Functional Connectivity Network Analysis with Discriminative Hub Detection for Brain Disease Identification

Mingliang Wang,[1] Jiashuang Huang,[1] Mingxia Liu,[2*] Daoqiang Zhang[1*]
[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[*]Corresponding authors (mxliu@med.unc.edu, dqzhang@nuaa.edu.cn)

Abstract

Brain network analysis can help reveal the pathological basis of neurological disorders and facilitate automated diagnosis of brain diseases, by exploring connectivity patterns in the human brain. Effectively representing the brain network has always been the fundamental task of computer-aided brain network analysis. Previous studies typically utilize human-engineered features to represent brain connectivity networks, but these features may not be well coordinated with subsequent classifiers. Besides, brain networks are often equipped with multiple hubs (i.e., nodes occupying a central position in the overall organization of a network), providing essential clues to describe connectivity patterns. However, existing studies often fail to explore such hubs from brain connectivity networks. To address these two issues, we propose a Connectivity Network analysis method with discriminative Hub Detection (*CNHD*) for brain disease diagnosis using functional magnetic resonance imaging (fMRI) data. Specifically, we incorporate both feature extraction of brain networks and network-based classification into a unified model, while discriminative hubs can be automatically identified from data via ℓ_1-norm and $\ell_{2,1}$-norm regularizers. The proposed CNHD method is evaluated on three real-world schizophrenia datasets with fMRI scans. Experimental results demonstrate that our method *not only* outperforms several state-of-the-art approaches in disease diagnosis, *but also* is effective in automatically identifying disease-related network hubs in the human brain.

Introduction

Recent advances in network analysis have allowed new insights for our understanding of complex networks and interactions (i.e., connectivity patterns) within each network. A human brain can be modeled as a complex network, containing a number of structurally/functionally interconnected brain regions (Power et al. 2010; Liu, Kong, and Ragin 2017). Generally, a brain network (a.k.a., connectome) can be characterized by a set of nodes and edges, where nodes represent regions-of-interest (ROIs) in the brain and edges denote the connectivity strength or correlation between brain regions. The characterization of brain networks can not only help uncover the pathological basis of brain disorders, but also lead to the development of biomarkers that quantify reorganizational mechanisms of brain diseases (Fox and Greicius 2010; Yang et al. 2015). Brain connectivity networks derived from resting-state functional magnetic resonance imaging (fMRI) scans have shown to be effective in revealing functional connectivity patterns of the brain (Sporns 2012; Bolton et al. 2017; Liu, Kong, and Philip 2017). In computer-aided brain network analysis, it remains a fundamental but challenging task to effectively represent functional brain networks constructed on fMRI data.

Existing fMRI-based studies first extract human-engineered features (e.g., clustering coefficient) to represent brain connectivity networks, and then feed these features to a pre-defined machine learning model for classification (Shervashidze et al. 2011; Wee et al. 2012; Jie et al. 2014). Since feature extraction and classifier training are treated as stand-alone tasks, the heterogeneity between features and classifiers would lead to a sub-optimal performance in brain disease diagnosis.The network embedding technology (Cao et al. 2017; Liu et al. 2018) is applied to brain network analysis by performing feature extraction and classifier training jointly. However, features of a brain network in these methods are usually manifested in the latent representation space, thus lacking of good interpretation of brain connectivity patterns.

On the other hand, network hubs have proven to be vulnerable to pathogenic processes of brain disorders (Dai, Bi, and He 2015), by effectively revealing communication and information integration across different brain regions. Here, the term "hub" refers to a particular node that occupies a central position in a brain connectivity network, reflecting the global structure of this network (Heuvel, Sporns, and Olaf 2013; He et al. 2016). Several network hubs have been shown to reveal significant abnormalities in brain connectivity networks of patients with schizophrenia and Alzheimer's disease (Huang et al. 2011; Stam 2014). Intuitively, exploring such hubs could bring prior knowledge for modeling global structure of brain networks, thus promoting the learning performance of network-based brain disease diagnosis. However, existing studies cannot effectively identify network hubs from brain connectivity networks.

To address these issues, in this paper, we propose a Connectivity Network analysis method with discriminative Hub Detection (*CNHD*) for brain disease diagnosis using fMRI

scans. Specifically, we first construct functional connectivity networks based on fMRI scans, with each subject corresponding to a specific network. Then, we develop a functional connectivity network based classification model, where both feature extraction and classifier training are incorporated into a unified framework. In particular, discriminative hubs can be automatically identified from brain connectivity networks, via l_1-norm and $l_{2,1}$-norm regularization terms in our CNHD model. To solve the proposed non-convex optimization problem, we develop an alternating direction method of multipliers (ADMM) (Boyd et al. 2011) algorithm. We evaluate the proposed method on three real schizophrenia datasets with fMRI scans, with experimental results demonstrating its effectiveness in both tasks of brain disease diagnosis and network hub detection.

The major contributions of this work are three-fold. *First*, we develop a functional connectivity network based classification model for brain disease diagnosis, with feature extraction and model training incorporated into a unified framework. *Second*, we propose to detect discriminative hubs from brain networks via both l_1-norm and $l_{2,1}$-norm based regularizers. To our knowledge, this is the first attempt to incorporate hub detection and classifier training into a unified model for functional connectivity network analysis. *Third*, we evaluate the proposed method on three real datasets, with results demonstrating the superiority of our method.

Related Work

Network-based Brain Disease Diagnosis Many network-based classification methods have been proposed for brain disease diagnosis. Based on different network representations, these methods can be roughly divided into two categories, including 1) topology-based representation methods, and 2) subgraph-based representation methods.

Topology-based representation methods typically directly compute global or local similarity between a pair of brain connectivity networks, using graph kernel (Jie et al. 2014) or clustering coefficient (CC) (Wee et al. 2012). The obtained similarities are used as feature representation of different networks and then fed into a pre-defined classifier (e.g., support vector machines, SVM) for brain disease diagnosis. For example, Jie et al. (Jie et al. 2014) utilize the topology-based Weisfeiler-Lehman graph kernel (Shervashidze et al. 2011) to measure the similarity between paired connectivity networks, followed by an SVM classifier for mild cognitive impairment identification (with graph kernel based representation as features of networks). Wee et al. (Wee et al. 2012) extract weighted local clustering coefficients between each ROI (w.r.t., a node) and the remaining ROIs as features to quantify the cliquishness of each ROI.

Subgraph-based representation methods usually first mine a set of discriminative subgraphs (with different pruning rules) to represent each network, and then resort to a particular classifier for disease diagnosis with such subgraph features as input data. For example, Zhang et al. (Zhang et al. 2018) design an ordinal pattern based network descriptor to mine frequent subgraphs for disease classification. This method can simultaneously modeling both weight information (i.e., connectivity strength) and ordinal relationship of weighted edges in a brain connectivity network while relying on the subsequent SVM for final prediction. Kong et al. (Kong et al. 2013) develop a subgraph feature selection method via dynamic programming, followed by an SVM for network-based disease diagnosis. To utilize the complementary information of multiple views, Cao et al. (Cao et al. 2016) introduce a subgraph mining algorithm for disease diagnosis by using multiple side views (e.g., cognitive measures) as guidance information.

It's worth noting that both topology-based and subgraph-based representation methods usually first extract particular network representation from brain connectivity networks, and then reply on pre-defined classifiers (e.g., SVM) for brain disease diagnosis. That is, feature extraction and model learning are treated as two separate tasks in these methods, so potential heterogeneity between features and classifiers may degrade the final performance of these methods. To address this issue, the network embedding technology (Cao et al. 2017; Liu et al. 2018) has been proposed for deriving representations from brain connectivity networks, attempting to integrate network feature learning and classifier training into a unified optimization problem. For example, Cao et al. (Cao et al. 2017) propose a tensor-based brain network embedding (t-BNE) method for automated diagnoses of anxiety disorder, by fusing the processes of tensor factorization and classifier learning on electroencephalogram brain networks. Liu et al. (Liu et al. 2018) develop a multi-view multi-graph embedding (M2E) method for multiple modalities brain network clustering analysis on human immunodeficiency virus and bipolar disorder. Although good performance can be produced by network embedding, those derived network features are often manifested in latent representation spaces and lack interpretability.

Hub Detection from Networks In the field of network analysis, several studies have recognized that exploring the hub structure in networks helps to better understand complex networks (Ruan and Parthasarathy 2014; He et al. 2016; Ma et al. 2017). For instance, to simultaneously identify community and structural role (e.g., hub nodes) assignments in social networks, Ruan et al. (Ruan and Parthasarathy 2014) employ a principled approach to guide the hub detection process in a nonparametric fashion. He et al. (He et al. 2016) design a harmonic modularity (HAM) method to simultaneously detect potential communities and the top-k hub nodes, using topological structures of a social network. Ma et al. (Ma et al. 2017) propose an auto-weighted multi-view graph embedding method with hub detection (MVGE-HD) for brain network clustering, based on multi-modality data (i.e., fMRI and diffusion tensor imaging). This method aims to learn a unified graph embedding across multiple views while reducing the potential influence of hubs on blurring the boundaries between node clusters in the network, and thus network hubs can be automatically identified from data. However, these methods cannot jointly identify discriminative hubs from functional brain connectivity networks and learn network-based classification models.

Methodology

In this section, we first introduce notations, and then present the proposed CNHD method in detail. We also describe our alternating optimization algorithm for solving the proposed problem and analyze its computational complexity.

Notations We use lower case letters (e.g., x), bold lower case letters (e.g., $\mathbf{x}$), and bold capital letters (e.g., $\mathbf{X}$) to denote scalars, vectors, and matrices, respectively. For a matrix $\mathbf{X}$, its (i,j)-th entry is denoted as $x_{i,j}$ and its i-th column is denoted as $\mathbf{x}_i$. Denote $tr(\cdot)$ as the trace of a matrix, which is defined to be the sum of the diagonal elements. The transpose of $\mathbf{X}$ is denoted as $\mathbf{X}^T$, and its Frobenius norm is denoted as $\|\mathbf{X}\|_F$. Also, the ℓ_1-norm of $\mathbf{X} \in \mathbb{R}^{n\times m}$ is defined as $\|\mathbf{X}\|_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}|x_{i,j}|$, while its $\ell_{2,1}$-norm is denoted as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{m} x_{i,j}^2}$. For functional connectivity networks derived from different subjects, we define the same set of M nodes, with each node corresponding to a particular ROI in the brain. Each network can be represented by its adjacency matrix $\mathbf{A} \in \mathbb{R}^{M\times M}$. For simplicity, we only consider undirected symmetric networks (i.e., $A_{i,j} = A_{j,i}$) without self-loops (i.e., $A_{i,i} = 0$).

Connectivity Network Analysis with Discriminative Hub Detection (CNHD) Denote a network set as $\{(\mathbf{A}^{(1)}, Y_1), (\mathbf{A}^{(2)}, Y_2), \cdots, (\mathbf{A}^{(N)}, Y_N)\}$, containing N training subjects and M nodes (i.e., $\mathbf{A}^{(n)} \in \mathbb{R}^{M\times M}$). Note that each network is constructed based on the fMRI scan of a particular subject. We denote the class label of the network $\mathbf{A}^{(n)}$ as Y_n, with $Y_n = 1$ denoting the n-th subject is a patient with a particular brain disease, while $Y_n = -1$ representing that this subject is a normal control (NC). Our goal is to learn a predictor to identify patients from the whole population based on their brain connectivity networks. A general form of network-based linear predictor can be described as

$$\arg\min_{\mathbf{\Theta}} \{\ell(\mathbf{\Theta}) + \Omega(\mathbf{\Theta})\}, \tag{1}$$

where $\ell(\mathbf{\Theta}) = \frac{1}{N}\sum_{n=1}^{N}\widetilde{\ell}(Y_n, \mathbf{A}^{(n)}; \mathbf{\Theta})$ is an empirical loss on the training set, and $\Omega(\mathbf{\Theta})$ is a regularization term. Here, the predictor $\mathbf{\Theta} \in \mathbb{R}^{M\times M}$ is a symmetric matrix, i.e., $\mathbf{\Theta} = \mathbf{\Theta}^T$. Also, the diagonal elements of $\mathbf{\Theta}$ are set to be zeros, i.e., $diag(\mathbf{\Theta}) = \mathbf{0}$. Although many types of loss functions can be used, for the sake of optimization, we mainly consider the logic loss function in this work:

$$\widetilde{\ell}(Y_n, \mathbf{A}^{(n)}; \mathbf{\Theta}, b) = \log(1 + \exp(-Y_n(tr(\mathbf{\Theta}^T\mathbf{A}^{(n)}) + b))), \tag{2}$$

where b is a bias term. In Eq. (2), we incorporate the classifier learning of $\mathbf{\Theta}$ into the process of learning network representation (i.e., $\mathbf{A}^{(n)}$). In this way, the weight matrix $\mathbf{\Theta}$ and the network data $\mathbf{A}^{(n)}$ can interact with each other in the unified learning framework.

Furthermore, to explicitly model the hub structure in a network, we decompose the weight matrix $\mathbf{\Theta}$ into the sum of two components $\mathbf{Z} \in \mathbb{R}^{M\times M}$ and $\mathbf{V} \in \mathbb{R}^{M\times M}$ (Tan et al. 2014; Gong, Ye, and Zhang 2012). As shown in Figure 1, $\mathbf{Z}$ is a sparse symmetric matrix, and $\mathbf{V}$ is a matrix whose columns are either entirely zeros or entirely non-zeros. The sparse entries of $\mathbf{Z}$ represent edges between non-hub nodes, and the non-zero columns of $\mathbf{V}$ are regarded as hub nodes. Formally, our CNHD model is formulated as:

$$\begin{aligned}\arg\min_{\mathbf{\Theta},b,\mathbf{Z},\mathbf{V}} \ & \frac{1}{N}\sum_{n=1}^{N}\log(1+\exp(-Y_n(tr(\mathbf{\Theta}^T\mathbf{A}^{(n)})+b))) \\ & + \lambda\|\mathbf{Z}\|_1 + \beta\|\mathbf{V}\|_1 + \gamma\|\mathbf{V}\|_{2,1} \\ \text{s.t.} \ & \mathbf{\Theta} = \mathbf{Z} + \mathbf{V}, \mathbf{Z} = \mathbf{Z}^T, \mathbf{V} = \mathbf{V}^T, \\ & diag(\mathbf{\Theta}) = \mathbf{0}\end{aligned} \tag{3}$$

where λ, β and γ are nonnegative parameters. Sparsity on $\mathbf{Z}$ is encouraged by its ℓ_1-norm. The $\ell_{2,1}$-norm in Eq. (3) is designed to induce group sparsity (Yuan and Lin 2006) using M groups, with each group denoting a specific columns of $\mathbf{V}$ that corresponds to an ROI. Due to the constraint of $\mathbf{V} = \mathbf{V}^T$, there will be overlap among these M groups. That is, the (i,j)-th element of $\mathbf{V}$ is contained in both the i-th and j-th groups. With overlapping groups, the group lasso penalty can estimate whose support is the complement of the union of groups, thus helping identify the most discriminative hub nodes. Besides, the ℓ_1-norm of $\mathbf{V}$ in Eq. (3) promotes sparsity inside the group, allowing us to select a subset of edges for a hub node.

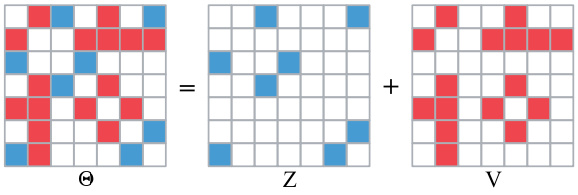


Figure 1: Illustration of a symmetric matrix $\mathbf{\Theta}$ decomposed into $\mathbf{Z} + \mathbf{V}$ for CNHD, where $\mathbf{Z}$ is sparse and $\mathbf{V}$ is a matrix whose columns are either entirely zeros (w.r.t., non-hub regions) or not entirely zeros (w.r.t., hub regions). White represents zero entry, blue denotes non-zero element, and red denotes non-zero element due to two hubs (red columns).

Optimization Algorithm It is difficult to jointly optimize all variables in Eq. (3) due to its non-convex nature. Here, we propose to solve the above problem using an alternating direction method of multipliers (ADMM) algorithm. ADMM is an attractive algorithm for solving this problem, because it allows us to decouple the terms in Eq. (3) that are difficult to be optimized directly.

To facilitate the optimization of Eq. (3), we first introduce some relaxation variables. Specifically, we denote $\|\mathbf{V}\|_1$ as $\mathbf{V}_1$, $\|\mathbf{V}\|_{2,1}$ as $\mathbf{V}_2$, $\mathbf{V}^T$ as $\mathbf{V}_3$, $\|\mathbf{Z}\|_1$ as $\mathbf{Z}_1$, and $\mathbf{Z}^T$ as $\mathbf{Z}_2$. Hence the constraints in Eq. (3) can be rewritten as:

$$\begin{aligned}&\mathbf{\Theta} = \mathbf{Z} + \mathbf{V}, \mathbf{V} = \mathbf{V}^T = \mathbf{V}_3, \mathbf{Z} = \mathbf{Z}^T = \mathbf{Z}_2, \\ &\mathbf{Z} = \mathbf{Z}_1, \mathbf{V} = \mathbf{V}_1, \mathbf{V} = \mathbf{V}_2, diag(\mathbf{\Theta}) = \mathbf{0}.\end{aligned} \tag{4}$$

Then we can solve the Eq. (3) by minimizing the follow-

ing Augmented Lagrangian function $\mathcal{L}$, which is defined as:

$$\begin{aligned}\mathcal{L} &= \ell(\mathbf{\Theta}) + \lambda\|\mathbf{Z}_1\|_1 + \beta\|\mathbf{V}_1\|_1 + \gamma\|\mathbf{V}_2\|_{2,1} \\ &+ tr[\mathbf{U}_1^T(\mathbf{Z}-\mathbf{Z}_1)] + tr[\mathbf{U}_2^T(\mathbf{V}-\mathbf{V}_1)] \\ &+ tr[\mathbf{U}_3^T(\mathbf{V}-\mathbf{V}_2)] + tr[\mathbf{U}_4^T(\mathbf{V}^T-\mathbf{V}_3)] \\ &+ tr[\mathbf{U}_5^T(\mathbf{Z}^T-\mathbf{Z}_2)] + tr[\mathbf{U}_6^T(\mathbf{\Theta}-\mathbf{V}-\mathbf{Z})] \\ &+ tr[\mathbf{U}_7^T(\mathbf{V}-\mathbf{V}_3)] + tr[\mathbf{U}_8^T(\mathbf{Z}-\mathbf{Z}_2)] \\ &+ \frac{\mu}{2}(\|\mathbf{Z}-\mathbf{Z}_1\|_F^2 + \|\mathbf{V}-\mathbf{V}_1\|_F^2 + \|\mathbf{V}-\mathbf{V}_2\|_F^2 \\ &+ \|\mathbf{V}^T-\mathbf{V}_3\|_F^2 + \|\mathbf{Z}^T-\mathbf{Z}_2\|_F^2 + \|\mathbf{\Theta}-\mathbf{V}-\mathbf{Z}\|_F^2 \\ &+ \|\mathbf{V}-\mathbf{V}_3\|_F^2 + \|\mathbf{Z}-\mathbf{Z}_2\|_F^2),\end{aligned} \tag{5}$$

where $\mathbf{U}_i$ $(i \in [1, 2, \cdots, 8])$ is Lagrange multiplier and $\mu > 0$ is a penalty parameter. $\|\cdot\|_F$ denotes the matrix Frobenius norm. To find a minimal point for $\mathcal{L}$, we alternatively update for each of variable, while keeping the other variables fixed. We define the t-th iteration optimized variables as $\mathbf{\Theta}_t, \mathbf{Z}_t, \mathbf{V}_t, \mathbf{Z}_{1,t}, \mathbf{Z}_{2,t}, \mathbf{V}_{1,t}, \mathbf{V}_{2,t}, \mathbf{V}_{3,t}$ and b_t. Then, we can achieve the sub-solution at the $t+1$ $(t \geq 0)$-th iteration as:

Update $\mathbf{\Theta}$:

$$\mathbf{\Theta}_{t+1} = \frac{\partial \ell}{\partial \mathbf{\Theta}} + \mathbf{U}_6 + \frac{\mu}{2}[2\mathbf{\Theta} - 2(\mathbf{V}_t + \mathbf{Z}_t)], \tag{6}$$

where $\frac{\partial \ell}{\partial \mathbf{\Theta}} = \frac{1}{N}\sum_{n=1}^{N} \frac{\exp[-Y_n(tr(\mathbf{\Theta}^T\mathbf{A}^{(n)})+b_t)]\cdot[-Y_n\mathbf{A}^{(n)}]}{1+\exp[-Y_n(tr(\mathbf{\Theta}^T\mathbf{A}^{(n)})+b_t)]}$. Then the minimum value of $\mathbf{\Theta}_{t+1}$ can be computed by gradient descent.

Update Z:

$$\mathbf{Z}_{t+1} = \frac{1}{4\mu}(\mu\mathbf{G}_1 + \mathbf{G}_2), \tag{7}$$

where $\mathbf{G}_1 = \mathbf{Z}_{1,t} + \mathbf{Z}_{2,t}^T + \mathbf{\Theta}_t - \mathbf{V}_t + \mathbf{Z}_{2,t}$ and $\mathbf{G}_2 = \mathbf{U}_6 - \mathbf{U}_1 - \mathbf{U}_5^T - \mathbf{U}_8$.

Update V:

$$\mathbf{V}_{t+1} = \frac{1}{5\mu}[\mu\mathbf{G}_3 + \mathbf{G}_4], \tag{8}$$

where $\mathbf{G}_3 = \mathbf{V}_{1,t} + \mathbf{V}_{2,t} + \mathbf{V}_{3,t}^T + \mathbf{\Theta}_t - \mathbf{Z}_t + \mathbf{V}_{3,t}$ and $\mathbf{G}_4 = \mathbf{U}_6 - \mathbf{U}_2 - \mathbf{U}_3 - \mathbf{U}_4^T - \mathbf{U}_7$.

Update $\mathbf{Z}_1$:

$$\mathbf{Z}_{1,t+1} = \arg\min_{\mathbf{Z}_1} \frac{\lambda}{\mu}\|\mathbf{Z}_1\|_1 + \frac{1}{2}\|\mathbf{Z}_1 - (\mathbf{Z}_t + \frac{\mathbf{U}_1}{\mu})\|_F^2. \tag{9}$$

According to the shrinkage operator (Yang et al. 2009), the above problem has the following closed form solution:

$$\mathbf{Z}_{1,t+1} = \mathcal{S}_{\frac{\lambda}{\mu}}(\mathbf{Z}_t + \frac{\mathbf{U}_1}{\mu}), \tag{10}$$

where $\mathcal{S}_{(\eta)}[x] = sign(x)\max(|x| - \eta, 0)$.

Update $\mathbf{Z}_2$:

$$\mathbf{Z}_{2,t+1} = \frac{1}{2\mu}[\mathbf{U}_5 + \mu\mathbf{Z}_t^T + \mathbf{U}_8 + \mu\mathbf{Z}_t]. \tag{11}$$

Update $\mathbf{V}_1$:

$$\mathbf{V}_{1,t+1} = \arg\min_{\mathbf{V}_1} \frac{\beta}{\mu}\|\mathbf{V}_1\|_1 + \frac{1}{2}\|\mathbf{V}_1 - (\mathbf{V}_t + \frac{\mathbf{U}_2}{\mu})\|_F^2. \tag{12}$$

Problem (12) can also be solved by using the shrinkage operator (Yang et al. 2009) in the same way as problem (10).

Update $\mathbf{V}_2$:

$$\mathbf{V}_{2,t+1} = \arg\min_{\mathbf{V}_2} \frac{\gamma}{\mu}\|\mathbf{V}_2\|_{2,1} + \frac{1}{2}\|\mathbf{V}_2 - (\mathbf{V}_t + \frac{\mathbf{U}_3}{\mu})\|_F^2, \tag{13}$$

which is easily addressed with (Yuan and Lin 2006) and the closed form solution can be rewritten as:

$$[\mathbf{V}_{2,t+1}]_{:,i} = \begin{cases} \frac{\|\mathbf{Q}_{:,i}\|_2 - \frac{\gamma}{\mu}}{\|\mathbf{Q}_{:,i}\|_2}\mathbf{Q}_{:,i}, & if\ \|\mathbf{Q}_{:,i}\|_2 > \frac{\gamma}{\mu}; \\ 0, & otherwise. \end{cases} \tag{14}$$

where $\mathbf{Q} = \mathbf{V}_t + \frac{\mathbf{U}_3}{\mu}$ and $[\mathbf{V}_{2,t+1}]_{:,i}$ is the i-th column of $\mathbf{V}_{2,t+1}$.

Update $\mathbf{V}_3$:

$$\mathbf{V}_{3,t+1} = \frac{1}{2\mu}[\mathbf{U}_4 + \mathbf{U}_7 + \mu\mathbf{V}_t^T + \mu\mathbf{V}_t]. \tag{15}$$

Update b:

$$\frac{\partial \ell}{\partial b} = \frac{1}{N}\sum_{n=1}^{N} \frac{\exp[-Y_n(tr(\mathbf{\Theta}_t^T\mathbf{A}^{(n)}) + b)]\cdot[-Y_n]}{1+\exp[-Y_n(tr(\mathbf{\Theta}_t^T\mathbf{A}^{(n)}) + b)]}. \tag{16}$$

Then it is easy to compute b_{t+1} via gradient descent.

Update multipliers: Multipliers $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4, \mathbf{U}_5, \mathbf{U}_6, \mathbf{U}_7, \mathbf{U}_8$ and parameter μ are updated by using (17):

$$\begin{cases} \mathbf{U}_1 = \mathbf{U}_1 + \mu(\mathbf{Z}_t - \mathbf{Z}_{1,t}) \\ \mathbf{U}_2 = \mathbf{U}_2 + \mu(\mathbf{V}_t - \mathbf{V}_{1,t}) \\ \mathbf{U}_3 = \mathbf{U}_3 + \mu(\mathbf{V}_t - \mathbf{V}_{2,t}) \\ \mathbf{U}_4 = \mathbf{U}_4 + \mu(\mathbf{V}_t^T - \mathbf{V}_{3,t}) \\ \mathbf{U}_5 = \mathbf{U}_5 + \mu(\mathbf{Z}_t^T - \mathbf{Z}_{2,t}) \\ \mathbf{U}_6 = \mathbf{U}_6 + \mu(\mathbf{\Theta}_t - \mathbf{V}_t - \mathbf{Z}_t) \\ \mathbf{U}_7 = \mathbf{U}_7 + \mu(\mathbf{V}_t - \mathbf{V}_{3,t}) \\ \mathbf{U}_8 = \mathbf{U}_8 + \mu(\mathbf{Z}_t - \mathbf{Z}_{2,t}) \\ \mu = min(\mu\rho, \mu_{max}) \end{cases} \tag{17}$$

The overall algorithm is outlined in Algorithm 1. Those parameters μ, ρ, μ_{max}, and ε are set empirically, while other balanced parameters α, β and γ are tuned in the experiment.

Computational Complexity The major computational complexity of Algorithm 1 lies in Step 2 and Step 10 that contain gradient derivation. We assume the number of iteration for ADMM and gradient descent is τ, and the computational complexity of Step 2 and Step 10 is $\tau N\mathcal{O}(M^3)$. Thus, the overall computational complexity of Algorithm 1 is $\tau^2 N\mathcal{O}(M^3)$, where M is usually a fixed number (e.g., 90) of brain regions. This indicates that the time cost of the proposed optimization algorithm will increase with the increase of the number of samples and optimization iterations.

Algorithm 1 Solving Problem (5) by ADMM

Require: $\mathbf{A}, \mathbf{Y}, \alpha, \beta$ and γ;
init: $\mathbf{\Theta} = \mathbf{0}, b = 0, \mathbf{Z} = \mathbf{0}, \mathbf{V} = \mathbf{0}, \mathbf{Z}_1 = \mathbf{0}, \mathbf{Z}_2 = \mathbf{0}, \mathbf{V}_1 = \mathbf{0}, \mathbf{V}_2 = \mathbf{0}, \mathbf{V}_3 = \mathbf{0}, \mathbf{U}_1 = \mathbf{0}, \mathbf{U}_2 = \mathbf{0}, \mathbf{U}_3 = \mathbf{0}, \mathbf{U}_4 = \mathbf{0}, \mathbf{U}_5 = \mathbf{0}, \mathbf{U}_6 = \mathbf{0}, \mathbf{U}_7 = \mathbf{0}, \mathbf{U}_8 = \mathbf{0}, \mu = 10^{-5}, \rho = 1.1, \mu_{max} = 10^7, \varepsilon = 10^{-7}$;
while not converged **do**
  Fix the other variables and update $\mathbf{\Theta}$ by solving (6);
  Fix the other variables and update $\mathbf{Z}$ by solving (7);
  Fix others variables and update $\mathbf{V}$ by solving (8);
  Fix the other variables and update $\mathbf{Z}_1$ by solving (9);
  Fix the other variables and update $\mathbf{Z}_2$ by solving (11);
  Fix the other variables and update $\mathbf{V}_1$ by solving (12);
  Fix the other variables and update $\mathbf{V}_2$ by solving (13);
  Fix the other variables and update $\mathbf{V}_3$ by solving (15);
  Fix the other variables and update b by solving (16);
  Update the multipliers and parameters by (17);
  Check the convergence conditions
  $\|\mathbf{Z} - \mathbf{Z}_1\|_\infty < \epsilon;\ \|\mathbf{V} - \mathbf{V}_1\|_\infty < \epsilon;$
  $\|\mathbf{V} - \mathbf{V}_2\|_\infty < \epsilon;\ \|\mathbf{V}^T - \mathbf{V}_3\|_\infty < \epsilon;$
  $\|\mathbf{Z}^T - \mathbf{Z}_2\|_\infty < \epsilon;\ \|\mathbf{\Theta} - (\mathbf{V} + \mathbf{Z})\|_\infty < \epsilon;$
  $\|\mathbf{V} - \mathbf{V}_3\|_\infty < \epsilon;\ \|\mathbf{Z} - \mathbf{Z}_2\|_\infty < \epsilon.$
end while
Ensure: $\mathbf{\Theta}, b, \mathbf{Z}, \mathbf{V}$

Experiments

Datasets To evaluate the effectiveness of our CNHD method, we perform experiments on three real schizophrenia (SZ) datasets with fMRI scans, collected from the Affiliated Nanjing Brain Hospital of Nanjing Medical University (NBH) containing 24 SZ patients and 21 normal controls (NCs), the Center for Biomedical Research Excellence (COBRE) having 67 SZ patients and 53 NCs[1], and the National Taiwan University Hospital (Taiwan) including 69 SZ patients and 62 NCs. We first preprocess fMRI data using Statistical Parametric Mapping software[2] (SPM8), and detailed procedures follow previously studies (Wang et al. 2017; Li et al. 2017). We then partition the whole brain into 90 pre-defined ROIs based on the anatomical automatic labeling (AAL) template[3]. For each ROI, we compute its mean rs-fMRI time series by averaging the gray matter (GM) masked blood oxygen-level dependent (BOLD) signals among all voxels within this ROI. Finally, we compute Pearson correlation coefficients to build functional connectivity between a pair of ROIs. Thus, a functional connectivity network is generated for each subject, with each node corresponding to a particular ROI and each edge denoting the correlation between a pair of ROIs. Given $M = 90$ ROIs, we can obtain a 90×90 functional connectivity network for each subject.

Competing Methods We first compare the proposed CNHD model with a baseline approach, i.e., Lasso (Tibshirani 2011). Besides, we also compare CNHD with six state-of-the-art methods, including 4 topology-based representation approaches (i.e., CC (Wee et al. 2012), Average Degree (Dai and He 2014), WL Graph Kernel (Shervashidze et al. 2011), and t-BNE (Cao et al. 2017)), and 2 subgraph-based representation approaches (i.e., Graph Boosting (Kudo, Maeda, and Matsumoto 2005), and Ordinal Pattern (Zhang et al. 2018)).

- **Lasso:** The *Lasso* (Tibshirani 2011) method first selects a discriminative subset of features from the vectorized representation of a brain network. Thus, each network is converted into a feature vector in this method.
- **Clustering Coefficient (CC):** The *CC* (Wee et al. 2012) method extracts local clustering coefficients as features of brain network, by measuring the degree to which nodes in a network tend to cluster together. Similar to Lasso, each network is converted into a feature vector in this method.
- **Average Degree (Avg-DE):** The *Avg-DE* (Dai and He 2014) method reflects the average number of connections any network node may have, which measures the overall level of brain network connectivity. With Avg-DE, each network can be represented as a scalar.
- **WL Graph Kernel (WLGK):** The *WLGK* (Shervashidze et al. 2011) method measures the similarity of networks based on graph isomorphism testing. That is, each network is represented by a feature vector corresponding to the similarity between this network with the other ones.
- **Tensor-based Brain Network Embedding (t-BNE):** The *t-BNE* (Cao et al. 2017) method embeds each network to a vector via tensor factorization, and the new vectorized features are used for classification.
- **Graph Boosting (GB):** The *GB* (Kudo, Maeda, and Matsumoto 2005) method first mines subgraphs from each network as features, and then learns subgraph-based decision stumps as weak learners. Finally, a Boosting algorithm is used for network classification.
- **Ordinal Pattern (OP):** The *OP* (Zhang et al. 2018) method first mines informative subgraphs via ordinal patterns that frequently appear in network sets, and then extracts features based on the selected ordinal patterns for representing the original network.

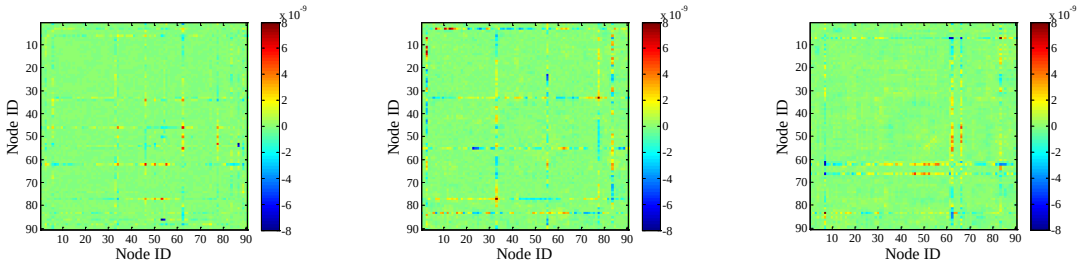
Since five methods (i.e., Lasso, CC, Avg-DE, WLGK and OP) can only be used for extracting network representation, we employ the linear SVM as the base classifier in these methods for disease classification. Here, we resort to the LIBSVM toolbox[4] with a default penalty parameter (i.e., $C = 1$). Similar to CNHD, GB and t-BNE can simultaneously perform feature extraction and classifier training.

Evaluation Metric We employ seven metrics to evaluate the performance of different methods, including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic (ROC) curve (AUC). Denote TP, TN, FP and FN as True Positive, True Negative,

[1] http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html
[2] https://www.fil.ion.ucl.ac.uk/spm/software/spm8/
[3] http://www.gin.cnrs.fr/en/tools/aal-aal2/
[4] https://www.csie.ntu.edu.tw/∼cjlin/libsvm/

Table 1: Network-based disease classification results achieved by eight different methods on three datasets.

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)	BAC (%)	PPV (%)	NPV (%)
NBH	Lasso	72.88 ± 0.59	67.50 ± 1.12	80.00 ± 2.09	80.50 ± 1.61	73.75 ± 0.69	82.00 ± 1.75	69.17 ± 0.56
	CC	62.69 ± 1.95	77.50 ± 2.56	48.00 ± 1.82	72.00 ± 1.77	62.75 ± 1.94	61.33 ± 1.26	69.05 ± 3.13
	Avg-DE	58.85 ± 0.85	56.25 ± 1.77	68.75 ± 3.26	71.75 ± 1.60	60.00 ± 1.05	73.33 ± 2.53	54.54 ± 0.44
	WLGK	61.35 ± 1.13	55.00 ± 1.12	70.00 ± 3.26	72.25 ± 1.23	62.50 ± 1.25	73.33 ± 2.53	56.44 ± 0.71
	t-BNE	81.25 ± 2.40	87.50 ± 2.50	75.00 ± 2.89	65.94 ± 2.19	81.25 ± 2.39	79.17 ± 2.50	87.50 ± 2.50
	GB	75.00 ± 1.25	68.33 ± 1.61	75.00 ± 4.33	68.33 ± 2.32	71.67 ± 1.77	80.00 ± 3.46	71.11 ± 0.77
	OP	83.66 ± 0.59	78.13 ± 2.14	88.75 ± 1.32	94.37 ± 0.66	83.44 ± 0.59	$\mathbf{91.87 \pm 0.99}$	81.67 ± 1.37
	CNHD	$\mathbf{90.96 \pm 1.04}$	$\mathbf{92.50 \pm 1.12}$	$\mathbf{90.00 \pm 1.37}$	$\mathbf{98.25 \pm 0.27}$	$\mathbf{91.25 \pm 1.05}$	91.00 ± 1.25	$\mathbf{91.67 \pm 1.18}$
COBRE	Lasso	65.03 ± 0.81	63.23 ± 1.76	66.77 ± 2.32	76.97 ± 0.34	65.00 ± 0.67	63.25 ± 1.39	71.03 ± 0.72
	CC	62.02 ± 0.90	57.54 ± 1.71	59.44 ± 0.70	63.79 ± 1.20	58.49 ± 0.94	63.33 ± 2.47	61.34 ± 0.65
	Avg-DE	59.25 ± 1.73	43.69 ± 2.31	71.59 ± 2.90	70.51 ± 1.73	57.64 ± 1.64	63.78 ± 2.63	60.15 ± 1.67
	WLGK	60.68 ± 1.59	48.77 ± 2.18	70.05 ± 2.83	72.31 ± 1.74	59.41 ± 1.50	65.37 ± 2.42	61.99 ± 1.63
	t-BNE	69.91 ± 1.22	75.38 ± 0.92	65.51 ± 1.58	53.53 ± 2.19	70.45 ± 1.19	64.55 ± 1.24	$\mathbf{76.19 \pm 1.21}$
	GB	71.68 ± 0.22	$\mathbf{77.54 \pm 0.56}$	64.31 ± 0.52	72.82 ± 0.33	70.92 ± 0.20	$\mathbf{73.45 \pm 0.26}$	69.61 ± 0.46
	OP	69.78 ± 0.46	61.38 ± 1.39	76.21 ± 0.96	68.38 ± 0.64	68.79 ± 0.51	67.62 ± 0.63	72.15 ± 0.50
	CNHD	$\mathbf{73.11 \pm 0.38}$	64.31 ± 0.52	$\mathbf{80.21 \pm 0.91}$	$\mathbf{78.67 \pm 0.65}$	$\mathbf{72.26 \pm 0.35}$	72.84 ± 0.99	74.07 ± 0.20
Taiwan	Lasso	67.46 ± 0.48	70.59 ± 1.18	64.29 ± 0.58	72.61 ± 0.40	67.44 ± 0.44	68.59 ± 0.26	67.27 ± 0.85
	CC	66.81 ± 0.71	75.70 ± 1.15	55.71 ± 2.18	72.95 ± 1.00	65.71 ± 0.80	65.25 ± 0.96	76.00 ± 1.57
	Avg-DE	62.01 ± 0.75	61.72 ± 1.52	62.62 ± 0.85	71.90 ± 0.46	62.17 ± 0.72	64.36 ± 0.63	60.71 ± 0.95
	WLGK	61.37 ± 0.70	63.62 ± 1.57	59.29 ± 1.13	70.67 ± 0.57	61.45 ± 0.66	63.37 ± 0.49	60.64 ± 1.05
	t-BNE	69.24 ± 0.40	71.04 ± 0.64	67.38 ± 0.93	78.65 ± 0.23	69.21 ± 0.41	71.00 ± 0.62	67.90 ± 0.36
	GB	72.28 ± 0.44	78.01 ± 1.24	65.95 ± 0.76	80.01 ± 0.31	71.98 ± 0.41	71.81 ± 0.37	74.63 ± 0.94
	OP	75.64 ± 0.29	78.73 ± 0.35	70.95 ± 0.42	82.17 ± 0.49	74.66 ± 0.18	75.02 ± 0.25	74.83 ± 0.33
	CNHD	$\mathbf{80.44 \pm 0.57}$	$\mathbf{84.52 \pm 0.56}$	$\mathbf{75.95 \pm 1.11}$	$\mathbf{84.80 \pm 0.58}$	$\mathbf{80.24 \pm 0.59}$	$\mathbf{80.22 \pm 0.77}$	$\mathbf{81.65 \pm 0.65}$



(a) NBH (b) COBRE (c) Taiwan

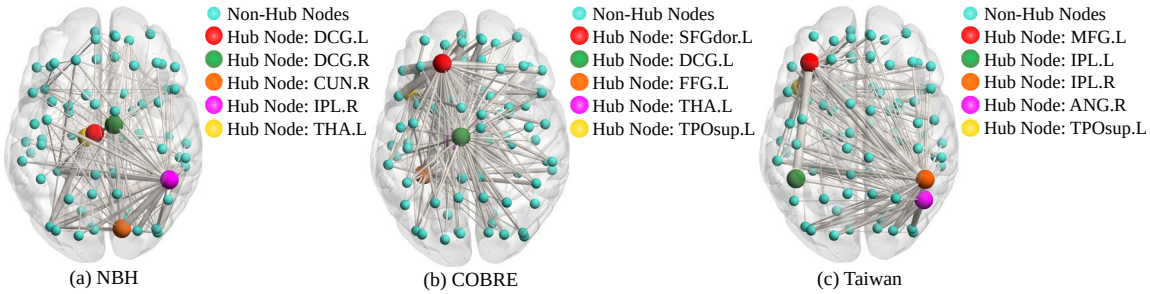
Figure 2: The weight maps of network predictor learned by CNHD on three datasets, (a) NBH, (b) COBRE, and (c) Taiwan.

False Positive, and False Negative, respectively. These metrics is defined as follows: $ACC = (TP + TN)/(TP + TN + FP + FN)$, $SEN = TP/(TP + FN)$, $SPE = TN/(TN + FP)$, $BAC = (SEN + SPE)/2$, $PPV = TP/(TP + FP)$, and $NPV = TN/(TN + FN)$. For all these metrics, higher values indicate better performance.

Experimental Settings We employ a cross-validation (CV) strategy to evaluate the performance of different methods. Specifically, we first randomly divide the data into 5 folds. Then, we alternatively choose 4 folds as the training set and the rest as the testing set, until all 5 folds have been used as the testing set. The mean and standard deviation of classification results are recorded for each method. To select optimal parameters of our approach and all competing methods, we further perform an inner 5-fold CV on the training data. That is, in each fold of 5-fold CV, we find the optimal parameters for each method via cross-validation on the training subset. Note that no testing data is used in such inner CV process. Parameters in six competing methods are set as suggested in their papers. For the proposed CNHD method, the parameters λ, β and γ are selected from $\{10^{-4}, 10^{-3}, \cdots, 10^{2}\}$ via inner CV.

Experimental Results and Analysis

Results of Brain Disease Classification Table 1 reports the disease classification results achieved by eight different methods on three datasets. From Table 1, one could have the following observations. *First*, in terms of the AUC value, our CNHD method consistently outperforms seven competing methods. For example, CNHD produces an AUC of 98.25% on NBH dataset, while the second-best result is only 94.37% (achieved by OP). This suggests that CNHD has a stronger discriminative power for network-based disease identification than baseline and state-of-the-art methods. *Second*, our CNHD method is superior to four topology-based methods



(a) NBH (b) COBRE (c) Taiwan

Figure 3: Visual plots of hub structure identified by CNHD from brain connectivity networks on three datasets, (a) NBH, (b) COBRE, and (c) Taiwan. Here, each node represents a brain region, and each edge represents the connections between a pair of brain regions. Also, the thickness of an edge represents the strength of a specific connection.

(i.e., CC, Avg-DE, WLGK and t-BNE) regarding seven evaluation metrics. It implies that global structure information conveyed in hubs (as explored in CNHD) helps promote the learning performance, compared with local structure information (e.g., clustering coefficient) captured by topology-based methods. *Third*, subgraph-based methods (i.e., GB and OP) generally obtain better performance than the baseline and topology-based methods, but its overall performance is worse than CNHD. This may be due to those subgraphs mined by GB and OP could cover the essential hub structure that is discriminative for network identification, while such hub structure is not employed in topology-based methods. In particular, although the OP method utilizes both connection strength and ordinal relationship among edges, its feature extraction is still independent of the classifier learning. This could partly explain why our CNHD method outperforms OP in most cases.

Analysis of Detected Network Hubs We investigate the identified hub nodes (i.e., non-zero columns in $\mathbf{V}$) from functional connectivity networks of all training subjects in three datasets. Since a 5-fold CV strategy is used in the experiments, the identified hub nodes may be different in different folds. Hence, we simply show the hub nodes that are detected from one fold. On three datasets, we plot the matrix $\mathbf{V}$ that contains hub structures learned by the proposed CNHD method in Figure 2. As can be seen from Figure 2, there are approximately 5 hubs detected from each of three datasets. Also, Figure 2 shows that connections within each hub node (i.e., a column) are densely-connected, while connections between the remaining nodes are sparse.

We further visualize the hub structure identified by our method in Figure 3, using the Brain Net Viewer toolbox[5]. In Figure 3, each node represents a brain region (i.e., ROI), and each edge denotes the connections between a pair of ROIs. Also, nodes with turquoise color refer to the non-hub brain regions, while those with other colors correspond to hub nodes detected by our CNHD method. Figure 3 indicates that four hub nodes are consistently detected from three datasets, including *DCG*, *THA*, *IPL*, and *TPOsup*. These hub nodes have been reported in previous studies (Heuvel, Sporns, and Olaf 2013; Mikail and Ed 2013), suggesting that our method can produce reliable results in hub detection from functional brain connectivity networks.

[5]https://www.nitrc.org/projects/bnv/

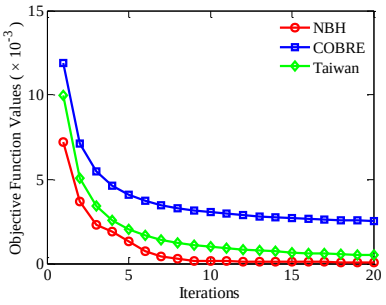


Figure 4: Objective function value vs. iteration number for optimization, achieved by our CNHD model on three datasets: (a) NBH, (b) COBRE, and (c) Taiwan.

Convergence Evaluation We now study the efficacy of the proposed optimization algorithm for solving the objective function in CNHD. Figure 4 shows the objective function values (i.e., $\mathcal{L}$ in Eq. (5)) with respect to the number of iterations on different datasets. Figure 4 shows that the objective function value decreases steadily with more iterations. Importantly, the proposed algorithm can converge within 20 iterations, indicating its fast convergence rate.

Conclusion

In this paper, we present a Connectivity Network analysis method with discriminative Hub Detection named CNHD for brain network analysis. We incorporate the feature extraction of brain networks, network-based classifier training into a unified framework so that the two tasks could benefit each other. Also, CNHD can automatically identify discriminative hub structures from functional connectivity network. We evaluated our method on three real fMRI datasets, with experimental results suggest the superiority of our method in disease diagnosis and network hub identification.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos: 61876082, 61861130366,

61703301, 61473149) and the University Science and Technology Project of Shandong Province (No. J17KA086).

References

Bolton, T. A.; Tarun, A.; Sterpenich, V.; Schwartz, S.; and Ville, D. V. D. 2017. Interactions between large-scale functional brain networks are captured by sparse coupled HMMs. *IEEE Trans. on Medical Imaging* 37(1):230–240.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1):1–122.

Cao, B.; Kong, X.; Zhang, J.; Yu, P. S.; and Ragin, A. B. 2016. Mining brain networks using multiple side views for neurological disorder identification. In *ICDM*, 709–714.

Cao, B.; He, L.; Xiaokai; Xing, M.; Yu, P. S.; Klumpp, H.; and Leow, A. D. 2017. t-BNE: Tensor-based brain network embedding. In *SDM*, 189–197.

Dai, Z., and He, Y. 2014. Disrupted structural and functional brain connectomes in mild cognitive impairment and Alzheimer's disease. *Neuroscience Bulletin* 30(2):217–232.

Dai, Z. J.; Bi, Y. C.; and He, Y. 2015. With great brain hub connectivity comes great vulnerability. *CNS Neuroscience and Therapeutics* 21(7):541–542.

Fox, M. D., and Greicius, M. 2010. Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience* 4(19):1–13.

Gong, P.; Ye, J.; and Zhang, C. 2012. Robust multi-task feature learning. In *KDD*, 895–903.

He, L.; Lu, C. T.; Ma, J.; Cao, J.; Shen, L.; and Yu, P. S. 2016. Joint community and structural hole spanner detection via harmonic modularity. In *KDD*, 875–884.

Heuvel, M. P. V. D.; Sporns; and Olaf. 2013. Network hubs in the human brain. *Trends in Cognitive Sciences* 17(12):683–696.

Huang, S.; Li, J.; Ye, J.; Fleisher, A.; Chen, K.; Wu, T.; and Reiman, E. 2011. Brain effective connectivity modeling for Alzheimer's disease by sparse gaussian bayesian network. In *KDD*, 931–939.

Jie, B.; Zhang, D.; Gao, W.; Wang, Q.; Wee, C. Y.; and Shen, D. 2014. Integration of network topological and connectivity properties for neuroimaging classification. *IEEE Trans. on Biomedical Engineering* 61(2):576–589.

Kong, X.; Yu, P. S.; Wang, X.; and Ragin, A. B. 2013. Discriminative feature selection for uncertain graph classification. In *SDM*, 82–93.

Kudo, T.; Maeda, E.; and Matsumoto, Y. 2005. An application of boosting to graph classification. In *NIPS*, 729–736.

Li, T.; Wang, Q.; Zhang, J.; Rolls, E. T.; Yang, W.; Palaniyappan, L.; Zhang, L.; Cheng, W.; Yao, Y.; and Liu, Z. 2017. Brain-wide analysis of functional connectivity in first-episode and chronic stages of schizophrenia. *Schizophrenia Bulletin* 43(2):436–448.

Liu, Y.; He, L.; Cao, B.; Yu, P. S.; Ragin, A. B.; and Leow, A. D. 2018. Multi-view multi-graph embedding for brain network clustering analysis. In *AAAI*, 117–124.

Liu, X.; Kong, X.; and Philip, S. Y. 2017. Collective discovery of brain networks with unknown groups. In *IJCNN*, 3569–3576.

Liu, X.; Kong, X.; and Ragin, A. B. 2017. Unified and contrasting graphical lasso for brain network discovery. In *SDM*, 180–188.

Ma, G.; Lu, C.; He, L.; Yu, P. S.; and Ragin, A. B. 2017. Multi-view graph embedding with hub detection for brain network analysis. In *ICDM*, 967–972.

Mikail, R., and Ed, B. 2013. Schizophrenia and abnormal brain network hubs. *Dialogues in Clinical Neuroscience* 15(3):339–349.

Power, J. D.; Fair, D. A.; Schlaggar, B. L.; and Petersen, S. E. 2010. The development of human functional brain networks. *Neuron* 67(5):735–748.

Ruan, Y., and Parthasarathy, S. 2014. Simultaneous detection of communities and roles from large networks. In *ACM Conference on Online Social Networks*, 203–214.

Shervashidze, N.; Schweitzer, P.; Leeuwen, E. J. V.; Mehlhorn, K.; and Borgwardt, K. M. 2011. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research* 12(3):2539–2561.

Sporns, O. 2012. From simple graphs to the connectome: Networks in neuroimaging. *NeuroImage* 62(2):881–886.

Stam, C. J. 2014. Modern network science of neurological disorders. *Nature Reviews Neuroscience* 15(10):683–695.

Tan, K. M.; London, P.; Mohan, K.; Lee, S. I.; Fazel, M.; and Witten, D. 2014. Learning graphical models with hubs. *Journal of Machine Learning Research* 15:3297–3331.

Tibshirani, R. 2011. Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society* 73(3):273–282.

Wang, M.; Hao, X.; Huang, J.; Wang, K.; Xu, X.; and Zhang, D. 2017. Multi-level multi-task structured sparse learning for diagnosis of schizophrenia disease. In *MICCAI*, 46–54.

Wee, C. Y.; Yap, P. T.; Zhang, D.; Denny, K.; Browndyke, J. N.; Potter, G. G.; Welshbohmer, K. A.; Wang, L.; and Shen, D. 2012. Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage* 59(3):2045–2056.

Yang, J.; Yin, W.; Zhang, Y.; and Wang, Y. 2009. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM Journal on Imaging Sciences* 2(2):569–592.

Yang, S.; Sun, Q.; Ji, S.; Wonka, P.; Davidson, I.; and Ye, J. 2015. Structural graphical lasso for learning mouse brain connectivity. In *KDD*, 1385–1394.

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

Zhang, D.; Huang, J.; Jie, B.; Du, J.; Tu, L.; and Liu, M. 2018. Ordinal pattern: A new descriptor for brain connectivity networks. *IEEE Trans. on Medical Imaging* 37(7):1711–1722.