Structural Deep Brain Network Mining

Shen Wang
University of Illinois at Chicago
Chicago, USA
swang224@uic.edu

Lifang He*
Shenzhen University
Shenzhen, China
lifanghescut@gmail.com

Bokai Cao
University of Illinois at Chicago
Chicago, USA
caobokai@uic.edu

Chun-Ta Lu
University of Illinois at Chicago
Chicago, USA
clu29@uic.edu

Philip S. Yu
University of Illinois at Chicago
Chicago, USA
Tsinghua University
Beijing, China
psyu@cs.uic.edu

Ann B. Ragin
Northwestern University
Chicago, USA
ann-ragin@northwestern.edu

ABSTRACT

Mining from neuroimaging data is becoming increasingly popular in the field of healthcare and bioinformatics, due to its potential to discover clinically meaningful structure patterns that could facilitate the understanding and diagnosis of neurological and neuropsychiatric disorders. Most recent research concentrates on applying subgraph mining techniques to discover connected subgraph patterns in the brain network. However, the underlying brain network structure is complicated. As a shallow linear model, subgraph mining cannot capture the highly non-linear structures, resulting in sub-optimal patterns. Therefore, how to learn representations that can capture the highly non-linearity of brain networks and preserve the underlying structures is a critical problem.

In this paper, we propose a Structural Deep Brain Network mining method, namely SDBN, to learn highly non-linear and structure-preserving representations of brain networks. Specifically, we first introduce a novel graph reordering approach based on module identification, which rearranges the order of the nodes to preserve the modular structure of the graph. Next, we perform structural augmentation to further enhance the spatial information of the reordered graph. Then we propose a deep feature learning framework for combining supervised learning and unsupervised learning in a small-scale setting, by augmenting Convolutional Neural Network (CNN) with decoding pathways for reconstruction. With the help of the multiple layers of non-linear mapping, the proposed SDBN approach can capture the highly non-linear structure of brain networks. Further, it has better generalization capability for high-dimensional brain networks and works well even for small sample learning. Benefit from CNN's task-oriented learning style, the learned hierarchical representation is meaningful for the clinical task. To evaluate the proposed SDBN method, we conduct extensive experiments on four real brain network datasets for disease diagnoses. The experiment results show that SDBN can capture discriminative and meaningful structural graph representations for brain disorder diagnosis.

CCS CONCEPTS

•Information systems →Data mining; •Computing methodologies →Machine learning;

KEYWORDS

brain network, deep learning, graph reordering

1 INTRODUCTION

In many neurological and neuropsychiatric disorders, brain involvement, including irreversible loss of brain tissue and deterioration in cognitive function [19], has deleterious consequences for judgment and function. Early detection in subclinical periods is critical for aiding clinical diagnosis, clarifying underlying mechanisms and informing neuroprotective interventions to slow or reverse neural injury for a broad spectrum of brain disorders, e.g., HIV infection, attention-deficit/hyperactivity disorder (ADHD), and bipolar disorder (BP). The last several decades have witnessed rapid progress in noninvasive neuroimaging techniques, e.g., functional resting-state magnetic resonance imaging (fMRI) and structural diffusion tensor imaging (DTI), from which an underlying brain connectivity network (a.k.a, connectome [42]) can be interrogated. New capacilities for investigating the vast patterns of connectivity involving billions of neurons, promise to yield new insights concerning the normal brain and many disorders. Brain connectomes are often represented as graphs, where nodes represent brain regions and edges represent connections between pairs of regions.

Learning discriminative and clinically meaningful representations from brain networks has the following challenges:

- **High non-linearity**: [30] found that the underlying structure of graph data is often non-linear. Brain network, as a typical example of graph data, has highly non-linear interactions between cortical regions [36]. How to build the model to capture these highly non-linear structure is challenging.

*Corresponding author. This work was done while the author was at the University of Illinois at Chicago.

KDD'17, August 13–17, 2017, Halifax, NS, Canada.
 ISBN 978-1-4503-4887-4/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3097983.3097988

- **Structure preserving**: The structure of brain network plays a key role in brain network analysis. However underlying structure of brain network is highly complicated [37]. Therefore, preserving the structure is difficult during the brain network mining.
- **Noise**: The analysis of real brain networks usually suffers from noise. For example, some brain connections may be wrongly placed during fMRI analysis. This kind of noise may introduce errors into diagnosis of the brain disorders.
- **Limited quantity and high dimensionality**: imaging studies of neurological disorders often involve a limited number of samples due to the limited number of available patients of interest and the high costs of acquiring the data, which significantly limits the graph mining method that can be used.

In the literature, many brain network mining methods have been proposed. Most of them concentrate on selecting frequent substructures, such as connected subgraph patterns[6, 9, 22, 26]. However, since connections between brain regions are characterized by massive complexity, it is difficult for traditional subgraph mining models to accurately capture or approximate the complicated highly non-linear structure of brain networks. Moreover, these methods only search for most frequent simple patterns, and do not consider any underlying structural information of brain networks. Additionally, as the number of subgraph patterns is exponential to the size of graphs, the subgraph mining process is both time consuming and memory intensive. Recently, some researchers leverage kernel techniques as a way to perform brain network learning [23, 46], where graph kernel is used to integrate local connectivity and global topological properties of brain networks. From the perspective of model architecture depth, all the above methods are shallow models. As stated in [47] and [50], shallow models are recognized as encountering the curse of dimensionality, and having limited capability in learning the feature representation in complex situations.

To address the above challenges, we propose a Structural Deep Brain Network Mining method, namely SDBN, to learn highly non-linear and structure-preserving representations of brain networks. Specifically, we consider to extract the locally connected ROIs as the receptive fields in the brain network and construct a multilayer convolutional neural network (CNN) architecture, which consists of multiple non-linear gating functions, for learning the graph representation in the highly non-linear latent space. To solve the issues of structural preserving and graph noise, we exploit the underlying graph structure in threefold. First, we propose a novel graph reordering approach based on spectral clustering to rearrange the order of the nodes, which can preserve the modular structure of the graph. Second, we perform structural augmentation to further enhance the spatial information of the reordered graph for structure-preserving and noise-robust feature learning. Third, we propose a deep feature learning framework for joint supervised and unsupervised learning in a small-scale setting by augmenting multilayer CNN with decoding pathways for reconstruction. Benefiting from the multiple layers of non-linear mapping, the proposed SDBN method can capture the highly non-linear structure of brain networks. With the help of the regularization power of unsupervised learning augmentation, the proposed deep model is robust to small datasets. Moreover, CNN learns features in a task-oriented style and can capture the meaningful feature for clinical tasks.

In summary, this paper makes the following contributions:

- We propose a Structural Deep Brain Network mining method, namely SDBN, to learn discriminative and meaningful graph representation from brain networks. The learned graph representation residing in the highly non-linear latent space can preserve the graph structure and be robust to the noise.
- A graph reordering technique for brain networks is proposed with the help of spectral clustering. This technique corresponds brain networks from different subjects and makes them comparable. The CNN-based deep feature learning method becomes more reliable and the underlying graph structure can be exposed.
- A structural augmentation is introduced to enhance the structure information by adding the spatial information and to alleviate the effect of the noise.
- A deep feature learning framework for joint supervised and unsupervised learning is proposed in a small-scale setting by augmenting existing neural networks with decoding pathways for reconstruction. The unsupervised learning augmentation has a regularization effect to make the learned features have better generalization from the high dimensional data.
- The proposed method is extensively evaluated on brain disorder detection with four real brain network datasets. The experiments show that SDBN can capture discriminative and meaningful structural graph representations for brain disorder diagnosis.

2 PRELIMINARY AND PROBLEM FORMULATION

In this section, we first establish the key notational conventions that we will use in the sequel. We then define the problem we are concerned with brain network mining.

Notations. Throughout this paper, vectors are denoted by boldface lowercase letters (*e.g.*, $\mathbf{a}$), matrices are denoted by boldface capital letters (*e.g.*, $\mathbf{A}$), and cubes are denoted by calligraphic letters (*e.g.*, $\mathcal{A}$). For a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$, its elements are denoted by a_{ij}, and its i-th row and j-th column are denoted by $\mathbf{a}_i$ and $\mathbf{a}^j$, respectively. We denote a weighted graph by a triple $G = (V, E, \mathbf{W})$, where $V = \{v_1, v_2, \cdots, v_N\}$ is the set of nodes, $E \subseteq V \times V$ is the set of edges connecting nodes, and $\mathbf{W}$ is the connectivity weight matrix encoding connectivity strengths of the edges. In particular, $N = |V|$ denotes the number of nodes and $M = |E|$ denotes the number of edges of the graph. We will often use calligraphic letters $(\mathcal{G}, \mathcal{H}, \cdots)$ to denote general sets or spaces, regardless of their specific nature.

Before proceeding to define our problem, we first introduce some related concepts of brain network. Hereafter we use "brain network" and "graph" interchangeably whenever there is no ambiguity.

Definition 2.1. **(Brain Network)** A brain network (or connectome), is a weighted graph $G = (V, E, \mathbf{W})$ with $|V|$ nodes and $|E|$ edges reflecting brain regions of interest (ROIs) and connectivities between ROIs, respectively. The weights or adjacency matrix in $\mathbf{W}$ represent the connectivity degree, where a larger weight corresponds to a higher connectivity degree, reflecting stronger functional correlations in fMRI and tighter fiber connections in DTI.

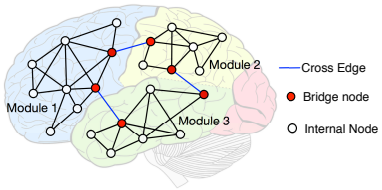


Figure 1: A brain network example, associated with modules, cross edges, internal and bridge nodes

The module structure of brain network is critical in the study of brain. To facilitate brain analysis, brain network mining is required to preserve the module structure. It can be characterized by locality and similarity. The locality describes that nodes in the same module are connected tightly while the similarity describes that nodes in the same module often have similar neighborhood structure. The definition of "module" is given as follows.

Definition 2.2. **(Module)** A module (also called community or group) in a graph is a subset of nodes which are densely connected to each other, but sparsely connected to the nodes in other modules. The node with all its neighboring nodes in the same module is internal node. The node with neighboring nodes belong to the different modules is bridge node.

To better understand and illustrate the module structure, an example of brain network is shown in Figure 1, in which lobe Frontal, lobe Parietal, lobe Temporal are represented by the modules in blue, yellow and green respectively. There are a number of nodes representing brain regions and a number of edges representing connections between pairs of regions. Among these nodes, white nodes indicate the internal nodes and red nodes indicate the bridge nodes. Between different modules, bridge nodes are connected by cross edge marked in blue. One can notice that within each module, the nodes are connected tightly to each other than the nodes in other regions, which performs strong locality and similarity.

Now we formally define the problem studied in this paper. Given N brain networks $\mathcal{G} = \{G_1, G_2, \cdots, G_N\}$, we denote their corresponding connectivity matrices $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_N\}$, where $\mathbf{W}_i \in \mathbb{R}^{N \times N}$. Let $\mathcal{H} = \{\mathbf{h}_1, ..., \mathbf{h}_N\}$. We are concerned with discovering a good vector representation $\mathbf{h}_i$ for each brain network G_i. Specifically, we are interested in learning a representation function $f : \mathcal{G} \rightarrow \mathcal{H}$, which maps each brain network $G_i \in \mathcal{G}$ into a discriminative, highly-nonlinear, structure-preserving, and clinically meaningful representation $\mathbf{h}_i \in \mathcal{H}$. The representation function is estimated on a collection of brain networks and their connectivity matrices.

Deep learning methods have powerful representation learning capability, especially in capturing the highly-nonlinear representation. Among them Convolutional Neural Network (CNN) provides an efficient framework to discover the highly-nonlinear, discriminative and task-focus representation. It can capture the local stationary structures hierarchically, which has achieved successes in many fields, such as computer vision, signal processing and nature language processing. However, applying CNN straightforwardly to brain network is not reliable. The connection between different brain regions lies on an irregular or non-Euclidean domain. This kind of edge is a universal representation of heterogeneous pairwise relation. In terms of the connectivity matrices, the receptive field of CNN can not encode the module structure well. Therefore it is usually difficult to identify common patterns and different patterns among different subjects. Here we formulate a graph reordering technique to generalize CNN to learn a structure-preserving representation $\mathcal{H}$. Its definition is as follows.

Definition 2.3. **(Graph Reordering)** Given a collection of unlabeled graphs $\mathcal{G} = \{G_1, G_2, \cdots, G_N\}$, the goal of graph reordering is to find a labeling ℓ such that for any two graphs $G_i, G_j \in \mathcal{G}$ drawn uniformly at random from $\mathcal{G}$, the expected difference between the distance of the graph connectivity matrices based on ℓ and the distance of the graphs in graph space is minimized. Let $\mathbf{d}_{\mathbf{G}}$ be a distance measure on graphs $\mathcal{G}$, and $\mathbf{d}_{\mathbf{W}}$ be a distance measure on connectivity matrices $\mathcal{W}$. It can be formulated as the following optimization problem:

$$\arg\min_{\ell} \mathbb{E}_{\mathbf{G}}[\|\mathbf{d}_{\mathbf{W}}(\mathbf{W}_i^{\ell}, \mathbf{W}_j^{\ell}) - \mathbf{d}_{\mathbf{G}}(G_i, G_j)\|] \quad (1)$$

Graph reordering is essentially a graph labeling procedure, which assigns labels to the nodes or edges, or both. In general, Eq. (1) is a difficult combinatorial problem. Solving this problem would require $n!$ trials in the worst case. Therefore optimal graph reordering is NP-hard and computationally infeasible. Alternatively, we borrow the idea of graph compression [8], handling this problem by performing module identification in brain.

In the following, we will elaborate our approach towards above concepts.

3 PROPOSED METHOD

As discussed in the introduction, learning good feature representations from brain networks is challenging due to its complex and non-linear structure. To overcome these challenges, we propose a structural deep brain network mining (SDBN) method. An overview of SDBN is shown in Figure 2, which consists of three main steps: (1) Graph Reordering; (2) Structural Augmentation; and (3) Convolutional Deep Feature Learning. These steps collectively address the challenges of brain network representation learning.

3.1 Graph Reordering

The objective of graph reordering is to leverage graph labeling procedure to reorder the nodes in the connectivity matrix, where the difference of distances from the connectivity matrices of graphs is minimized, in the sense that nodes close to each other are more likely to have strong connections. In this way, a permutation of the nodes across different brain networks is established. By graph reordering, nodes belonging to the same module will be located close to each other [13] and the structural topology within the graphs is very similar [33]. This would make CNN more able to bring out the latent network structures shared by the different networks. However, as the optimal graph reordering is NP-hard, it is infeasible to solve Eq. (1) directly. Alternatively, we borrow the idea from graph compression [2] to address this problem. The intuitive idea of graph compression is to produce a compressed graph of a given weighted graph by finding an ordering of nodes, such that the connectivity matrix is close to block-diagonal. Specifically, it was noted that finding modules is a useful conceptual tool for analyzing

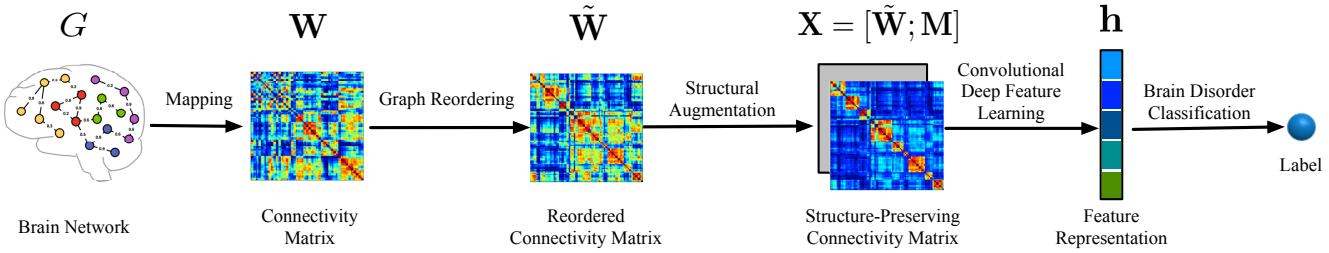


Figure 2: An overview of the proposed SDBN method. Connectivity matrix, reordered connectivity matrix and structure-preserving connectivity matrix are represented in heat map, where high heat color indicates strong connection. The structure-preserving connectivity matrix consists of two channels: the first channel is the reordered connectivity matrix with the weights of the off-block diagonal regions being lowered. The second channel is the module identification matrix M

relations between the original and compressed graphs. Based on this insight, we identify the modules among brain networks to approximate the optimal graph reordering.

Currently various methods have been used for module identification. Among the existing methods, spectral clustering often yields more superior performance compared to other methods. Spectral methods have also been successfully applied to graph reordering problems [24]. Therefore, we employ the spectral clustering to perform module identification. The key idea of spectral clustering is to convert a clustering problem into a graph partitioning problem and then solve this problem by means of matrix theory. Let K be the number of modules to be identified, and $\hat{\mathbf{W}} \in \mathbb{R}^{N\times N}$ be the average connectivity matrix across all graphs. Then, spectral clustering can be formulated as follows:

$$\min_{\mathbf{F}} \sum_{i,j=1}^{N} \hat{w}_{ij} \left\|\mathbf{f}_i - \mathbf{f}_j\right\|_2^2 = \mathrm{tr}\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F}\right)$$
$$\text{s.t. } \mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{I}_K \tag{2}$$

where $tr(\cdot)$ denotes the trace function, a superscript T denotes transposition, $\mathbf{I}_K$ denotes the identity matrix with size K, and $\mathbf{L}$ is the Laplacian matrix [45] obtained from $\hat{\mathbf{W}}$.

By applying K-means to the eigenvectors of the Laplacian matrix $\mathbf{L}$, we can obtain K modules $\mathcal{M} = \{M_1, M_2, \cdots, M_K\}$, with $V = M_1 \cup M_2 \cup \cdots \cup M_K$ and $M_i \cap M_j = \varnothing$ for every pair i, j with $i \neq j$. We treat the module sequence $1, 2, \cdots, K$ as a permutation of nodes. Based on this, the reordered connectivity matrix $\tilde{\mathbf{W}}$ for each brain network can be established. Since all reordered connectivity matrices are obtained according to the global module structure, different subjects are more comparable. An example of the reordered connectivity matrices for abnormal and normal subjects is shown in the second column (b) and (e) of Figure 3.

3.2 Structural Augmentation

According to the graph reordering procedure above, we obtain the reordered connectivity matrix for each brain network. This representation is more meaningful and beneficial to CNN architecture. However, there are still two problems affecting the use of CNN. The first one is the spatial information loss caused by CNN itself. The second one is the noise in brain network. To address these problems, we propose a structural augmentation approach to further enhance the spatial information of the reordered graph for structure-preserving and noise-robust feature learning.

CNN adopts max pooling to perform invariant properties to translation, rotation and shifting, which results in the spatial information loss. However, the spatial relation of each receptive field is also important structure information for brain network mining. Given the top layer of the network, the original spatial information is lost. Naturally, the clinical structure patterns only occur in specific brain regions and those occurring in the other brain regions are usually invalid.

To tackle this problem, we augment the reordered connectivity matrix with an additional channel. Specifically, we define a module identification matrix $\mathbf{M}$ to further encode the module structure information, whose element is:

$$m_{ij} = \begin{cases} k & \text{for} \quad v_i, v_j \in M_k, i, j = 1, ..., N \\ 0 & \text{for} \quad otherwise \end{cases} \tag{3}$$

The constructed $\mathbf{M}$ is a block-diagonal matrix. It is concatenated with the reordered connectivity matrix as an additional channel as shown in Figure 2. With such an information augmentation, we could preserve the spatial information during the feature learning procedure.

On the other hand, brain networks usually suffer from noise, which is introduced by the error in the acquisition and in the image analysis. For the reordered connectivity matrix obtained from graph reordering, the intra-module neighbor edges reside in the block-diagonal region and the cross edges lie in the off-diagonal region. The intra-module neighbor edges preserve the structure within each module, which are more reliable to learn structure-preserving graph representation, while the cross edges are more complicated which may have negative effects to study the modular structure and some of them may be invalid connections [18]. To alleviate the effects of cross edges, we further augment the reordered connectivity matrix $\tilde{\mathbf{W}}$ by lowering the weights of the off-block diagonal regions such that

$$\tilde{w}_{ij} = \begin{cases} 1 & \text{for} \quad i, j \in M_k, k = 1, ..., K \\ \epsilon & \text{for} \quad i, j \notin M_k, k = 1, ..., K \end{cases} \tag{4}$$

An example of the resulted new reordered connectivity matrices for abnormal and normal subjects is shown in the third column (c) and

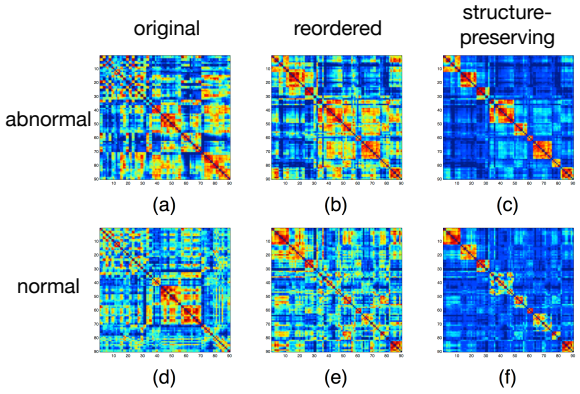


Figure 3: The first row and second row show the connectivity matrices of the abnormal subjects and that of the normal subjects, respectively. From left to right, the first column Figures (a) and (d) present the original connectivity matrices mapped from the brain network, the second column Figures (b) and (e) indicate the reordered connectivity matrices after the graph reordering and the third column Figures (c) and (f) show the first channel of structure preserving connectivity matrices.

(f) of Figure 3. Together with module information augmentation, we obtain the structure-preserving connectivity matrix $\mathbf{X} = [\tilde{\mathbf{W}}; \mathbf{M}]$ as shown in Figure 2, which encodes more structural information and is beneficial for CNN.

3.3 Unsupervised Learning Augmented CNN

In this section, the resulting connectivity matrices are fed into the CNN architecture for feature learning. Specifically, we introduce a novel CNN architecture to learn highly non-linear feature representations from the brain network data while dealing with the problem of small sample size and high dimensionality.

CNN is a supervised feature learning technique, which has demonstrated state-of-the-art performance for image analysis [27]. Recent works [14] [39] [43] commonly nest a group of convolutional layers and a pooling layer to construct the basic feature extractor in different scales. The extracted internal representation preserves highly-nonlinear properties. Consider the feature extractors in different scales as cells, the input data goes over a number of convolution-pooling cells and fully connected layer. Finally, they are fed to the top-layer classifier. The CNN with L cells can be defined as follows

$$\mathbf{X}^l = f_l(\mathbf{X}^{l-1}; \Theta_l) \quad \text{for} \quad l = 1, 2, ..., L+1 \tag{5}$$

where $\mathbf{X}^0 = \mathbf{X}$ is the input,which is the resulted structure-preserving connectivity matrix. $f_l(\cdot)$ for $l = 1, 2, ..., L$ are the feature mapping functions of the $l-th$ cell with the parameter Θ_l. The final classification loss can be represented as follows

$$C(\mathbf{X}^0, y) = \ell(\mathbf{X}^{L+1}, y) \tag{6}$$

where y corresponds to the ground truth label, which indicates the patients' disease diagnose result, and ℓ indicates the classification loss function.

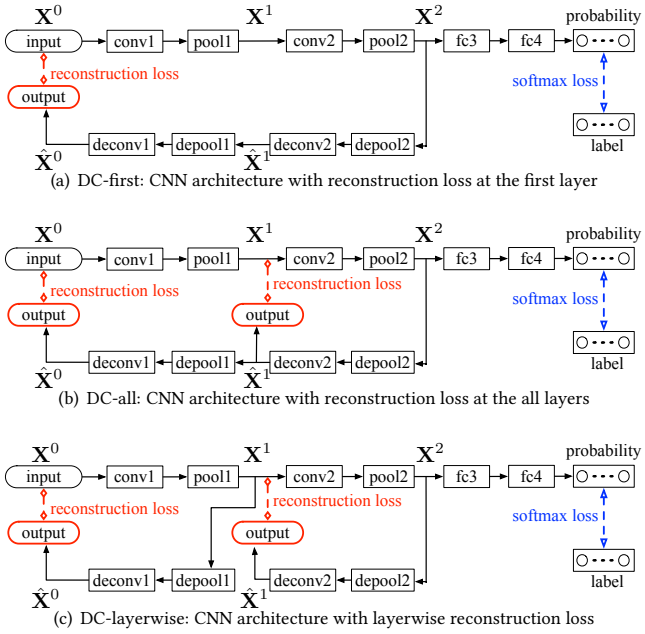


Figure 4: CNN architecture with three types of unsupervised learning augmentation

Given a set of subject data $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$ with corresponding labels $y_1, y_2, ..., y_N$, the CNN is trained by minimizing the objective function as follows

$$\frac{1}{N}\sum_{i=1}^{N} C(\mathbf{X}_i, y_i) \tag{7}$$

However, CNN usually requires a large amount of labeled data for training, which is infeasible for current brain health research, due to the limited number of available patients of interest and the high costs of acquiring the data. It could potentially limit the performance of CNN.

Motivated by the recent attempts [28, 35, 41, 49] to couple the unsupervised and supervised learning in the same phase, making the unsupervised objective being able to impact the network training after supervised learning took place, we augment the auxiliary unsupervised learning objective function to the supervised learning objective function to address the above problem. The joint objective function is as follows

$$\frac{1}{N}\sum_{i=1}^{N}(C(\mathbf{X}_i, y_i) + \lambda U(\mathbf{X}_i)) \tag{8}$$

where $U(\cdot)$ indicates the unsupervised objective function, which can be one or more auxiliary pathways that are attached to the convolution-pooling cells in the original CNN.

Specifically, given the CNN architecture for classification, we take the sub-network consists of all the convolution-pooling cells as the encoding pathway, and build a mirrored decoder network as an auxiliary pathway of the original network. The fully connected layers are not used for reconstruction, since they are more related to supervised learning objective. For a CNN with 2 convolution-pooling cells, we consider three types of unsupervised learning

augmentation via decoding: DC-first, DC-all and DC-layerwise. Their frameworks are shown in Figure 4. The decoding pathway is defined as follows

$$\hat{\mathbf{X}}^L = \mathbf{X}^L, \quad \hat{\mathbf{X}}^{l-1} = f_l^{dec}(\hat{\mathbf{X}}^l; \Theta_l'), \quad \hat{\mathbf{X}} = \hat{\mathbf{X}}^0 \tag{9}$$

where Θ_l' are decoder parameters.

Figure 4(a) shows the framework of DC-first. In DC-first, the CNN is just augmented with a stacked autoencoder. The decoding starts from the pooled feature generated by the second pooling layer and goes to reconstruct the original input. Reconstruction errors are measured between the final reconstruction input and the CNN input. The auxiliary training signals of DC-first come from the bottom of the decoding pathway and are incorporated with the top-down signals from the supervised learning into the encoder pathway. The corresponding unsupervised loss is as follows:

$$U_{DC-first}(\mathbf{X}) = ||\hat{\mathbf{X}} - \mathbf{X}||_2^2 \tag{10}$$

Figure 4(b) shows the framework of DC-all. Different from DC-first, it allows more gradients to flow directly to the previous cells. This model makes the autoencoder perform better mirrored architecture by matching activations for all the cells. Its unsupervised learning loss is as follows:

$$U_{DC-all}(\mathbf{X}) = \sum_{l=1}^{L-1} ||\hat{\mathbf{X}}^l - \mathbf{X}^l||_2^2 \tag{11}$$

Figure 4(c) shows the framework of DC-layerwise, which is another autoencoder variant with layerwise decoding architecture. It reconstructs the output activations of every cell to its input. Its unsupervised learning loss is the same as DC-all. Different from DC-all, its decoding pathway is defined as follows:

$$\hat{\mathbf{X}}^{l-1} = f_l^{dec}(\mathbf{X}^l; \Theta_l') \tag{12}$$

The first two methods encourage top-level convolution features to preserve as much information as possible. In contrast, the decoding pathways in DC-layerwise concentrates on inverting the clean intermediate activations to the input of the associated cell-layer, performing parallel layerwise training.

The training procedure of our proposed method is as follows: (1) Train the CNN with labels; (2) Train "layerwise" decoding pathways; (3) Train the top-down decoding pathways; and (4) Fine-tune the entire augmented network. With the help of the proposed deep convolutional feature learning framework, we can address the three problems above and learn discriminative graph feature representations.

4 EXPERIMENTS AND RESULTS

In this section, we evaluate the proposed methods on four real-world datasets. We first introduce the datasets used in the experiments, the compared methods and the experimental setting. Next, we show the performance of compared methods on four brain network datasets described below. We further conduct several case studies, including the effectiveness of graph reordering using different spectral clustering configurations (corresponding to different Laplacian matrices) and the effectiveness of unsupervised learning augmentation.

Table 1: The statistics for each dataset

Name	HIV-fMRI	HIV-DTI	BP-fMRI	ADHD-fMRI
Category	fMRI	DTI	fMRI	fMRI
Quantity	77	77	97	776
Size	90 x 90	90 x 90	82 x 82	90 x 90

4.1 Data Collection

In the experiments, we evaluate the proposed methods on four brain network datasets from three different neurological disorders.

- **HIV-fMRI-77 & HIV-DTI-77**: Chicago Early HIV Infection Study at Northwestern University [34] collected both fMRI and DTI data for 56 HIV (positive) and 21 seronegative controls (negative). By propagating the Automated Anatomical Labeling (AAL) to each image [44], each brain is represented as a graph with 90 nodes corresponding to cerebral regions (45 for each hemisphere). A detailed description about data acquisition and preprocessing is available in [5].
- **BP-fMRI-97**: This dataset consists of 52 bipolar I subjects who are currently in euthymia and 45 age and gender matched healthy controls. Using the Freesurfer-generated cortical/subcortical gray matter regions, functional brain networks were derived using pairwise BOLD signal correlations of 82 brain regions. A detailed description about data acquisition and preprocessing is available in [6].
- **ADHD-fMRI-776**: This dataset is collected from ADHD-200 global competition dataset[1]. The dataset contains records of resting-state fMRI images for 776 subjects, which are labeled as real patients (positive) and normal controls (negative). Similarly, each brain is parcellated into 90 brain regions.

The statistics for each dataset is summarized in Table 1. The HIV-fMRI, HIV-DTI, BP-fMRI are relatively small datasets, while the ADHD-fMRI is a relatively big dataset.

4.2 Compared Methods

To evaluate the performance of the proposed method, we compare it with other state-of-the-art methods. All the compared methods are summarized as follows:

- **gSSC**: It is a semi-supervised subgraph feature selection method [25] based upon both labeled and unlabeled graphs.
- **Freq**: This method is an unsupervised subgraph feature selection method based upon frequency. The top-n frequent subgraph features are selected.
- **Conf**, **Ratio**, **Gtest**, **HSIC**: These methods are supervised subgraph selection [26] based upon confidence, frequency ratio, G-test score and HSIC, respectively. The top-n discriminative subgraph features are selected in terms of different discrimination criteria.
- **AE**: It is an unsupervised feature learning method based on autoencoder [48] which learns the latent representation of all the connectivity data without considering any structural information.

[1] http://neurobureau.projects.nitrc.org/ADHD200

- **CNN**: It is a plain vanilla convolutional architecture stacking the convolution layer, pooling layer, non-linear gating layer and fully connected layer. The graph reordering and structural augmentation is not applied.

Since the proposed SDBN consists of several components, we consider several variations of SDBN as follows:

- **CNN-GN**: It is a convolutional architecture with graph reordering of nodes. It has three versions: **SP**, **NJW** and **SM** [45], each of which uses a different spectral clustering method. SP indicates the graph reordering using the unnormalized spectral clustering, which aims to minimize ratio cut. NJW indicates the graph reordering with the normalized spectral clustering by Ng-Jordan-Weiss Algorithm [32]. SM indicates the graph reordering with normalized spectral clustering by Shi and Malik [38], which aims to minimize normalized cut.
- **CNN-GN-A**: It is a convolutional architecture with graph reordering and structural augmentation.
- **CNN-GN-A-DC**: It is a convolutional architecture with graph reordering, structural augmentation and unsupervised learning augmentation. It has three variations: **DC-first**, **DC-all**, **DC-layerwise**. DC-first uses the augmentation with reconstruction loss at the first layer. DC-all uses the augmentation with reconstruction loss at all layers. DC-layerwise uses the augmentation with a layerwise architecture.
- **SDBN**: It is the method proposed in this work which can learn structural-preserving graph feature representation that lies in a highly non-linear latent space. There are two versions of SDBN: one with a shallow architecture (**SDBN-S**) and one with a deep architecture (**SDBN-D**). We select the best parameter configuration for them.

4.3 Experimental Setting

To evaluate the quality of the learned brain network representation, we feed the learned feature representation to a softmax classifier for each compared method. For the subgraph mining methods, the threshold for four dataset is 0.9, 0.3, 0.5, 0.5, respectively. We select top-100 features for classification as in [48].

For SDBN, we first select the configuration of the graph reordering. By empirical study, we choose the SM spectral clustering method and set the cluster size $K = 9$. Then we choose the configurations to build CNN. We build a convolutional architecture with different layers as shown in Table 2. It stacks the convolution layer, pooling layer, non-linear gating layer and fully connected layer. For a fair comparison, we set the number of the neurons in the fully connected layer to 100. SDBN-S has the following setup: LayerC-LayerP-LayerN-LayerF-Layer-N, while SDBN-D has the following setup: LayerC-LayerP-LayerN-LayerC-LayerP-LayerN-LayerF-LayerN. To prevent the overfiting problem, the dropout is added after the fully connected layer with dropout rate being 0.5. In the experiment, we apply AdaGrad [11], an adaptive gradient-based optimization approach that automatically determines a per-parameter learning rate, for parameter learning. Since there are limited number of subjects in the datasets, we perform a 3-fold cross validation as in [3] on balanced datasets and report the average results.

Table 2: The configuration of different cell-layers in the convolution architecture. LayerC, LayerP, LayerN and LayerF indicate the convolution layer, the pooling layer, the non-linear gating function and the fully connected layer, respectively.

Name	LayerC	LayerP	LayerN	LayerF
Type	conv	max-pool	gating	fc
Parameter	$5 \times 5 \times 50$	2×2	relu	100

4.4 Performance on Brain Disorder Detection

In this section, we investigate the effectiveness of the learned structure-preserving graph feature representations for brain disease detection. The average classification performance are shown in Table 3. Two evaluation metrics are used, including accuracy and F1 score. It is easy to see that SDBN outperforms all the other baseline methods on all four datasets. Compared with subgraph mining methods that search in a potentially exponential space, the proposed methods achieve better performance and consume around 83% less memory and computation resources. In additional we can see that AE has a little better performance than the subgraph mining methods in most of cases. This is probably because, the connectivity patterns learned from AE can capture the non-linear complex structure in brain network. CNN is nearly the worst one among all the models, since applying CNN directly to brain network is not reliable. Compared with the shallow feature learning methods, our shallow model SDBN-S outperforms AE. This is due to the fact that SDBN can identify more complex graph representations, while AE can only identify simple representations. It supports our premise that more discriminative graph feature representation can be identified by SDBN. Considering the deepness, our deep architecture SDBN-D achieves a better result than the shallow architecture SDBN-S, since the features learned from the deep architecture can capture hierarchical and highly non-linear structure patterns. It is also worth noticing that deep model has limited improvement compared with shallow model on small data set such as HIV-fMRI, HIV-DIT, BP-fMRI, while shows significant improvement in the a relative big dataset such as ADHD-fMRI. When the size of dataset is big, the deep learning methods show their superiority.

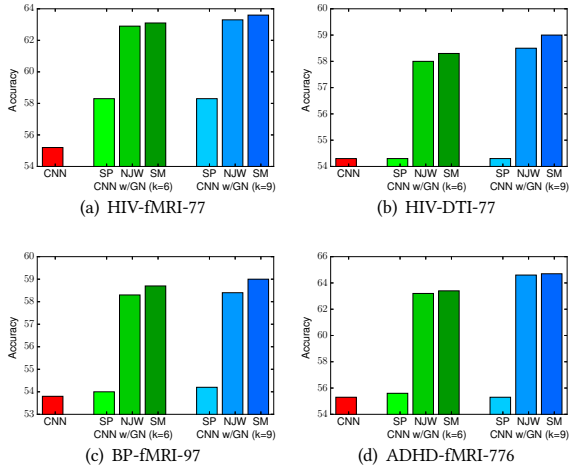
4.5 Case Study of Graph Reordering

In this part, we study the importance of graph reordering in the proposed method. We compare the performance of CNN with and without graph reordering. In addition, we study how the graph reordering using different spectral clustering methods and cluster sizes affects the quality of learned feature representation.

We conduct experiments for brain disorder detection with three different spectral clustering methods and two cluster numbers on all four datasets. The clustering methods used in the experiments include SP, NJW, and SM [45] as mentioned above. Two cluster sizes 6 and 9 are considered. The average accuracies are reported in Figure 5. Compared the CNN without graph reordering, CNNs with graph reordering have a significant gain except for the ones using SP. It indicates the importance of graph reordering. As an unnormalized spectral clustering method, SP has a limited clustering ability, especially for brain networks. It has limited improvement

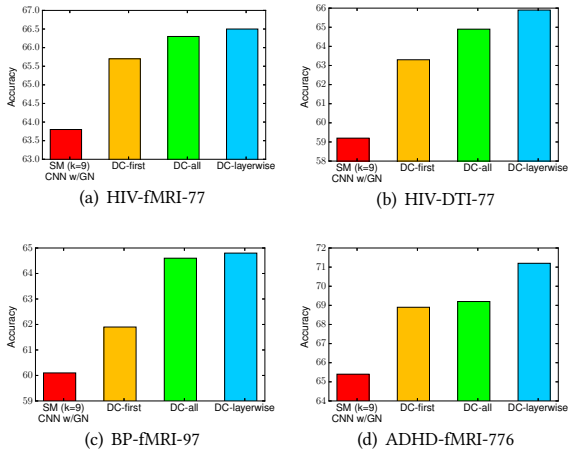
Table 3: Performances of the compared methods. The results are reported as "average performance + (rank)".

Dataset	Measure	Method									
		gSSC	Freq	Conf	Ratio	Gtest	HSIC	AE	CNN	SDBN-S	SDBN-D
HIV-fMRI-77	ACC	60.0 (4)	54.3 (8)	58.6 (6)	54.3 (8)	50 (9)	58.7 (5)	62.9 (3)	55.2 (9)	65.3 (2)	**66.5 (1)**
	F1	62.5 (4)	58.2 (8)	64.2 (3)	62.0 (5)	52.5 (10)	59.5 (7)	60.4 (6)	56.7 (9)	65.4 (2)	**66.7 (1)**
HIV-DIT-77	ACC	59.5 (4)	64.6 (3)	52.4 (9)	59.3 (5)	59.3 (5)	57.9 (6)	57.8 (7)	54.3 (8)	64.7 (2)	**65.9 (1)**
	F1	59.6 (4)	63.9 (3)	46.1 (10)	57.9 (7)	58.5 (5)	58.3 (6)	56.8 (8)	53.5 (9)	64.5 (2)	**65.6 (1)**
BP-fMRI-97	ACC	56.7 (5)	56.8 (4)	50.8 (10)	54.2 (8)	55.2 (6)	54.9 (7)	60.4 (3)	53.8 (9)	63.0 (2)	**64.8 (1)**
	F1	57.8 (4)	57.6 (5)	49.1 (10)	53.7 (8)	53.9 (7)	55.8 (6)	59.8 (3)	52.1 (9)	62.5 (2)	**63.7 (1)**
ADHD-fMRI-776	ACC	55.7 (4)	48.5 (7)	54.5 (6)	48.5 (7)	54.5 (6)	46.1 (8)	61.3 (3)	55.3 (5)	68.9 (2)	**71.2 (1)**
	F1	55.9 (4)	55.0 (5)	51.4 (7)	55.0 (5)	51.4 (7)	50.5 (8)	60.1 (3)	54.2 (6)	67.6 (2)	**69.9 (1)**



(a) HIV-fMRI-77 (b) HIV-DTI-77 (c) BP-fMRI-97 (d) ADHD-fMRI-776

Figure 5: Case study of graph reordering. In each bar chart, the CNN methods with and without graph reordering (GR) are compared. For CNN w/GN, different cluster numbers m are tested.



(a) HIV-fMRI-77 (b) HIV-DTI-77 (c) BP-fMRI-97 (d) ADHD-fMRI-776

Figure 6: Case study of unsupervised learning augmentation. In each bar chart, the method with good graph reordering (SM, k=9), and the methods with additional unsupervised learning augmentation are compared.

on HIV-DTI, BP-fMRI and ADHD-fMRI compared with CNN without graph reordering. The clusters it produces cannot preserve the underlying structure well and are not helpful for CNN feature learning. For the rest of two clustering methods, SM is slightly better than NJW, due to its better clustering ability. Hence, the spectral clustering method is the key to graph reordering. Compared with graph reordering using different cluster sizes, the ones with 9 clusters are slightly better than the ones with 6 clusters, which confirms the same observation in [31]. It indicates that the cluster size is also an important factor for graph reordering. Overall, the quality of spectral clustering is important to graph reordering and good spectral clustering will produce a graph reordering close to the optimal graph reordering.

Spectral clustering not only supports graph reordering but also exposes the rich graph structure of brain networks. The generated clustering information can be reused for structural augmentation. In order to test its effectiveness, we add the modular structure matrix during the feature learning process. It is found that with structural augmentation, we usually have about 1% improvement based on CNN with good graph reordering.

In conclusion, graph reordering is critical to learn graph representation directly from the brain network data using CNN architecture. After graph reordering, the structure-preserving locally connectivity regions can be extracted and serve as the receptive fields of CNN, allowing the CNN framework to learn effective graph representations.

4.6 Case Study of Unsupervised Learning Augmentation

In this part, we investigate the effectiveness of unsupervised learning augmentation in the proposed method. With the best reordering configuration from previous section, we compare CNN with and without unsupervised learning augmentation. We conduct experiments for brain disease detection with three different unsupervised learning augmentations on all four datasets. We take the sub-network of CNN composed of all the convolution-pooling layers as the encoder and generate a fully mirrored decoder network as an auxiliary pathway of the original network. We measure the reconstruction loss in three ways: DC-first, DC-all, DC-layerwise as described above and augment the supervised loss with these

reconstruction loss. The average accuracies are shown as in Figure 6. We can see that on four datasets, CNNs with unsupervised learning augmentation outperform the ones without augmentation. It demonstrates the effectiveness of unsupervised learning augmentation, and the performance of different augmentation methods are comparable. Among them, the DC-layerwise is slightly better than the other two methods, the DC-all is slightly better than DC-first.

Overall, unsupervised learning augmentation is helpful for learning graph representation. This can be explained in two folds. First, the spatial information is critical for brain network analysis. The clinical disrupted conductivities is module inspired and usually happens among certain ROIs. The traditional CNN's max pooling operation preserves the spatial invariance properties but removes the spatial information. Reconstruction objective is spatial corresponded, which can help to keep the spatial information. Second, the decoding pathways help the supervised objective reach a better optimum and regularize the final objective solution. Third, the layer-wise reconstruction loss is an effective way to regularize the network training.

5 RELATED WORK

Our work is related to deep learning, graph reordering and brain data mining. We briefly discuss them in the followings.

5.1 Deep Learning

Representation learning has long been an important problem of machine learning and many works aim at learning representations for samples [1]. Recent years, deep learning is becoming more and more popular, due to its powerful ability of learning feature with different kinds properties and can be applied for different kind of data. It made success in different fields, such as computation vision [14, 27, 39, 43], nature language processing [40] and speech recognition [21]. Several recent works have been proposed which show neural architectures for graph input data [12, 20, 29], these methods are trying to apply the convolutional neural network on the simple graph structure data. However, to the best of our knowledge, there have been few deep learning works on brain network [48], especially learning brain network representations.

5.2 Graph Reordering

Graph Reordering is very active in the field of data base and data mining. It is widely used for the graph compression task. which focus on finding homogeneous regions in the graph so that nodes inside a region are tightly connected to each other than to nodes in other regions. In terms of the adjacency matrix, the target was to find an ordering of nodes so that the adjacency matrix is close to block-diagonal, containing more "square" blocks. Spectral clustering [32, 38], co-clustering [10], cross-associations [7], and shingle-ordering [8] are typical examples for such approaches. Graph reordering also has some other applications. Recently, [33] proposed a reordering method for the nodes of the neighborhood graph so as to map from the unordered graph space to a vector space with a linear order and try to learn from graph with the help of CNN with considering the node attribute.

5.3 Brain Data Mining

Applications of data mining techniques to brain network data analysis are related to our work as well. Wee et al. used clustering coefficients of each brain region and a multi-kernel SVM for early detection of brain abnormalities in Alzheimer's disease [46]. Jie et al. used a graph kernel and integrate local connectivity and global topological properties of fMRI brain networks [23]. Kong et al. proposed a discriminative subgraph selection approach for uncertain fMRI networks [6, 26]. Cao et al. proposed a supervised tensor factorization technique to analyze EEG brain networks [4]. Ma et al. investigate the interior-node topology of brain networks [31]. Zhang et al. presented an autoencoder architecture with the guidance of side information for brain network analysis [48]. He et al. proposed several tensor based methods to analyze brain data [15–17]. In contrast to the existing works, this work presents a novel approach to discover highly non-linear patterns from brain network data based on a deep convolutional architecture.

6 CONCLUSION AND OUTLOOK

In this paper, a structural deep brain network mining method, namely SDBN, is proposed to learn structural preserving and clinically meaningful representation from brain network. SDBN consists of several components: (1) Graph Reordering (2) Structural Augmentation, and (3) Convolutional Deep Feature Learning. Based on spectral clustering, graph reordering is proposed to impose an order on the nodes of the neighborhood graph so as to preserve the global and local structure of the brain network. Next, we perform structural augmentation to further enhance the spatial information of the reordered graph. Then a deep feature learning framework is introduced, which joint supervised and unsupervised learning in a small-scale setting by augmenting existing supervised CNN with decoding pathways for unsupervised reconstruction. With the help of the multiple layers of non-linear mapping, it can capture the highly non-linear network structure of the brain network. To evaluate our methods, we conduct the experiment on four real brain network datasets for disease diagnoses. The experiments show that SDBN can discover discriminative and meaningful structural graph representation for brain disorder diagnosis. Since the proposed deep feature learning framework is end-to-end and task-oriented, its application is not limited to binary disease classification. It can be easily extended to the other clinical task with objectives such as multi-class classification, clustering, regression and ranking. We plan to apply our framework for the other medical task.

7 ACKNOWLEDGEMENTS

This work is supported in part by NSF through grants IIS-1526499 and CNS-1626432, and NSFC through grants 61503253 and 61672313, and NIH through grant R01-MH080636. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research.

REFERENCES

[1] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[2] Paolo Boldi and Sebastiano Vigna. The webgraph framework i: compression techniques. In *Proceedings of the 13th International Conference on World Wide*

Web, pages 595–602, 2004.

[3] Bokai Cao, Lifang He, Xiangnan Kong, Philip S. Yu, Zhifeng Hao, and Ann B. Ragin. Tensor-based multi-view feature selection with applications to brain diseases. In *IEEE International Conference on Data Mining*, pages 40–49, 2014.

[4] Bokai Cao, Lifang He, Xiaokai Wei, Mengqi Xing, Philip S. Yu, Heide Klumpp, and Alex D Leow. t-bne: Tensor-based brain network embedding. In *Proceedings of the SIAM International Conference on Data Mining*, 2017.

[5] Bokai Cao, Xiangnan Kong, Jingyuan Zhang, Philip S. Yu, and Ann B. Ragin. Identifying HIV-induced subgraph patterns in brain networks with side information. *Brain Informatics*, 2015.

[6] Bokai Cao, Liang Zhan, Xiangnan Kong, Philip S. Yu, Nathalie Vizueta, Lori L Altshuler, and Alex D Leow. Identification of discriminative subgraph patterns in fmri brain networks in bipolar affective disorder. In *International Conference on Brain Informatics and Health*, pages 105–114, 2015.

[7] Deepayan Chakrabarti, Spiros Papadimitriou, Dharmendra S Modha, and Christos Faloutsos. Fully automatic cross-associations. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 79–88, 2004.

[8] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Michael Mitzenmacher, Alessandro Panconesi, and Prabhakar Raghavan. On compressing social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 219–228, 2009.

[9] Ian Davidson, Sean Gilpin, Owen Carmichael, and Peter Walker. Network discovery via constrained tensor analysis of fmri data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 194–202, 2013.

[10] Inderjit S Dhillon, Subramanyam Mallela, and Dharmendra S Modha. Information-theoretic co-clustering. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 89–98, 2003.

[11] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[12] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, pages 2224–2232, 2015.

[13] Daniel Gamermann, Arnau Montagud, RA Infante, Julian Triana, PF de Córdoba, and JF Urchueguía. Pynetmet: Python tools for efficient work with networks and metabolic models. *arXiv preprint arXiv:1211.7196*, 2012.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[15] Lifang He, Xiangnan Kong, Philip S. Yu, Xiaowei Yang, Ann B. Ragin, and Zhifeng Hao. Dusk: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimages. In *Proceedings of the SIAM International Conference on Data Mining*, pages 127–135, 2014.

[16] Lifang He, Chun-Ta Lu, Hao Ding, Shen Wang, Linlin Shen, Philip S. Yu, and Ann B. Ragin. Multi-way multi-level kernel modeling for neuroimaging classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[17] Lifang He, Chun-Ta Lu, Guixiang Ma, Shen Wang, Linlin Shen, Philip S. Yu, and Ann B. Ragin. Kernelized support tensor machines. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[18] Lifang He, Chun-Ta Lu, Jiaqi Ma, Jianping Cao, Linlin Shen, and Philip S. Yu. Joint community and structural hole spanner detection via harmonic modularity. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 875–884, 2016.

[19] RK Heaton, DB Clifford, DR Franklin, SP Woods, C Ake, F Vaida, RJ Ellis, SL Letendre, TD Marcotte, JH Atkinson, et al. Hiv-associated neurocognitive disorders persist in the era of potent antiretroviral therapy charter study. *Neurology*, 75(23):2087–2096, 2010.

[20] Mikael Henaff, Joan Bruna, and Yann LeCun. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

[21] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 2012.

[22] Shuai Huang, Jing Li, Jieping Ye, Adam Fleisher, Kewei Chen, Teresa Wu, and Eric Reiman. Brain effective connectivity modeling for alzheimer's disease by sparse gaussian bayesian network. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 931–939, 2011.

[23] Biao Jie, Daoqiang Zhang, Wei Gao, Qian Wang, Chong-Yaw Wee, and Dinggang Shen. Integration of network topological and connectivity properties for neuroimaging classification. *IEEE Transactions on Biomedical Engineering*, 2014.

[24] Martin Juvan and Bojan Mohar. Optimal linear labelings and eigenvalues of graphs. *Discrete Applied Mathematics*, 36(2):153–168, 1992.

[25] Xiangnan Kong and Philip S. Yu. Semi-supervised feature selection for graph classification. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 793–802, 2010.

[26] Xiangnan Kong, Philip S. Yu, Xue Wang, and Ann B. Ragin. Discriminative feature selection for uncertain graph classification. In *Proceedings of the SIAM International Conference on Data Mining*, pages 82–93, 2013.

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[28] Hugo Larochelle and Yoshua Bengio. Classification using discriminative restricted boltzmann machines. In *Proceedings of the 25th International Conference on Machine Learning*, pages 536–543, 2008.

[29] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015.

[30] Dijun Luo, Feiping Nie, Heng Huang, and Chris H Ding. Cauchy graph embedding. In *Proceedings of the 28th International Conference on Machine Learning*, pages 553–560, 2011.

[31] Guixiang Ma, Lifang He, Bokai Cao, Jiawei Zhang, Philip S. Yu, and Ann B. Ragin. Multi-graph clustering based on interior-node topology with applications to brain networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 476–492, 2016.

[32] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.

[33] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *Proceedings of the 33rd International Conference on Machine Learning*, 2016.

[34] Ann B. Ragin, Hongyan Du, Renee Ochs, Ying Wu, Christina L Sammet, Alfred Shoukry, and Leon G Epstein. Structural brain alterations can be detected early in hiv infection. *Neurology*, 2012.

[35] Marc'Aurelio Ranzato and Martin Szummer. Semi-supervised learning of compact document representations with deep networks. In *Proceedings of the 25th International Conference on Machine Learning*, pages 792–799, 2008.

[36] Mikail Rubinov, Stuart A Knock, Cornelis J Stam, Sifis Micheloyannis, Anthony WF Harris, Leanne M Williams, and Michael Breakspear. Small-world properties of nonlinear brain activity in schizophrenia. *Human Brain Mapping*, 30(2):403–416, 2009.

[37] Blake Shaw and Tony Jebara. Structure preserving embedding. In *Proceedings of the 26th International Conference on Machine Learning*, pages 937–944, 2009.

[38] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[40] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, Christopher Potts, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 1631, page 1642, 2013.

[41] Kihyuk Sohn, Guanyu Zhou, Chansoo Lee, and Honglak Lee. Learning and selecting features jointly with point-wise gated boltzmann machines. In *Proceedings of the 30th International Conference on Machine Learning*, pages 217–225, 2013.

[42] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 2005.

[43] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[44] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 2002.

[45] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 2007.

[46] Chong-Yaw Wee, Pew-Thian Yap, Daoqiang Zhang, Kevin Denny, Jeffrey N Browndyke, Guy G Potter, Kathleen A Welsh-Bohmer, Lihong Wang, and Dinggang Shen. Identification of mci individuals using structural and functional connectivity networks. *Neuroimage*, 2012.

[47] Zhaoquan Yuan, Jitao Sang, Changsheng Xu, and Yan Liu. A unified framework of latent feature learning in social media. *IEEE Transactions on Multimedia*, 16(6):1624–1635, 2014.

[48] Jingyuan Zhang, Bokai Cao, Sihong Xie, Chun-Ta Lu, Philip S. Yu, and Ann B. Ragin. Identifying connectivity patterns for brain diseases via multi-side-view guided deep architectures. In *Proceedings of the SIAM International Conference on Data Mining*, pages 36–44, 2016.

[49] Yuting Zhang, Kibok Lee, Honglak Lee, and UMICH EDU. Augmenting supervised neural networks with unsupervised objectives for large-scale image classification. In *Proceedings of the 33rd International Conference on Machine Learning*, 2016.

[50] Jinfeng Zhuang, Ivor W Tsang, and Steven CH Hoi. Two-layer multiple kernel learning. In *AISTATS*, pages 909–917, 2011.