great flexibility for modeling the edge-dependent regression coefficients in the scalar-on-network regression integrating multiple sources of information into the model. Deep learning methods have been shown to have a very good predictive performance in many different applications and the benefit of a flexible framework [65, 67].

4.1.2 Motivating Data: ABCD Study

The ABCD study aims to provide more insight into brain development in adolescents, and specifically to show the ways in which it can be a marker of neurocognitive functioning. This study has collected rfMRI and task-oriented fMRI data for about 11,000 children aged nine or ten in the first release of data, as well as demographic information and behavior tasks aiming to understand higher-order cognitive functions [50].

Previous work [78] shows that DNN have the potential to identify brain regions that are highly associated with cognitive functioning with greater precision than existing methods. Extending this work to incorporate multiple imaging modalities provides the framework to combine multiple sources of information that are believed to be important in the underlying associations between brain functioning and neurocognitive disease. The flexible nature of DNNs allows the incorporation of additional node and edge information as inputs in the neural nets, such as summary information from task-oriented imaging. In this work we will propose an algorithm to perform scalar on network regression with DNN, extending the previous work, while also incorporating multiple imaging modalities. By using DNN, we hope to relax the assumption of homogeneity within groups that was integral to ensure good performance in the second chapter.

This application could particularly benefit from incorporating task fMRI, because of the known associations between differences in performance of certain tasks and addiction [35]. For example, one study of adolescents detected differences in activation patterns among teens performing an inhibition task predictive of substance use in the next 18 months [51]. Another study found patterns of deactivation in the default mode network when performing verbal Stroop tests for subjects with internet addiction compared to other subjects [18]. The monetary incentive delay task has also been established to be effective in demonstrating differences in brain activity among addicted or at-risk subjects [4].

In this analysis, we aim to identify connectivity patterns that are associated with the Child Behavior Checklist (CBCL) score among children in the ABCD study. To do this, we will utilize both task-oriented fMRI and rfMRI, summarized at the cortical surface level. Unlike the previous two chapters, which utilized volumetric data. Surface-based metrics have some advantages over volumetric summaries, such as better accounting for differences in individuals' cortical structure [2, 8].

Figure 4.1 presents the process of incorporating multiple sources of data into the scalar-on-network regression model. Panel A shows the location of nodes in the cortical surface rfMRI data. The nodes are colored by membership in 13 functional networks, like the default mode or cingulo-opercular networks. Panel B shows how the time series corresponding to each node are used to establish the connectivity patterns (Panel C) for each subject. Panels A-C are the processing steps for the rfMRI data. Panel D shows the images from task MRI for each subject, and the activation levels for each location on the cortex are averaged across subjects. These activation patterns are included in the data used as an input in the neural networks (Panels E and F). Other node or edge-level characteristics like the coordinates of the parcels and an indicator of membership in the 13 functional networks are also used as inputs for the neural networks (Panel E). Panel F presents the structure of the

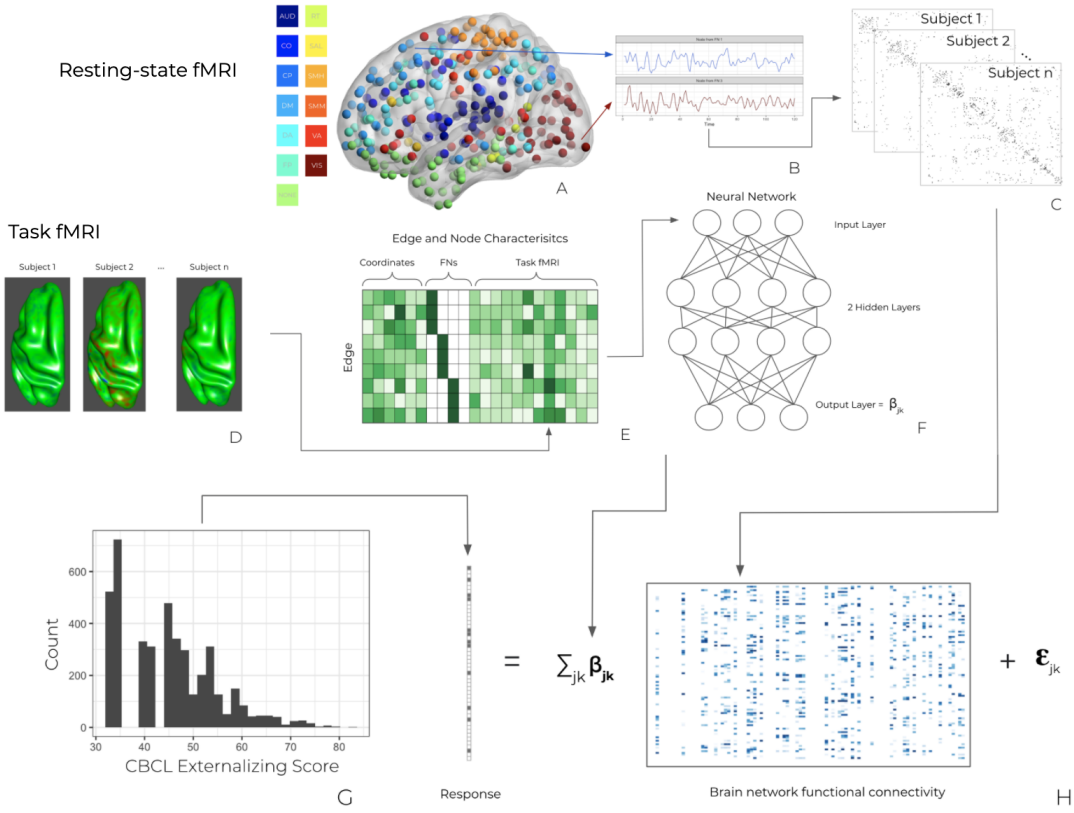


Figure 4.1: Flowchart depicting how the three sources of data, rfMRI, task fMRI, and patient characteristics, relate to estimate the final equation (panel H).

neural networks: inputs are those in Panel E, two hidden layers, and the output are the coefficients for the eventual scalar-on-network regression. The outcome of the data application in this work is the Child Behavior Checklist Externalizing score, and Panel G shows a histogram of this outcome across subjects. Finally Panel H shows how these components relate in the scalar-on-network regression model.

4.2 Methods

The proposed method provides a framework for using deep neural networks to model functional connectivity patterns using brain network characteristics, and cortical surface coordinates. This framework, though developed for the analysis of functional brain connectivity, can be generalized to networks of other types as well. In the cortical surface context, nodes in the network are parcels along the cortical surface defined by the Gordon spatial parcellation [27] and edges are connectivity patterns between any two parcels.

Suppose the data consists of n subjects and let $i(i = 1, \ldots, n)$ denote the subject index. We are interested in relating functional connectivity measures of q brain regions, denoted as $\{x_{ijk} : (j,k) \in S\}$, to a scalar outcome, denoted as y_i, where $S = \{(j,k) : 1 \leq k < j \leq q\}$ is the whole edge set of interest in the functional connectivity network.

Consider the scalar-on-network regression model

$$y_i = \beta_0 + \sum_{l=1}^{p_A} \alpha_l z_{il} + \sum_{(j,k)\in S} \beta_{jk} x_{ijk} + \epsilon_i \tag{4.1}$$

where $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. We consider regression coefficients β_{jk} to be edge-dependent which is specific to a node pair (j,k).

In some network applications, the entire network can be partitioned into several subnetworks or groups, i.e., $S = \bigcup_{r=1}^{R} S_r$ and $S_r \cap S_{r'} = \emptyset$. For example, the Gordon

spatial parcellation used with cortical surface data, categorizes each parcel into one of thirteen functional networks. This partition may be leveraged in the modeling stage if it is assumed that there are similar effects within groups, or it may be used to aid in interpretation of results. In this analysis, to relax the assumption made in the previous chapter of a similar effect within groups, we don't directly use the partition when estimating the neural network.

In many cases the linear relationship between y_i and x_{ijk} forced in (4.1) is too rigid. Model performance may be improved if we relax the assumption of linearity and model with a more flexible nonlinear model. Consider estimating β_{jk} as follows

$$\beta_{jk} = g(u_{jk}; \boldsymbol{\theta}); \quad u_{jk} = \{v_j, v_k, w_{jk}\}. \tag{4.2}$$

The function $g(\cdot)$ is modeled as a multi-layer feed forward neural network, where $u_{jk} = \{v_j, v_k, w_{jk}\}$ and $\boldsymbol{\theta}$ represents the weight parameters. The input of the neural network contains the coordinates of the parcels as well as other characteristics specific to the locations j and k, denoted by vectors $\{v_j, v_k\}$. Other inputs can be edge specific characteristics, such as an indicator of membership to functional regions. Additionally task-oriented fMRI characteristics can be summarized at the node or edge-level and incorporated as inputs in the neural network for each β_{jk}. Node-level task fMRI information is summarized as the average activation across subjects at that parcel, and edge-level information is summarized as the average full or partial correlation between two parcels across subjects. Including task fMRI statistics as inputs in the neural network will improve estimation if the task images impact the association between clinical outcome and the resting-state images, essentially an interaction.

4.2.1 Estimation

To estimate parameters in (4.1), we consider the following constrained optimization problem with the L_1 penalty to induce the sparsity in regression coefficients.

$$Q(\boldsymbol{\Theta}) = \frac{1}{2}\|\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \quad \text{subject to} \quad \boldsymbol{\beta} = \mathbf{g}(u_{jk}; \boldsymbol{\theta}), \tag{4.3}$$

where λ is a tuning parameter and $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$. In order to solve the objective function we use the ADMM algorithm which introduces η constrained by $\eta = \beta$, and then solves the objective function in (4.4).

$$\begin{aligned} Q(\boldsymbol{\Theta}) =& \frac{1}{2}\|\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\eta}\|_1 + \boldsymbol{\tau}^\top(\boldsymbol{\beta} - \boldsymbol{\eta}) \\ &+ \frac{\rho_1}{2}\|\boldsymbol{\eta} - \boldsymbol{\beta}\|_2^2 + \boldsymbol{\zeta}^\top(\boldsymbol{\beta} - \mathbf{g}(u_{jk}; \boldsymbol{\theta})) + \frac{\rho_2}{2}\|\boldsymbol{\beta} - \mathbf{g}(u_{jk}; \boldsymbol{\theta})\|_2^2, \end{aligned} \tag{4.4}$$

where λ, ρ_1, and ρ_2 are tuning parameters and $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta})$. The neural network used to estimation $g(u_{jk}; \theta)$ is a feed forward network with two hidden layers, using the ReLU activation function. The first hidden layer has 50 nodes and the second hidden layer has 25 nodes.

Similar to the solution using ADMM in Chapter II and ADMM for lasso [5], we can obtain the following solutions for the parameters at iteration t for $t = 1, 2, \ldots,$:

$$\begin{aligned} \beta^{(t)} &= \left\{\mathbf{X}^\top\mathbf{X} + (\rho_1 + \rho_2)\mathbf{I}\right\}^{-1}\left\{\mathbf{X}^\top(\mathbf{y} - \mathbf{Z}\alpha^{(t-1)}) + \rho_1\boldsymbol{\eta}^{(t-1)} - \boldsymbol{\tau}^{(t-1)} + \right. \\ &\qquad \left. \rho_2\mathbf{g}(u_{jk}; \boldsymbol{\theta}^{(t-1)}) - \boldsymbol{\zeta}^{(t-1)}\right\}, \\ \eta^{(t)} &= S_{\lambda/\rho_1}\left\{\boldsymbol{\beta}^{(t-1)} + \frac{\boldsymbol{\tau}^{(t-1)}}{\rho_1}\right\}, \\ \zeta^{(t)} &= \zeta^{(t-1)} + \rho_2\{\beta^{(t-1)} - g(u_{jk}; \theta)\}, \\ \tau^{(t)} &= \tau^{(t-1)} + \rho_1\{\beta^{(t-1)} - \eta^{(t-1)}\}, \end{aligned}$$

where $S_\lambda(x) = (x - \lambda)I(x > \lambda) + (x + \lambda)I(x < -\lambda)$ for $\lambda > 0$ is the soft thresholding

function. The ADMM algorithm is more efficient if the starting values of the parameters are closer to the final estimates. In order to obtain a reasonable starting value, lasso estimates are used for $\hat{\alpha}_l, (l = 1, \ldots, p_A)$ and $\hat{\beta}_{jk}, (1 \leq k \leq j \leq q)$.

Data: $\{\mathbf{x}_i, y_i, \mathbf{z}_i\}_{i=1}^n$; Max number of iterations T; Tuning parameter λ, ρ_1, ρ_2.

Result: $\{\beta_{jk}\}$

begin

Initialize $\beta_{jk} = \hat{\beta}_{jk,\text{lasso}}$ and θ.

for $t = 1, \ldots, T$ **do**

Train the NN

$\text{Loss} = \frac{1}{2}\|\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\eta}\|_1 + \boldsymbol{\tau}^\top(\boldsymbol{\beta} - \boldsymbol{\eta})$

$+\frac{\rho_1}{2}\|\boldsymbol{\eta} - \boldsymbol{\beta}\|_2^2 + \boldsymbol{\zeta}^\top(\boldsymbol{\beta} - \mathbf{g}(u_{jk}; \boldsymbol{\theta})) + \frac{\rho_2}{2}\|\boldsymbol{\beta} - \mathbf{g}(u_{jk}; \boldsymbol{\theta})\|_2^2$

Update $\boldsymbol{\alpha} = (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Update $\boldsymbol{\beta} =$

$\left(\mathbf{X}^\top\mathbf{X} + (\rho_1 + \rho_2)\mathbf{I}\right)^{-1}\left(\mathbf{X}^\top(\mathbf{y} - \mathbf{Z}\alpha) + \rho_1\boldsymbol{\eta} - \boldsymbol{\tau} + \rho_2\mathbf{g}(u_{jk}; \boldsymbol{\theta}) - \boldsymbol{\zeta}\right)$

Update $\eta = S_{\lambda/\rho_1}(\boldsymbol{\beta} + \frac{\boldsymbol{\tau}}{\rho_1})$

Update $\boldsymbol{\tau} = \boldsymbol{\tau} + \rho_1(\boldsymbol{\beta} - \boldsymbol{\eta})$

Update $\boldsymbol{\theta} = \text{argmin}_{\boldsymbol{\theta}}\|\boldsymbol{\beta} + \frac{\boldsymbol{\zeta}}{\rho_2} - \mathbf{g}(u_{jk}; \boldsymbol{\theta})\|_2^2$

Update $\boldsymbol{\zeta} = \boldsymbol{\zeta} + \rho_2(\boldsymbol{\beta} - \mathbf{g}(u_{jk}; \boldsymbol{\theta}))$

end

end

Both ρ_1 and ρ_2 may depend on the sample size. From our experiences, $\rho_1 = 1.5n$ and $\rho_2 = 0.01n$ are a good choice in practice. The feature selection is based on the choice of λ so this is chosen using the 5-fold cross validation. The proposed method is referred to as SoNR-NN (scalar on network regression - neural networks). The Adam algorithm [43], a stochastic gradient descent method, is used to solve for $\boldsymbol{\theta}$ in the algorithm. In this application, the neural network has two hidden layers and uses the ReLU activation function. The first hidden layer has 50 nodes and the second hidden

layer has 25 nodes.

4.3 Simulations

Simulations were used to evaluate the predictive performance of the proposed method. The simulations were designed according to the rfMRI data in the ABCD study. The outcome variable was simulated from the model $y_i = \sum_{l=0}^{p_A} \alpha_l z_{il} + \sum_{(j,k)\in S} \beta_{jk} x_{ijk} + \epsilon_i$ with variance of ϵ_i set to define a signal to noise ratio of about 0.60. The predictors, i.e. the functional connectivity networks of 418 regions for $\{x_{ijk}\}$ and 16 additional individual characteristics for $\{z_{il}\}$, were extracted from the ABCD data, see details in Section 4.4. The true regression coefficients were specified in light of the analysis of rfMRI data in the ABCD study. In particular, the lasso estimate was used as the true value of α_l. For the true value of β_{jk}, the largest 1,000 estimated effect size were adopted as the non-zero effects and all other effects were set to be zero. The sample size is 4,322, the same number of subjects analyzed in the ABCD study.

The proposed method is compared to Lasso, and evaluated in terms of variable selection accuracy as well as predictive performance. The predicted R^2 is obtained by simulating a testing set used to test 100 iterations of different training sets. Selection accuracy is evaluated using sensitivity, specificity, and false discovery rate (FDR), which is defined as the ratio of correctly selected variables to the total number of selected variables.

One simulation design is based on rfMRI volumetric data from the ABCD study and does not incorporate task fMRI data. This simulation demonstrates that the method can achieve much higher sensitivity compared to existing methods like Lasso, even without including multiple imaging modalities (see Table 4.1). Another important note is that FDR is well controlled, less than 10% for both methods.

The other simulation design does incorporate task fMRI from the ABCD data.

Table 4.1: Results from two simulations based on ABCD imaging. The evaluation metrics are presented as proportions and the table presents the average of 100 iterations.

	Method	Predicted R^2	Sensitivity	Specificity	FDR
Volumetric	SoNR-NN	0.82	0.24	0.83	0.04
	Lasso	0.76	0.02	0.99	0.07
Cortical Surface	SoNR-NN	0.23	0.41	0.97	0.12
	Lasso	0.22	0.12	0.99	0.19

This simulation setting gives insight into the results from the real data analysis, using the same data sources. We can see again that the proposed method has much higher sensitivity. FDR is still well controlled, with the proportion of falsely selected edges still less than 15% of all the selected edges.

4.4 Data Application

4.4.1 ABCD fMRI Data

Cortical surface data from rfMRI and task fMRI were obtained for a sample of 4,322 subjects from the ABCD study [12]. We limited the sample to subjects performed the same task in the first round of imaging, those who had the Monetary Incentive Delay (MID) task-oriented fMRI. This task was selected as one of three for the ABCD study because of previously observed differences in subjects suffering from addiction [4]. These images are summarized into several different contrast maps, here we focus on the contrast between small reward anticipation vs neutral anticipation.

Summarizing the fMRI data at the cortical surface rather than by volume may have some advantages, such as improved sensitivity and better alignment for individuals [8, 2]. The Gordon parcellation of about 32,000 parcels is used to determine the locations on the folded cortex [27]. Rather than 3D coordinates like in volumetric data, we use spherical coordinates to capture the locations on the folded cortex.

The functional labels for the Gordon parcellation are mostly consistent with those used with the Power parcellation of volumetric data, with a few exceptions [27]. Table 4.2 presents the functional label and corresponding abbreviation used in Figure 4.1.

Abrv.	Function	Abrv.	Function
DM	Default Mode	CP	Cingulo-parietal
SMH	Sensory/somatomotor Hand	CO	Cingulo-opercular Task Control
SMM	Sensory/somatomotor Mouth	RT	Retrosplenial Temporal
VIS	Visual	VA	Ventral attention
FP	Fronto-parietal Task Control	SAL	Salience
AUD	Auditory	DA	Dorsal attention
NONE	Uncertain		

Table 4.2: Functional brain networks and associated brain functions as identified using resting state and task oriented fMRI by *Gordon et al.* [27].

In the estimation of $\beta_{jk} = g(v_j, v_k, w_{jk}; \boldsymbol{\theta})$, the parcels defined by the Gordon 418 spatial parcellation represent nodes j and k. Then node level feature v_j may contain the spherical coordinates of parcel j and an indicator of membership in a functional network (see Table 4.2); v_j may also include average task fMRI information at location j across all subjects. In our application the node-level task fMRI features were summarized as the average contrast between small reward anticipation and neutral anticipation at each parcel across patients from the MID task. Edge characteristics included in w_{jk} are an indicator of whether the two nodes belong to the same functional network and the average correlation between the task contrast maps for those two nodes across subjects.

4.4.2 ABCD Clinical Data

In addition to the rfMRI images, we will adjust for five subject characteristics in the model ($\mathbf{Z}$). The variables include: sex, age (in months) race, parental education level, and parents' marital status. Several of these categorical variables are incorporated as indicators of having any of the categories so this results in the following 16 covariates: female or male, race is split into white, Black, Native American, Chinese, Japanese, or other race (note that subjects could choose more than one race category), age in months, parental education is split into less than a high school diploma, high school diploma or GED, attended some college, Associate's degree, Bachelor's degree, post-graduate degree, or unknown, and parents' marital status is simply married or not.

The outcome of interest is the CBCL externalizing score. The CBCL was developed in 1991 as a screening tool to identify behavioral and emotional problems in children or adolescents [1]. Children with severe affective and behavioral dysregulation can be identified using a profile from the CBCL, and these children are at greater risk of adverse outcomes like future substance use, suicide, and severe psychiatric symptoms [34]. Because one of the main long-term goals of the ABCD study is to evaluate brain development related to substance use, this score is an important metric to track, especially among the children when they are still young. At the time of this analysis, less than 2% of the subjects had reported consuming a full drink of alcohol or trying any substances, so a risk factor for future substance use such as the CBCL score is used instead.

4.4.3 Results

The SoNR-NN model fitting was compared to the performance of Lasso. The SoNR-NN model included three additional components of information from the task fMRI across subjects: (1) average value of the task contrast maps across subjects for

each node; (2) average correlation between two nodes of the task contrast map; and (3) the standard deviation of the task contrast maps for each node.

Method	Fitted R^2	Predicted R^2
SoNR-NN	0.105	0.095
Lasso	0.075	0.058

Table 4.3: Results from the ABCD analysis, comparing SoNR-NN to Lasso in terms of fitted and predicted R^2. Predicted R^2 is an average of 5 splits into 80% training and 20% testing splits.

From Table 4.3, we observe that this method can improve the fitted R^2 and predicted R^2 substantially compared to using Lasso. The model including all considered inputs to the neural network, including task fMRI, identified 172 edges that are associated with the CBCL externalizing score. The inputs with the largest impact are the coordinates of the nodes (see Table 4.4). We also note that including the standard deviation of task contrasts at each node and the correlation between task contrasts seem to contribute more information than simply the average task contrast at each node. This implies that the variation of task contrast across subjects and the correlation between nodes interacts more with the resting-state data impacting the CBCL score. Using other MID contrasts or other tasks, such as the stop signal task which is also collected by the ABCD study, may reveal a different association.

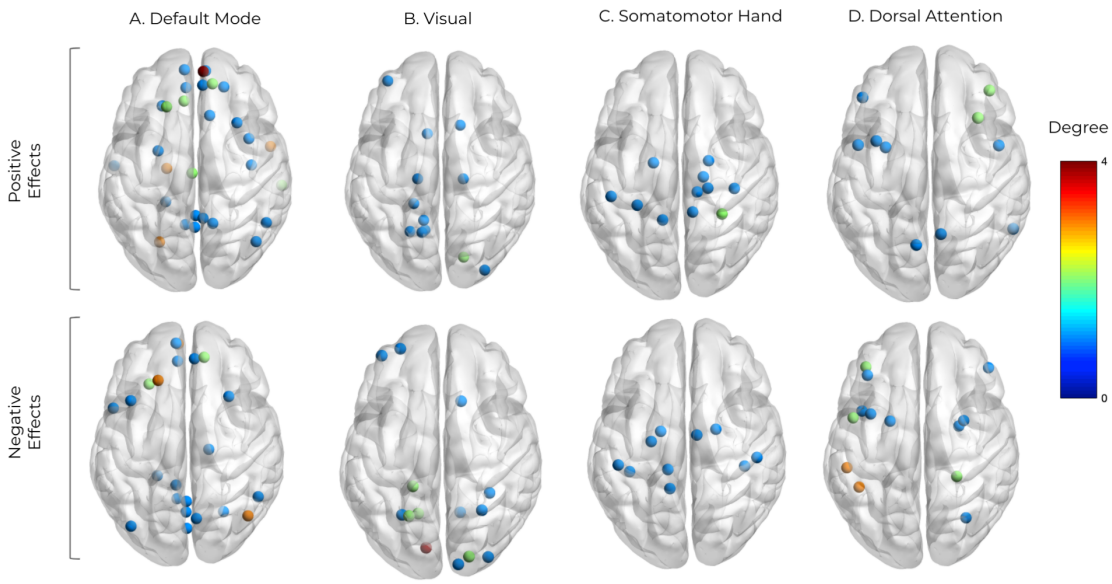


Figure 4.2: Parcels associated with the CBCL externalizing score, identified from the full data analysis that belong to the four most commonly associated functional networks. The nodes are colored according to the degree, ie the number of connections that node has with other nodes. The top row presents the nodes with a positive effect, ie higher connectivity is associated with higher CBCL score, and the bottom row presents the nodes with a negative effect, ie lower connectivity is associated with higher CBCL score.

Inputs	Proportion of Variation Explained (%)
Node level spatial coordinates	41.66
Node pair functional network indicator	6.65
Node level mean of task contrast map	2.60
Node pair correlation of task contrast map	17.16
Node level SD of task contrast map	13.61

Table 4.4: ANOVA of neural network inputs to evaluate importance of each component.

The functional networks most frequently associated with the CBCL externalizing score are: default mode (42% of selected edges are contained within or span this network), Cingulo-opercular (23%), visual (20%), and dorsal attention (19%).

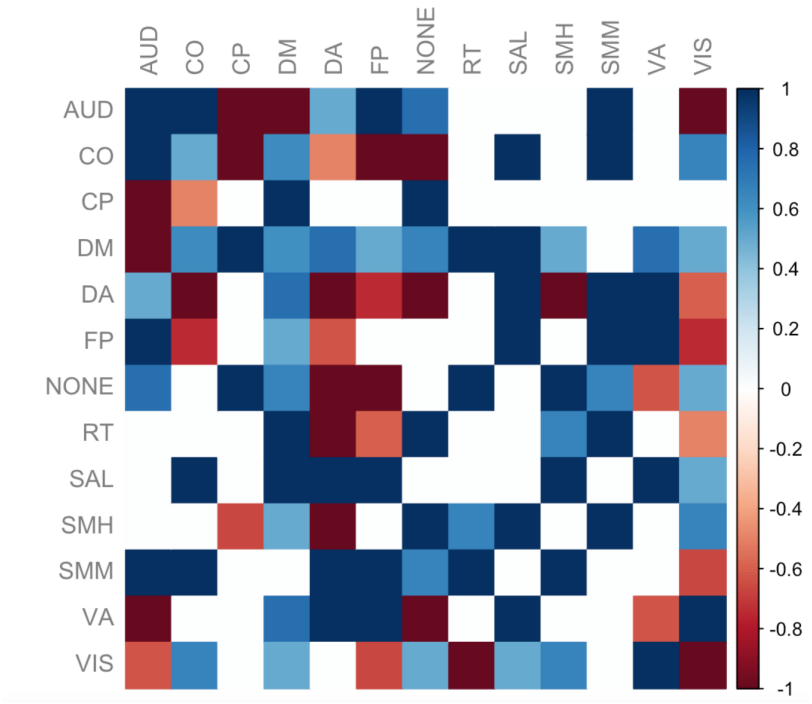


Figure 4.3: Heatmap showing the proportion of positive or negative edges detected within and across each of the functional networks. Color denotes the proportion of edges identified which have a positive (blue) or negative (red) association with the CBCL score.

Figure 4.3 shows on average which regions higher or lower connectivity is associated with an increased CBCL score. The blue regions are those in which a greater proportion of edges identified have a positive association with the behavioral score, meaning higher connectivity is associated with higher behavioral score. Conversely, red regions are those in which a larger proportion of edges identified in that functional network have a negative association with the CBCL score.

4.5 Discussion

This work proposes an approach to analyze resting-state fMRI data in conjunction with task fMRI, in order to identify regions of brain connectivity associated with a clinical outcome. Analysis of the ABCD data established association between the default mode, visual, Cingulo-opercular, dorsal attention, and somatomotor hand

networks and the child behavior checklist outcome, which is known to be associated with future substance use [34]. The proposed method is able to detect more associated edges in the brain network, and reveals that node level variation and node pair correlation of the task contrast maps have more impact on the association between the behavioral outcome and the resting-state functional connectivity measures.

Additional analyses of different tasks or different contrast maps could provide further insight into the benefits of combining resting-state and task fMRI. Additionally, applying this method to substance use directly to the ABCD data in a few years, rather than the CBCL externalizing score, could reveal a different association.

CHAPTER V

Discussion and Future Work

This work has introduced three new methods for analyzing functional brain connectivity and identifying patterns that are associated with clinical characteristics and outcomes. Each chapter explores a different setting where these connections may provide further insights into the development or progression of cognitive ability and psychiatric diseases. Developments in big data analysis and brain imaging have paved the way for new research in this area, and these methods demonstrate the potential ways that brain imaging can be used to influence future research and eventually clinical practice. These methods may also be generalized other network settings outside of neuroimaging. The first method may be used in any setting where one would like to relate some vector of predictors to connectivity patterns in a network. The latter two methods may be applied to settings in which a scalar outcome is regressed on a sparse network. Though they have designed with functional connectivity in mind, these methods are not limited to brain imaging applications.

On Predictability of Individual Functional Connectivity Networks from Clinical Characteristics (Chapter II) introduces a joint modeling approach to estimating individual connectivity networks and associations between clinical characteristics the connectivity patterns. This chapter demonstrates how machine learning algorithms can be used to improve power to detect novel associations amid often noisy imaging

data. Estimating individual connectivity networks is still an active area of research and joint estimation is one way to leverage information across subjects to improve this estimation. Importantly, we are also able to identify clinical characteristics associated with connectivity patterns, which is an important and difficult feature of imaging studies. With noisy imaging data, often the false discovery rate of associations is high without dimension reduction methods. The proposed method does not rely on dimension reduction to identify reliable, interpretable associations between specific regions of interest in the brain and clinical characteristics.

When this method was applied to analysis of the PNC data, we identified regions of the brain where connectivity patterns were highly predictable given the subset of clinical characteristics collected in that study. Associations identified between the subcortical network and SIPS survey questions are consistent with previous literature [36, 82, 16], in addition to some novel associations identified. This gives further confidence in the methods used, as reproducibility is both difficult and critical in brain imaging studies.

The next chapter III, *Scalar on Network Regression via Boosting*, demonstrates the utility of a boosting algorithm to identify associations between clinical outcomes and functional connectivity patterns. The proposed method provides higher sensitivity to detect associations between connectivity and clinical outcomes, while ensuring the false discovery rate is low. Identifying associations that are reproducible is one of the main strengths of this method. It also leverages the natural group structure in light of the well-known functional brain networks, to efficiently screen out any regions that are not associated with the outcome of interest.

Applying this method the the ABCD study, we identified specific edges within several functional networks that are associated with general cognitive ability. The most frequently identified functional networks are consistent with previous studies of cognitive ability, yet identifying specific edges in the network established by the

Power spatial parcellation provides a more granular level of specificity compared to most existing methods.

Finally, *Scalar on Network Regression via Deep Neural Networks* (Chapter IV) builds upon the ideas of the boosting method in the previous chapter. This method similarly detects associations between connectivity patterns and a clinical outcome. The main difference is that by using neural networks, we can incorporate additional node and edge-wise information beyond simple membership to a group or functional network. In this context, task fMRI was incorporated to determine if there is an interaction between resting-state and task fMRI that could be leveraged to improve the prediction of clinical outcomes given connectivity patterns.

This method when applied to the ABCD study identifies 172 edges that are predictive of the CBCL score, an important risk factor for future substance use. This method was also able to show that including task fMRI information is possible, though in this setting not necessary to improve upon performance of existing methods. Using coordinates of parcels in the image alone, greatly improved the power to detect associations compared to Lasso.

In summary, this work presents three new methods for analyzing brain functional connectivity, and demonstrates their utility through simulations and with applications to the PNC and ABCD studies. Several future directions can be pursued. First, as one important consideration with brain imaging data analysis, image pre-processing procedure may impact the performance of methods. For example, registering the raw imaging data into the volumetric space or on the cortical surface may affect the ability to detect brain signals. Further research is needed to study how the proposed methods are sensitive to the different image pre-processing procedures. Second, incorporating more information and knowledge into the proposed methods may potentially obtain a more scientifically meaningful result. For example, in Chapter III using a more informative grouping of nodes in the brain network may improve the power of the

method and identify more region pairs that are strongly associated with the clinical outcome of interests, since it relies partially on the assumption of similar effects within groups. In addition, the computational algorithms for the proposed methods may be sensitive to the choices of tuning parameters, e.g., the learning rate in ADMM and the neural network architecture in the DNN method, as well as the initial values. The current proposed algorithm cannot guarantee the global optimal solution. It is of great interest to develop a data-adaptive and more robust method to choose those parameters.

APPENDICES

APPENDIX A

Theoretical Properties

A.1 Theoretical Properties of Chapter 3: Scalar on network regression via boosting

This section summarizes the main theoretical properties of the method proposed in Chapter 2: scalar on network regression via boosting (Section 3.2). Broadly, we show that the first stage of the group boosting algorithm provides a close approximation to the true values when edges in the network are not grouped. The deviation from the true value is a function of the average value of β, the max deviation within a group from β_g, and the expected value of the squared summation of all the elements in group g. The properties will be investigated separately for two cases: homogeneity of the effects within groups and heterogeneity of the effects within each group.

We will denote the summary of edges within a given group as Z_{ig} and the individual covariates as X_{ijk}, such that $Z_{ig} = \frac{\sum_{jk \in g} X_{ijk}}{m_g}$. The networks are defined by g and m_g is the number of edges in group g.

The properties of this method vary based on the assumptions made on the group effects. In some cases, one may expect the effect of all edges in a given region to be

the same, whereas in other cases more heterogeneity may be assumed. Consider the following cases: one homogeneous group, one heterogeneous group, multiple homogeneous groups, and multiple heterogeneous groups.

A.1.1 Scenario 1: Homogeneous effect within groups

Suppose for the simplest case that all the edges belong to one group. The only difference between this approach and standard L_2Boosting without group selection is that we force all the variables to have the same effect on the outcome - homogeneity within a single group. First we will assume this to the be the true underlying model, so we are estimating with a correctly specified model, where we assume $\beta_1 = \beta_2 = \ldots = \beta_{m_g}$. In the case that the true model is $\beta_1 = \ldots = \beta_{m_g}$, this approach may be more efficient because we force the estimates to be equal within the group.

We know the true relationship between Z_i and x_{ijk} is $Z_i = \frac{\sum_{(j,k)\in g} \beta_{jk} x_{ijk}}{\lambda_1}$. We will estimate Z_i with $\tilde{Z}_i = \frac{\sum_{(j,k)\in g} X_{ijk}}{m_g}$, the model we will use for group selection is

$$Y = \lambda_0 + \lambda_1 \tilde{Z}$$

Consider the expected difference between the true Z and the estimate $\tilde{Z}$:

$$E\{\|Z - \tilde{Z}\|_2^2\} = E\left\{\left(\frac{\sum_{(j,k)\in g} \beta_{jk} x_{ijk}}{\lambda_1} - \frac{\sum_{(j,k)\in g} x_{ijk}}{m_g}\right)^2\right\} \tag{A.1}$$

If we take the derivative with respect to $\frac{1}{\lambda_1}$ we can obtain:

$$\frac{\partial(Z-\tilde{Z})^2}{\partial\lambda_1} = 2\left(\frac{\sum_{(j,k)\in g}\beta_{jk}x_{ijk}}{\lambda_1} - \frac{\sum_{(j,k)\in g}x_{ijk}}{m_g}\right)\left(\sum_{(j,k)\in g}\beta_{jk}x_{ijk}\right) \tag{A.2}$$

$$0 = 2\left(\frac{\sum_{(j,k)\in g}\beta_{jk}x_{ijk}}{\lambda_1} - \frac{\sum_{(j,k)\in g}x_{ijk}}{m_g}\right)\left(\sum_{(j,k)\in g}\beta_{jk}x_{ijk}\right) \tag{A.3}$$

$$= \frac{(\sum_{(j,k)\in g}\beta_{jk}x_{ijk})^2}{\lambda_1} - \frac{(\sum_{(j,k)\in g}x_{ijk})(\sum_{(j,k)\in g}\beta_{jk}x_{ijk})}{m_g} \tag{A.4}$$

$$\frac{1}{\lambda_1} = E\left\{\frac{\sum_{(j,k)\in g}x_{ijk}}{m_g(\sum_{(j,k)\in g}\beta_{jk}x_{ijk})}\right\} \tag{A.5}$$

In the simplest case if $\beta_{jk} = \beta_g$ for all $(j,k) \in g$ then this simplifies to be

$$\frac{1}{\lambda_1} = E\left\{\frac{\sum_{(j,k)\in g}X_{ijk}}{m_g\beta_g\sum_{(j,k)\in g}X_{ijk}}\right\} = \frac{1}{m_g\beta_g} \tag{A.6}$$

We can see in this very simple case that the expected value of the difference between our estimated group effect and the true value is:

$$E\{\|Z-\tilde{Z}\|_2^2\} = E\left\{\sum_{i=1}^{n}\left(\frac{\sum_{(j,k)\in g}\beta_{jk}x_{ijk}}{\lambda_1} - \frac{\sum_{(j,k)\in g}x_{ijk}}{m_g}\right)^2\right\} \tag{A.7}$$

$$= E\left\{\sum_{i=1}^{n}\left(\frac{\beta_g\sum_{(j,k)\in g}x_{ijk}}{m_g\beta_g} - \frac{\sum_{(j,k)\in g}x_{ijk}}{m_g}\right)^2\right\} \tag{A.8}$$

$$= 0 \tag{A.9}$$

Therefore in scenario 1, where we assume the effect within group is entirely homogeneous and we have only one group, the expected squared error of estimating $\tilde{Z}$ $E\{\|Z-\tilde{Z}\|_2^2\}$ is zero. We can see that this result expands to the case where we have multiple groups and the effect within each group is homogeneous. In this case we also see that the expected squared error of $\|Z-\tilde{Z}\|$ is zero.

A.1.2 Scenario 2: Heterogeneous effect within groups

In this second scenario, we now relax the assumption that the effect of edges within a group are exactly the same. Allowing some heterogeneity of the effect within groups is more realistic, though we will still assume the effects are similar within a given group.

The general form of λ_g is

$$\frac{1}{\lambda_g} = E\left\{\frac{\sum_{(j,k)\in g} x_{ijk}}{m_g \sum_{(j,k)\in g} \beta_{jk} x_{ijk}}\right\}$$

If we consider the heterogeneity among a group of edges as the difference from the average effect, we can define $\beta_{jk} = \bar{\beta}_g + \delta_{jk}$, where $\bar{\beta}_g$ is the average effect of an edge in group g on the outcome. δ_{jk} represents the deviation of the effect of edge (j,k) on the outcome from the average of the group's effect. Now we can define λ in the following way,

$$\frac{1}{\lambda_g} = E\left\{\frac{\sum_{(j,k)\in g} X_{ijk}}{m_g \bar{\beta}_g \sum_{(j,k)\in g} X_{ijk} + m_g \sum_{(j,k)\in g} \delta_{jk} X_{ijk}}\right\} \tag{A.10}$$

$$\lambda_g = \frac{1}{E\left\{\frac{\sum_{(j,k)\in g} X_{ijk}}{m_g \bar{\beta}_g \sum_{(j,k)\in g} X_{ijk} + m_g \sum_{(j,k)\in g} \delta_{jk} X_{ijk}}\right\}} \tag{A.11}$$

We would like to obtain the upper bound for the difference between Z and $\tilde{Z}$ in order to quantify the approximation error of this approach.

Taylor series expansion can be used to show the following approximation

$$E\left(\frac{1}{X}\right) \approx E\left(\frac{1}{E(X)} - \frac{(X - E(X))}{E(X)^2} + \frac{(X - E(X))^2}{E(X)^3}\right) \tag{A.12}$$

$$= \frac{1}{E(X)} + \frac{Var(X)}{E(X)^3} \tag{A.13}$$

Now if we take $'X'$ to be $\frac{m_g \bar{\beta} \sum X + m_g \sum \delta X}{\sum X} = m_g \bar{\beta} + m_g u_g$, where $u_g = \frac{\sum_{j\in g} \delta_j X_j}{\sum X_j}$,

then we get

$$\frac{1}{\lambda} = E\left(\frac{\sum X}{m_g\bar{\beta}\sum X + m_g\sum \delta X}\right) \approx \frac{1}{E(m_g\bar{\beta} + m_g u_g)} + \frac{Var(m_g\bar{\beta} + m_g u_g)}{E(m_g\bar{\beta} + m_g u_g)^3} \quad (A.14)$$

$$= \frac{1}{m_g\bar{\beta} + m_g E(u_g)} + \frac{m_g^2 Var(u_g)}{[m_g\bar{\beta} + m_g E(u_g)]^3} \quad (A.15)$$

We can use this to evaluate an approximation of the distance between Z and $\tilde{Z}$.

$$E\|Z - \tilde{Z}\|_2^2 = E\left\{\sum_i \left(\frac{\sum X\beta}{\lambda_g} - \frac{\sum X}{m_g}\right)^2\right\} \quad (A.16)$$

$$\approx E\left\{\sum_i \left(\sum X(\bar{\beta}_g + \delta)\left(\frac{1}{m_g\bar{\beta}_g + m_g E(u_g)} + \frac{m_g^2 Var(u_g)}{[m_g\bar{\beta}_g + m_g E(u_g)]^3}\right) - \frac{\sum X}{m_g}\right)^2\right\} \quad (A.17)$$

$$= n\,E\left\{\left(\frac{\bar{\beta}_g\sum X + \sum \delta X}{m_g\bar{\beta}_g + m_g E(u_g)} + \frac{m_g^2 Var(u_g)\sum X(\bar{\beta}_g + \delta)}{[m_g\bar{\beta}_g + m_g E(u_g)]^3} - \frac{\sum X}{m_g}\right)^2\right\} \quad (A.18)$$

$$= n\,E\left\{\left(\frac{\bar{\beta}_g\sum X + u_g\sum X}{m_g\bar{\beta}_g + m_g E(u_g)} - \frac{\sum X}{m_g} + \frac{m_g^2 Var(u_g)\sum X(\bar{\beta}_g + \delta)}{[m_g\bar{\beta}_g + m_g E(u_g)]^3}\right)^2\right\} \quad (A.19)$$

$$= n\,E\left\{\left(\frac{\sum X}{m_g}\left(\frac{\bar{\beta}_g + u_g}{\bar{\beta}_g + E(u_g)} - 1\right) + \frac{m_g^2 Var(u_g)\sum X(\bar{\beta}_g + \delta)}{[m_g\bar{\beta}_g + m_g E(u_g)]^3}\right)^2\right\} \quad (A.20)$$

$$= n\,E\left\{\left(\frac{\sum X}{m_g}\left(\frac{\bar{\beta}_g + u_g}{\bar{\beta}_g + E(u_g)} - 1\right) + \frac{Var(u_g)(\bar{\beta}_g + u_g)\sum X}{m_g(\bar{\beta}_g + E(u_g))^3}\right)^2\right\} \quad (A.21)$$

$$= n\,E\left\{\left(\frac{\sum X}{m_g}\left(\frac{\bar{\beta}_g + u_g}{\bar{\beta}_g + E(u_g)} - 1 + \frac{Var(u_g)(\bar{\beta}_g + u_g)}{(\bar{\beta}_g + E(u_g))^3}\right)\right)^2\right\} \quad (A.22)$$

$$E(u_g) = E\left\{\frac{\sum_{(j,k)\in g}\delta_{jk}x_{ijk}}{\sum x_{ijk}}\right\} \tag{A.23}$$

$$Var(u_g) = Var\left\{\frac{\sum_{(j,k)\in g}\delta_{jk}x_{ijk}}{\sum x_{ijk}}\right\} \tag{A.24}$$

If we assume that $E(u_g) = 0$ then this becomes

$$E\left\{\sum_{i=1}^{n}(Z_i - \tilde{Z}_i)^2\right\} = E\left\{\sum_{i=1}^{n}\left(\frac{\sum_{jk}x_{ijk}}{m_g}\left(\frac{\bar{\beta}_g + u_g}{\bar{\beta}_g + E(u_g)} - 1 + \frac{Var(u_g)(\bar{\beta}_g + u_g)}{(\bar{\beta}_g + E(u_g))^3}\right)\right)^2\right\} \tag{A.25}$$

$$= n\,E\left\{\left(\frac{\sum_{jk}x_{ijk}}{m_g}\left(\frac{\bar{\beta}_g + u_g}{\bar{\beta}_g} - 1 + \frac{Var(u_g)(\bar{\beta}_g + u_g)}{(\bar{\beta}_g)^3}\right)\right)^2\right\} \tag{A.26}$$

$$= n\,E\left\{\left(\frac{\sum_{jk}X_{jk}}{m_g}\left(\frac{u_g}{\bar{\beta}_g} + \frac{Var(u_g)(\bar{\beta}_g + u_g)}{\bar{\beta}_g^3}\right)\right)^2\right\} \tag{A.27}$$

$$= n\,E\left\{\left(\frac{\sum_{jk}X_{jk}}{m_g\bar{\beta}_g}\left(u_g + \frac{Var(u_g)(\bar{\beta}_g + u_g)}{\bar{\beta}_g^2}\right)\right)^2\right\} \tag{A.28}$$

Under the assumption that $E(u_g) = 0$, $Var(u_g) = E(u_g^2) - E(u_g)^2 = E(u_g^2)$. If we

also take $|\delta_{jk}| < \epsilon$ for all $(j,k) \in g$. So the equation becomes:

$$= n\,E\left\{\left(\frac{\sum_{jk} x_{ijk}}{m_g\bar{\beta}_g}\left(u_g + \frac{E(u_g^2)(\bar{\beta}_g + u_g)}{\bar{\beta}_g^2}\right)\right)^2\right\} \tag{A.29}$$

$$= \frac{n}{m_g\bar{\beta}_g^3}E\left\{\left(\sum_{jk} x_{ijk}\left(u_g\bar{\beta}_g^2 + \bar{\beta}_g E(u_g^2) + u_g E(u_g^2)\right)\right)^2\right\} \tag{A.30}$$

$$\begin{aligned}
&= \frac{n}{m_g}\left(\bar{\beta}_g + \bar{\beta}_g^{-3}E(u_g^2)^2 + 2\,\bar{\beta}_g^{-1}E(u_g^2)\right)E\left\{\left(\sum_{jk\in g}\delta_{jk}x_{ijk}\right)^2\right\} \\
&+ \frac{n}{m_g}\left(2\,E(u_g^2) + 2\,\bar{\beta}_g^{-2}E(u_g^2)^2\right)E\left\{\left(\sum_{jk\in g}\delta_{jk}x_{ijk}\right)\left(\sum_{jk\in g}x_{ijk}\right)\right\} \\
&+ \frac{n}{m_g}\,\bar{\beta}_g^{-1}E(u_g^2)^2 E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\}
\end{aligned} \tag{A.31}$$

$$\begin{aligned}
&< \frac{n}{m_g}\left(\bar{\beta}_g + \bar{\beta}_g^{-3}\epsilon^4 + 2\,\bar{\beta}_g^{-1}\epsilon^2\right)\epsilon^2 E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\} \\
&+ \frac{n}{m_g}\left(2\,\epsilon^2 + 2\,\bar{\beta}_g^{-2}\epsilon^4\right)\epsilon E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\} \\
&+ \frac{n}{m_g}\,\bar{\beta}_g^{-1}\epsilon^4 E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\}
\end{aligned} \tag{A.32}$$

$$= \frac{n}{m_g}(\bar{\beta}_g\epsilon^2 + 2\epsilon^3 + 3\bar{\beta}_g^{-1}\epsilon^4 + 2\bar{\beta}_g^{-2}\epsilon^5 + \bar{\beta}_g^{-3}\epsilon^6)E\left\{\left(\sum_{jk\in g}x_{ijk}\right)^2\right\} \tag{A.33}$$

The deviation of Z from $\tilde{Z}$ can be summarized by the average value of β, the max deviation any β from the average (ϵ), and the expected value of the square of the summation of all elements in group g.

APPENDIX B

Sensitivity Analyses

B.1 Sensitivity Analysis of Grouping in ABCD analysis

Two additional node groupings were tested to understand how sensitive the results are to the choice of grouping. Both of these groups were determined using the node coordinates in combination with the functional network structure. One clustered each functional network into 3 groups based on the spatial location of nodes, and the other grouped them into 5 groups per functional network based on spatial coordinates.

We found that the edges identified did not directly overlap the ones identified in the primary analysis, however the functional networks represented did have substantial overlap. Additionally, these two groupings had a large proportion of overlap between each other, among groups and edges. Using spatial coordinates to group the nodes, regardless of the number of clusters, identified very similar edges as associated with cognition. The main functional networks identified by this partition were: default mode, cingul-opercular task control, and visual networks.

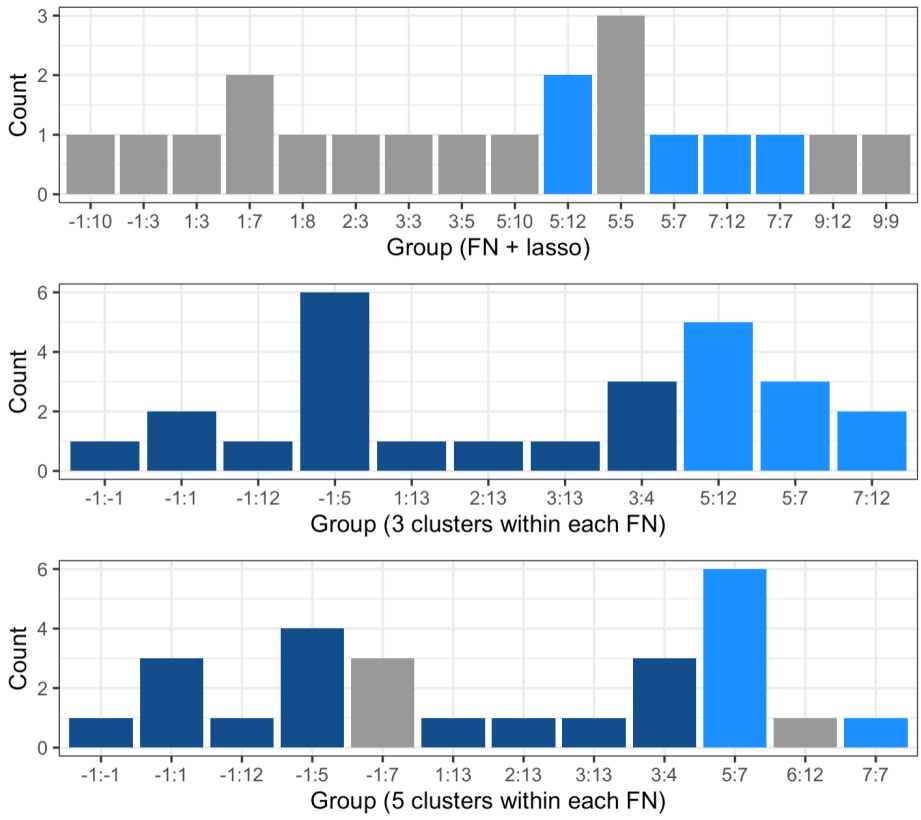


Figure B.1: Figure showing the functional networks identified by different group partitions of the network nodes. 'Count' shows the number of edges identified. The dark blue columns are those that overlap between the two groups determined using spatial coordinates, and the light blue denotes those that overlap with the groups identified in the primary analysis.

The figure shows how edges spanning the default mode network and the visual network or cingulo-opercular task control network are identified across all three group partitions. It also shows how the edges identified are quite consistent across the two groupings determined using spatial coordinates, despite dividing into different number of clusters. In terms of predicted R^2, all three of these performed similarly (FN + lasso had a fitted R^2 of 0.320; 3 clusters within each functional network had a fitted R^2 of 0.310; 5 clusters within each functional network had a fitted R^2 of 0.314).

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Achenbach, T. M. (1999), The child behavior checklist and related instruments.

[2] Anticevic, A., D. L. Dierker, S. K. Gillespie, G. Repovs, J. G. Csernansky, D. C. Van Essen, and D. M. Barch (2008), Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia, *Neuroimage*, *41*(3), 835–848.

[3] Anticevic, A., et al. (2014), Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness, *Cerebral cortex*, *24*(12), 3116–3130, doi:10.1093/cercor/bht165.

[4] Balodis, I. M., and M. N. Potenza (2015), Anticipatory reward processing in addicted populations: a focus on the monetary incentive delay task, *Biological psychiatry*, *77*(5), 434–444.

[5] Boyd, S., N. Parikh, and E. Chu (2011), *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc.

[6] Breiman, L. (2001), Random forests, *Machine learning*, *45*(1), 5–32, doi:10.1023/A:1010933404324.

[7] Brix, A., and J. Møller (2001), Space-time multitype log gaussian cox processes with a view to modeling weed data, *Scandinavian Journal of Statistics*, *28*, 471–488.

[8] Brodoehl, S., C. Gaser, R. Dahnke, O. W. Witte, and C. M. Klingner (2020), Surface-based analysis increases the specificity of cortical activation patterns and connectivity results, *Scientific reports*, *10*(1), 1–13.

[9] Bühlmann, P., and B. Yu (2003), Boosting with the l 2 loss: regression and classification, *Journal of the American Statistical Association*, *98*(462), 324–339.

[10] Bühlmann, P., et al. (2006), Boosting for high-dimensional linear models, *The Annals of Statistics*, *34*(2), 559–583.

[11] Calhoun, V. D., T. Adalı, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson (2006), A method for multitask fmri data fusion applied to schizophrenia, *Human brain mapping*, *27*(7), 598–610.

[12] Casey, B., et al. (2018), The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites, *Developmental cognitive neuroscience*, *32*, 43–54.

[13] Chen, J., and Z. Chen (2008), Extended bayesian information criteria for model selection with large model spaces, *Biometrika*, *95*(3), 759–771.

[14] Chen, Y., X. Wang, L. Kong, and H. Zhu (2016), Local region sparse learning for image-on-scalar regression, *arXiv preprint arXiv:1605.08501*.

[15] Cireşan, D. C., A. Giusti, L. M. Gambardella, and J. Schmidhuber (2013), Mitosis detection in breast cancer histology images with deep neural networks, in *International conference on medical image computing and computer-assisted intervention*, pp. 411–418, Springer.

[16] Clinton, S. M., and J. H. Meador-Woodruff (2004), Thalamic dysfunction in schizophrenia: neurochemical, neuropathological, and in vivo imaging abnormalities, *Schizophrenia research*, *69*(2-3), 237–253, doi:10.1016/j.schres.2003.09.017.

[17] Cortes, C., and V. Vapnik (1995), Support-vector networks, *Machine learning*, *20*(3), 273–297, doi:10.1007/BF00994018.

[18] Darnai, G., et al. (2019), Internet addiction and functional brain networks: task-related fmri study, *Scientific reports*, *9*(1), 1–10.

[19] Erhardt, E. B., S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adali, and V. D. Calhoun (2011), Comparison of multi-subject ica methods for analysis of fmri data, *Human brain mapping*, *32*(12), 2075–2095.

[20] Fan, J., Y. Feng, and Y. Wu (2009), Network exploration via the adaptive lasso and scad penalties, *The annals of applied statistics*, *3*(2), 521.

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

[22] Fox, M. D., M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle (2006), Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems, *Proceedings of the National Academy of Sciences*, *103*(26), 10,046–10,051, doi:10.1073/pnas.0604187103.

[23] Friedman, J., T. Hastie, and R. Tibshirani (2008), Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, *9*(3), 432–441.

[24] Friedman, J., T. Hastie, and R. Tibshirani (2010), Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, *33*(1), 1–22.

[25] Goldsmith, J., L. Huang, and C. M. Crainiceanu (2014), Smooth scalar-on-image regression via spatial bayesian variable selection, *Journal of Computational and Graphical Statistics*, *23*(1), 46–64.

[26] González-Recio, O., K. A. Weigel, D. Gianola, H. Naya, and G. J. Rosa (2010), L 2-boosting algorithm applied to high-dimensional problems in genomic selection, *Genetics research*, *92*(3), 227–237.

[27] Gordon, E. M., T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen (2016), Generation and evaluation of a cortical area parcellation from resting-state correlations, *Cerebral cortex*, *26*(1), 288–303.

[28] Greicius, M. D., B. Krasnow, A. L. Reiss, and V. Menon (2003), 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, doi:10.1073/pnas.0135058100.

[29] Greicius, M. D., G. Srivastava, A. L. Reiss, and V. Menon (2004), Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI, *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4637–4642, doi:10.1073/pnas.0308627101.

[30] Greicius, M. D., B. H. Flores, V. Menon, G. H. Glover, H. B. Solvason, H. Kenna, A. L. Reiss, and A. F. Schatzberg (2007), Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus, *Biological psychiatry*, *62*(5), 429–437.

[31] Guha, S., and A. Rodriguez (2020), Bayesian regression with undirected network predictors with an application to brain connectome data, *Journal of the American Statistical Association*, pp. 1–13.

[32] Gulshan, V., et al. (2016), Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *Jama*, *316*(22), 2402–2410.

[33] Hollon, T. C., et al. (2020), Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks, *Nature medicine*, *26*(1), 52–58.

[34] Holtmann, M., A. F. Buchmann, G. Esser, M. H. Schmidt, T. Banaschewski, and M. Laucht (2011), The child behavior checklist-dysregulation profile predicts substance use, suicidality, and functional impairment: A longitudinal analysis, *Journal of Child Psychology and Psychiatry*, *52*(2), 139–147.

[35] Hommer, D. W., J. M. Bjork, and J. M. Gilman (2011), Imaging brain response to reward in addictive disorders, *Annals of the New York Academy of Sciences*, *1216*(1), 50–61.

[36] Howes, O. D., and S. Kapur (2009), The dopamine hypothesis of schizophrenia: version III – the final common pathway, *Schizophrenia bulletin, Journal of Psychoses and Related Disorders*, *35*(3), 549–562, doi:10.1093/schbul/sbp006.

[37] Insel, T., B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, and P. Wang (2010), Research domain criteria (RDoC): toward a new classification framework for research on mental disorders, doi:10.1176/appi.ajp.2010.09091379.

[38] Jang, H., S. M. Plis, V. D. Calhoun, and J.-H. Lee (2017), Task-specific feature extraction and classification of fmri volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks, *NeuroImage*, *145*, 314–328.

[39] Kang, J., M. Körner, Y. Wang, H. Taubenböck, and X. X. Zhu (2018), Building instance classification using street view images, *ISPRS journal of photogrammetry and remote sensing*, *145*, 44–59.

[40] Kang, J., B. J. Reich, and A.-M. Staicu (2018), Scalar-on-image regression via the soft-thresholded gaussian process, *Biometrika*, *105*(1), 165–184.

[41] Kessler, D., M. Angstadt, and C. Sripada (2016), Growth charting of brain connectivity networks and the identification of attention impairment in youth, *JAMA Psychiatry*, *73*(5), 481–489, doi:10.1001/jamapsychiatry.2016.0088.

[42] Khaki, S., and L. Wang (2019), Crop yield prediction using deep neural networks, *Frontiers in plant science*, *10*, 621.

[43] Kingma, D. P., and J. Ba (2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.

[44] Li, F., T. Zhang, Q. Wang, M. Z. Gonzalez, E. L. Maresh, J. A. Coan, et al. (2015), Spatial bayesian variable selection and grouping for high-dimensional scalar-on-image regression, *The Annals of Applied Statistics*, *9*(2), 687–713.

[45] Li, X., T. Zhao, L. Wang, X. Yuan, and H. Liu (2019), *flare: Family of Lasso Regression*, r package version 1.6.0.2.

[46] Li, X., L. Wang, H. J. Wang, and A. D. N. Initiative (2020), Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression, *Journal of the American Statistical Association*, pp. 1–15.

[47] Lin, J., N. T. Clancy, J. Qi, Y. Hu, T. Tatla, D. Stoyanov, L. Maier-Hein, and D. S. Elson (2018), Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks, *Medical image analysis*, *48*, 162–176.

[48] Litjens, G., et al. (2016), Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, *Scientific reports*, *6*(1), 1–11.

[49] Liu, H., J. Lafferty, and L. Wasserman (2009), The nonparanormal: Semiparametric estimation of high dimensional undirected graphs., *Journal of Machine Learning Research*, *10*(10).

[50] Luciana, M., J. Bjork, B. Nagel, D. Barch, R. Gonzalez, S. Nixon, and M. Banich (2018), Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (abcd) baseline neurocognition battery, *Developmental cognitive neuroscience*, *32*, 67–79.

[51] Mahmood, O., D. Goldenberg, R. Thayer, R. Migliorini, A. Simmons, and S. Tapert (2013), Adolescents' fmri activation to a response inhibition task predicts future substance use, *Addictive behaviors*, *38*(1), 1435–1441.

[52] Martino, M., et al. (2020), Abnormal functional relationship of sensorimotor network with neurotransmitter-related nuclei via subcortical-cortical loops in manic and depressive phases of bipolar disorder, *Schizophrenia bulletin, Journal of Psychoses and Related Disorders*, *46*(1), 163–174.

[53] Mayr, A., H. Binder, O. Gefeller, and M. Schmid (2014), The evolution of boosting algorithms-from machine learning to statistical modelling, *arXiv preprint arXiv:1403.1452*.

[54] Meinshausen, N., P. Bühlmann, et al. (2006), High-dimensional graphs and variable selection with the lasso, *The annals of statistics*, *34*(3), 1436–1462.

[55] Mumford, J. A., and J. D. Ramsey (2014), Bayesian networks for fmri: a primer, *Neuroimage*, *86*, 573–582.

[56] Nassif, A. B., I. Shahin, I. Attili, M. Azzeh, and K. Shaalan (2019), Speech recognition using deep neural networks: A systematic review, *IEEE access*, *7*, 19,143–19,165.

[57] Orru, G., W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli (2012), Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review, *Neuroscience & Biobehavioral Reviews*, *36*(4), 1140–1152.

[58] Patel, R. S., F. D. Bowman, and J. K. Rilling (2006), A bayesian approach to determining connectivity of the human brain, *Human brain mapping*, *27*(3), 267–276.

[59] Power, J. D., et al. (2011), Functional network organization of the human brain, *Neuron*, *72*(4), 665–678, doi:10.1016/j.neuron.2011.09.006.

[60] Ramezani, M., K. Marble, H. Trang, I. S. Johnsrude, and P. Abolmaesumi (2014), Joint sparse representation of brain activity patterns in multi-task fmri data, *IEEE Transactions on Medical Imaging*, *34*(1), 2–12.

[61] Relión, J. D. A., D. Kessler, E. Levina, S. F. Taylor, et al. (2019), Network classification with applications to brain connectomics, *The Annals of Applied Statistics*, *13*(3), 1648–1677.

[62] Sandino, C. M., J. Y. Cheng, F. Chen, M. Mardani, J. M. Pauly, and S. S. Vasanawala (2020), Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging, *IEEE signal processing magazine*, *37*(1), 117–127.

[63] Satterthwaite, T. D., et al. (2014), Neuroimaging of the philadelphia neurodevelopmental cohort, *Neuroimage*, *86*, 544–553, doi:https://doi.org/10.1016/j.neuroimage.2013.07.064.

[64] Satterthwaite, T. D., et al. (2016), The philadelphia neurodevelopmental cohort: A publicly available resource for the study of normal and abnormal brain development in youth, *NeuroImage*, *124*, 1115–1119, doi:10.1016/j.neuroimage.2015.03.056.

[65] Schmidhuber, J. (2015), Deep learning in neural networks: An overview, *Neural networks*, *61*, 85–117.

[66] Schmidt, A., et al. (2013), Brain connectivity abnormalities predating the onset of psychosis: correlation with the effect of medication, *JAMA psychiatry*, *70*(9), 903–912.

[67] Schmidt-Hieber, J., et al. (2020), Nonparametric regression using deep neural networks with relu activation function, *Annals of Statistics*, *48*(4), 1875–1897.

[68] Schultz, W., P. Dayan, and P. R. Montague (1997), A neural substrate of prediction and reward, *Science*, *275*(5306), 1593–1599, doi:10.1126/science.275.5306.1593.

[69] Sheffield, J. M., et al. (2015), Fronto-parietal and cingulo-opercular network integrity and cognition in health and schizophrenia, *Neuropsychologia*, *73*, 82–93, doi:10.1016/j.neuropsychologia.2015.05.006.

[70] Sheline, Y. I., J. L. Price, Z. Yan, and M. A. Mintun (2010), Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus, *Proceedings of the National Academy of Sciences*, *107*(24), 11,020–11,025, doi:10.1073/pnas.1000446107.

[71] Smith, S. M., et al. (2009), Correspondence of the brain's functional architecture during activation and rest, *Proceedings of the national academy of sciences*, *106*(31), 13,040–13,045.

[72] Smith, S. M., et al. (2013), Functional connectomics from resting-state fMRI, *Trends in Cognitive Sciences*, *17*(12), 666 – 682, doi:https://doi.org/10.1016/j.tics.2013.09.016, special Issue: The Connectome.

[73] Solo, V., J.-B. Poline, M. A. Lindquist, S. L. Simpson, F. Du Bois Bowman, M. Chung, and B. Cassidy (2018), Connectivity in fMRI: Blind spots and

breakthroughs, *IEEE Transactions on Medical Imaging*, *37*(7), 1537 – 1550, doi:10.1109/TMI.2018.2831261.

[74] Sripada, C., S. Rutherford, M. Angstadt, W. K. Thompson, M. Luciana, A. Weigard, L. H. Hyde, and M. Heitzeg (2019), Prediction of neurocognition in youth from resting state fmri, *Molecular psychiatry*, pp. 1–9.

[75] Sripada, R. K., A. P. King, S. N. Garfinkel, X. Wang, C. S. Sripada, R. C. Welsh, and I. Liberzon (2012), Altered resting-state amygdala functional connectivity in men with posttraumatic stress disorder, *Journal of Psychiatry & Neuroscience: JPN*, *37*(4), 241–249, doi:10.1503/jpn.110069.

[76] Supekar, K., V. Menon, D. Rubin, M. Musen, and M. D. Greicius (2008), Network analysis of intrinsic functional brain connectivity in alzheimer's disease, *PLoS Comput Biol*, *4*(6), e1000,100.

[77] Thompson, W. K., D. M. Barch, J. M. Bjork, R. Gonzalez, B. J. Nagel, S. J. Nixon, and M. Luciana (2019), The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the abcd study's baseline neurocognitive battery, *Developmental cognitive neuroscience*, *36*, 100,606.

[78] Tureen, T., E. L. Morris, C. Sripada, and J. Kang (2021), Identifying resting-state functional network markers for predicting neurocognitive profiles in adolescence using deep neural networks [working paper].

[79] Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot (2002), Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, *NeuroImage*, *15*(1), 273–289, doi:10.1006/nimg.2001.0978.

[80] Wang, H., et al. (2012), Bayesian graphical lasso models and efficient posterior computation, *Bayesian Analysis*, *7*(4), 867–886.

[81] Wang, Y., J. Kang, P. B. Kemmer, and Y. Guo (2016), An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation, *Frontiers in neuroscience*, *10*, 123.

[82] Weinstein, J. J., M. O. Chohan, M. Slifstein, L. S. Kegeles, H. Moore, and A. Abi-Dargham (2017), Pathway-specific dopamine abnormalities in schizophrenia, *Biological psychiatry*, *81*(1), 31–42, doi:10.1016/j.biopsych.2016.03.2104.

[83] Xia, C. H., et al. (2018), Linked dimensions of psychopathology and connectivity in functional brain networks, *Nature Communications*, *9*(3003), 1–14, doi:10.1038/s41467-018-05317-y.

[84] Xia, M., J. Wang, and Y. He (2013), Brainnet viewer: a network visualization tool for human brain connectomics, *PloS one*, *8*(7), e68,910.

[85] Xia, Y., L. Li, S. N. Lockhart, and W. J. Jagust (2019), Simultaneous covariance inference for multimodal integrative analysis, *Journal of the American Statistical Association*.

[86] Yang, Y., H. Zou, and S. Bhatnagar (2020), *gglasso: Group Lasso Penalized Learning Using a Unified BMD Algorithm*, r package version 1.5.

[87] Yuan, M., and Y. Lin (2006), Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.

[88] Zhang, D., L. Li, C. Sripada, and J. Kang (2020), Image-on-scalar regression via deep neural networks, *arXiv preprint arXiv:2006.09911*.

[89] Zhang, J., M. Liu, and D. Shen (2017), Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks, *IEEE Transactions on Image Processing*, *26*(10), 4753–4764.

[90] Zhang, J., W. W. Sun, and L. Li (2018), Network response regression for modeling population of networks with covariates, *arXiv preprint arXiv:1810.03192*.

[91] Zhao, T., X. Li, H. Liu, K. Roeder, J. Lafferty, and L. Wasserman (2015), *huge: High-Dimensional Undirected Graph Estimation*, r package version 1.2.7.

[92] Zhao, Y., et al. (2017), Automatic recognition of fmri-derived functional networks using 3-d convolutional neural networks, *IEEE Transactions on Biomedical Engineering*, *65*(9), 1975–1984.

[93] Zhou, H., L. Li, and H. Zhu (2013), Tensor regression with applications in neuroimaging data analysis, *Journal of the American Statistical Association*, *108*(502), 540–552.

[94] Zhu, H., J. Fan, and L. Kong (2014), Spatially varying coefficient model for neuroimaging data with jump discontinuities, *Journal of the American Statistical Association*, *109*(507), 1084–1098.

[95] Zhu, Y., and I. Cribben (2018), Sparse graphical models for functional connectivity networks: best methods and the autocorrelation issue, *Brain connectivity*, *8*(3), 139–165.