Deep Graph Similarity Learning for Brain Data Analysis

Guixiang Ma
Intel Labs
guixiang.ma@intel.com

Nesreen K. Ahmed
Intel Labs
nesreen.k.ahmed@intel.com

Theodore L. Willke
Intel Labs
ted.willke@intel.com

Dipanjan Sengupta
Intel Labs
dipanjan.sengupta@intel.com

Michael W. Cole
Rutgers University
michael.cole@rutgers.edu

Nicholas B. Turk-Browne
Yale University
nicholas.turk-browne@yale.edu

Philip S. Yu
University of Illinois at Chicago
psyu@uic.edu

ABSTRACT

We propose an end-to-end graph similarity learning framework called *Higher-order Siamese GCN* for multi-subject fMRI data analysis. The proposed framework learns the brain network representations via a supervised metric-based approach with siamese neural networks using two graph convolutional networks as the twin networks. Our proposed framework performs higher-order convolutions by incorporating higher-order proximity in graph convolutional networks to characterize and learn the community structure in brain connectivity networks. To the best of our knowledge, this is the first community-preserving graph similarity learning framework for multi-subject brain network analysis. Experimental results on four real fMRI datasets demonstrate the potential use cases of the proposed framework for multi-subject brain analysis in health and neuropsychiatric disorders. Our proposed approach achieves an average AUC gain of 75% compared to PCA, an average AUC gain of 65.5% compared to Spectral Embedding, and an average AUC gain of 24.3% compared to S-GCN across the four datasets, indicating promising applications in clinical investigation and brain disease diagnosis.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning approaches**; **Neural networks**.

KEYWORDS

Metric learning, Graph Neural Networks, Graph Convolutional Networks, Higher-order Networks, Brain Networks, Functional Connectivity, Graph Similarity, Structural Similarity, Community Embedding, Graph Matching, Community-Preserving Embedding

1 INTRODUCTION

Learning a similarity metric has gained much attention recently in a variety of real-world applications, where the goal is to learn a function that maps input patterns to a target space while preserving the semantic distance in the input space [8, 17, 18, 23, 50]. For example, finding images that are similar to a query image is an indispensable problem in search engines [50], and an effective image similarity metric is the key for finding similar images. Similarity learning is particularly suitable for applications where the number of classes is very large, the number of samples is very small, and/or only a subset of classes is known at the time of training (e.g., face recognition applications). Metric learning systems are typically trained on pairs of patterns from the training set (e.g., pairs of images, graphs, etc), with the goal to find a function that maps the input patterns to a target space such that the distance among patterns in the target space approximates the semantic distance in the input space. Given a family of possible models (functions) with parameter weights $\mathbb{W}$, the goal is to find the value of optimal weights $\mathbb{W}$, such that the estimated similarity score of the input pair (X_1, X_2) is small if X_1 and X_2 belong to the same class, and large if they belong to different classes. The loss function used in training maximizes the mean similarity between input pairs belonging to the same class, and minimizes the mean similarity between pairs belonging to different classes.

While the majority of the similarity learning research is focused on images, our work focuses instead on learning a similarity metric for neuroimages, such as fMRI images and DTI images. Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes associated with blood flow, and it has been widely used in cognitive neuroscience, medical, and clinical applications. We particularly focus on brain connectivity networks (i.e., functional networks) extracted from the input fMRI images, where the observed brain activity is modeled as a network of inter-regional functional associations. The resulting functional networks can be

analyzed/modeled using graph-theoretic and graph machine learning techniques. The analysis of brain connectivity networks has revealed many key features of the brain organization, such as the modular structure of the brain, and the presence of hubs [11, 37].

Graph convolutional neural networks (GCNs) have emerged to learn deep representations of graph-structured data and they have been shown to outperform other traditional relational learning methods [19, 30]. However, these works focus mainly on social and information networks, to model the node/graph similarity, where the relationships/links are usually given or well-defined (e.g., friendship, retweet, following links). In this paper, however, we focus on learning deep representations from fMRI brain connectivity networks, where each brain network represents the brain activity patterns of a particular subject. Brain connectivity networks are widely used to model inter-regional functional connections and are typically inferred from the input fMRI data of each subject. Each brain connectivity network is usually represented by a graph constructed from input fMRI data where the nodes represent brain regions and the edges represent the pairwise region-to-region correlations. However, the organization of brain connectivity networks is constrained by various factors, such as the underlying brain anatomical network [24, 26], which has an important contribution in shaping the activity across the brain. These constraints pose significant challenges in characterizing the structure and organization of brain connectivity networks. In addition, these constraints lead to distance-dependent inter-regional correlations, where activity patterns from nearby brain regions are strongly correlated due to spatial factors than the activity from distant ones. Thus, it's crucial to distinguish between long-range functional connectivity that represents the brain inter-regional communications, which arise from functional necessity, from those that arise due to spatial constraints [26]. We propose an end-to-end similarity learning framework for multi-subject fMRI data. The proposed framework learns the brain network representations via a supervised metric-based approach with siamese neural networks and using two GCNs as the twin networks. Our proposed framework performs higher-order convolutions by incorporating higher-order proximity via random walks in graph convolutional networks. We use this approach to characterize the community and organizational structure in brain connectivity networks, and to emphasize the contribution of the long-range functional connectivity. To the best of our knowledge, this is the first community-preserving similarity learning framework for multi-subject brain network analysis.

The principal aim of this work is to build a framework that can learn discriminative structural features of brain connectivity networks for multi-subject analysis. However, the complex structure of these networks poses significant challenges to such a framework. For example, prior work has shown the importance of higher-order proximity structure such as the community structure of brain networks in health as well as neuropsychiatric disorders [33, 34]. However, current graph neural methods are capable of only learning local neighborhood information and cannot directly capture the higher-order proximity and community structure of graphs. Although it might be possible for GCNs to learn these higher-order patterns under certain conditions, for example, stacking more hidden GCN layers to allow more information to flow across the graph, these conditions are unlikely to work well in practice, which is due

to the non-linearity of the GCN layers, and the difficulty of fitting the larger number of parameters with a limited number of training samples (e.g., the limited samples in neuroimaging data) which leads to overfitting (see Figures 5 and 6). Thus, how to generalize GCNs to capture the community structure of brain networks is a key challenge. Moreover, how to leverage the structural features learned by GCNs for similarity learning and build an end-to-end similarity learning framework for multi-subject analysis is another challenging problem. Our proposed framework leverages higher-order graph convolutional networks to characterize the community and organizational structure of brain connectivity networks. We conjecture this framework would be useful for the multi-subject brain analysis in health and neuropsychiatric disorders [22]. Our main contributions are summarized as follows:

- We propose an end-to-end similarity learning framework called "Higher-order Siamese GCN" for metric learning of graph data. The framework learns the brain network representations via a supervised metric-based approach with siamese neural networks and using two GCNs as the twin networks. The framework leverages higher-order graph convolutions to characterize the community structure of brain connectivity networks.
- We apply the proposed framework on four real fMRI brain network datasets for similarity learning with respect to brain health status and cognitive abilities. The experimental results demonstrate the effectiveness and advantage of the proposed framework for neurological disorder analysis with respect to three kinds of disease (i.e., Autism, Bipolar, and HIV) and the cognitive analysis for the Human Connectome Project.
- Our proposed approach achieves an average AUC gain of 75% compared to PCA, 65.5% compared to Spectral Embedding, and 24.3% compared to S-GCN across a variety of datasets, indicating its promising use cases for multi-subject brain analysis in health and neuropsychiatric disorders.

2 SIMILARITY LEARNING FRAMEWORK

We propose an end-to-end framework called *higher-order Siamese GCN* for learning the similarity among fMRI brain connectivity networks extracted from multiple subjects. We assume each brain network is represented by a graph where the nodes represent brain regions and the edges represent the inter-regional associations. Our proposed framework (presented in Fig. 1) consists of the following: (1) a Siamese neural network architecture, (2) two identical higher-order graph convolutional networks used as the twin networks, (3) a dot product layer, and (4) a fully connected layer. The framework takes an input pair of graphs, such that each graph is processed by one branch of the siamese network (i.e., each graph is processed by one of the identical twin GCNs). Since the twin graph neural networks share the same weights, each input graph is guaranteed to be processed in the same manner by the networks. As such, similar input graphs would be embedded similarly in the latent space. Our proposed framework performs higher-order convolutions by incorporating higher-order proximity using random walks into graph convolutional networks. Finally, the dot product layer combines the outputs from the two branches of the siamese network, followed
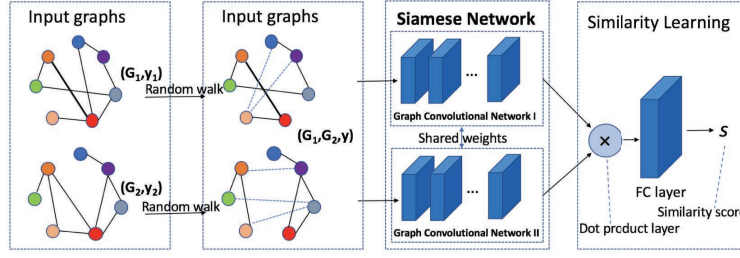
Figure 1: The Higher-order Siamese-GCN framework

by a single fully connected (FC) layer with one output to estimate the similarity score.

Given a multi-subject fMRI data set $\mathcal{G} = \{G_1, G_2, \cdots, G_N\}$ where $G_i = (V_i, E_i, \mathbf{A_i})$ is the fMRI brain network of subject i, V_i is the set of vertices in G_i, $E_i \subset V_i \times V_i$ is the set of edges in G_i, and $\mathbf{A_i} \in \mathbb{R}^{m \times m}$ is the affinity matrix of G_i, we assume that we have a model $\mathcal{M}$ with parameter weights $\mathbb{W}$ to map the brain networks to the feature/embedding space and the embedding of G_i can be obtained as $H_i = \mathcal{M}(A_i)$. Assume we have a function $\mathcal{S}$ for computing similarity between the embeddings of two brain networks, and the similarity score between a pair of brain networks G_i and G_j is denoted by $s_{ij} = \mathcal{S}(H_i, H_j)$, the goal of similarity learning for the multi-subject fMRI data is to learn the optimal parameter weights $\mathbb{W}$ for $\mathcal{M}$ such that the value of s_{ij} is large for an input pairs (G_i, G_j) belonging to the same class, and small for input pairs (G_i, G_j) belonging to different classes. The loss function used in training should maximize the mean similarity between input pairs belonging to the same class and minimize the mean similarity between pairs belonging to different classes.

From fMRI data to brain connectivity networks. A whole-brain fMRI image consists of a sequence of 3D brain image scans, where each volume consists of hundreds of thousands of voxels. Each voxel has an intensity value that is proportional to the strength of the Nuclear Magnetic Resonance (NMR) signal emitted at the corresponding location in the brain volume. The raw form of an fMRI brain image is a discrete time series of 3D images. In order to convert the original fMRI images to region-by-region brain networks, we extract a sequence of responses from each of the regions of interest (ROI), where each ROI represents a brain region. We perform standard fMRI brain image processing steps, including functional image realignment, slice timing correction, and normalization, etc. Then, we compute the region-to-region brain activity correlations. We only keep the positive correlations as the links among the brain regions. The final constructed network is a graph where the nodes/vertices represent brain regions and the edges are the region-to-region correlations.

Graph Convolutional Networks. Graph convolutional network (GCN), as a generalized convolutional neural network from grid-structure domain to graph-structure domain, has emerged as a powerful approach for graph mining [14, 19]. In GCNs, filters are defined in the graph spectral domain. In this paper, we employ the GCN with fast spectral filtering proposed in [19]. Given a graph $G = (V, E, \mathbf{A})$, where V is the set of vertices, $E \subset V \times V$ is the set of

edges, and $\mathbf{A} \in \mathbb{R}^{m \times m}$ is the adjacency matrix, the diagonal degree matrix $\mathbf{D}$ will have elements $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The graph Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{A}$, which can be normalized as $\mathbf{L} = \mathbf{I}_m - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{I}_m$ is the identity matrix. Assume the orthonormal eigenvectors of $\mathbf{L}$ are represented as $\{u_l\}_{l=0}^{m-1} \in \mathbb{R}^{m \times m}$, and their associated eigenvalues are $\{\lambda_l\}_{l=0}^{m-1}$, the Laplacian is diagonalized by the Fourier basis $[u_0, \cdots, u_{m-1}](= \mathbf{U}) \in \mathbb{R}^{m \times m}$ and $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathbf{T}}$ where $\mathbf{\Lambda} = diag([\lambda_0, \cdots, \lambda_{m-1}]) \in \mathbb{R}^{m \times m}$. The graph Fourier transform of a signal $x \in \mathbb{R}^m$ can then be defined as $\hat{x} = \mathbf{U}^{\mathbf{T}} x \in \mathbb{R}^m$ [44]. Suppose a signal vector $\mathbf{x} : V \to \mathbb{R}$ is defined on the nodes of graph G, where $\mathbf{x}_i$ is the value of $\mathbf{x}$ at the ith node. Then the signal $\mathbf{x}$ can be filtered by g_θ as

$$y = g_\theta * \mathbf{x} = g_\theta(\mathbf{L})\mathbf{x} = g_\theta(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathbf{T}})\mathbf{x} = \mathbf{U}g_\theta(\Lambda)\mathbf{U}^{\mathbf{T}}\mathbf{x} \quad (1)$$

where the filter $g_\theta(\Lambda)$ can be defined as $g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k$, and the parameter $\theta \in \mathbb{R}^K$ is a vector of polynomial coefficients [19]. The above filter is exactly K-localized, which means the nodes with shortest path length greater than K are not considered for the convolution. To further reduce computational complexity, we use the Chebyshev polynomials [19] which can be computed recursively by $T_k(\mathbf{x}) = 2\mathbf{x}T_{k-1}(\mathbf{x}) - T_{k-2}(\mathbf{x})$ with $T_0 = 1$ and $T_1 = \mathbf{x}$, and a filter of order $K - 1$ is parameterized as the truncated expansion

$$g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\hat{\Lambda}) \quad (2)$$

Then the filtering operation can be written as $y = g_\theta(\mathbf{L})\mathbf{x} = \sum_{k=0}^{K-1} \theta_k T_k(\hat{L})\mathbf{x}$, where $T_k(\hat{L}) \in \mathbb{R}^{n \times n}$ is the Chebyshev polynomial of order k with the Laplacian $\hat{L} = 2L/\lambda_{max} - I_n$. The jth output feature map of sample s is then given by

$$y_{s,j} = \sum_{i=1}^{F_{in}} g_{\theta_{i,j}}(L)x_{s,i} \in \mathbb{R}^m \quad (3)$$

where $x_{s,i}$ is the input feature map, and F_{in} represents the number of input filters. The $F_{in} \times F_{out}$ vectors of Chebyshev coefficients $\theta_{i,j} \in \mathbb{R}^K$ are the layer's trainable parameters. Then we build the GCN by stacking multiple convolutional layers in the form of Equation (3), with a non-linearity activation (ReLU) following each layer.

Higher-order Graph Convolutional Networks. As discussed above, GCN networks use spectral filterings, which consider localized convolutions while ignoring vertices with shortest path length beyond a threshold. This limits the ability of traditional GCNs to

learn the complex structural patterns of graphs, such as the community structure, motif patterns, higher-order proximity, etc. Recently, high-order structural patterns were shown to be crucial for learning more accurate graph representations [3–5, 10, 41, 42].

Thus, we modify the traditional GCN model to operate directly on graphs with higher-order proximity representations so that the higher-order information will be incorporated during the convolutions. Specifically, we employ a random walk sampling process on graphs to obtain the higher-order proximity representations. Random walk sampling has been widely used for learning graph embeddings that capture the higher-order proximity and community structure in the graph [5, 36]. In our framework, the random walk strategy helps incorporate the higher-order structural information into the graph representations, which further allows for the higher-order graph convolutions in GCNs to capture community and organizational structure of brain networks.

We denote a random walk rooted at vertex v_i as W_{v_i}, which is a stochastic process with random variables $W_{v_i}^1, W_{v_i}^2, \cdots, W_{v_i}^n$, and $W_{v_i}^{j+1}$ is a vertex chosen randomly from the neighbors of vertex W_{v_j}. Given a graph G, the random walk generator samples uniformly a random vertex v_i as the root of the random walk. The walk uniformly samples a vertex from the neighbors of the root, after which it continues sampling from the neighbors of the last vertex visited until the maximum walk length is reached. There could be multiple walks starting from each vertex, depending on the number of walks specified[36]. Line $2 - 9$ in Algorithm 1 illustrates the random walk sampling we use for capturing higher-order proximity information. Note that we slide a window with size w on each walk generated and record the frequency of nodes that co-occur within a window in $\mathbf{F}$, and w is a parameter that controls the extent of the higher-order proximity considered in the graph representation. After we obtain the adjacency matrix $\mathbf{A}$ in Line 12 of Algorithm 1 for the higher-order proximity representation, we use $\mathbf{A}$ for computing the Laplacian matrix $\mathbf{L}$, and use this $\mathbf{L}$ for the spectral filtering in Equation (1) to enable the higher-order graph convolutions.

Algorithm Description. Algorithm 1 shows the overall process of the Higher-order Siamese GCN approach. As the spectral graph convolutional networks filter signals are defined on a common graph Laplacian for all samples, we first estimate the mean functional connectivity matrix among the training samples by computing the mean affinity matrix over the graphs and obtain the k-nn graph $\bar{G}$ using the correlation distance between region pairs. Then we obtain the binary adjacency matrix $\bar{\mathbf{A}}$ for $\bar{G}$ as its graph representation. In order to obtain the higher-order proximity representation of the graph, we apply a random walk on $\bar{G}$, after which we slide a window with size w on each walk to get the co-occurrence frequency between two nodes and record the frequency in the matrix F. After the random walk sampling stage, we obtain the k-nn graph G' based on F, where the edges encode the higher-order proximity of the original graph $\bar{G}$. Then we merge the edges of G' into $\bar{G}$ to get the updated graph representation with higher-order proximity and obtain the adjacency matrix of the updated graph as $\mathbf{A}$. We compute the Laplacian matrix of the new graph representation and denote it with $\mathbf{L}$. Now we start the model learning of Higher-order Siamese GCN. We first prepare the pairs of training samples from $\mathcal{G}$ with label $Y_{ij} = 1$ for each pair of same class and $Y_{ij} = -1$ for each pair

---

Algorithm 1 Higher-order Siamese GCN

Input: $\mathcal{G} = G_1, G_2, \cdots, G_n$ (training graph samples); $\mathbf{y}$(class labels); random walk parameters: γ (number of walks), l (walk length), w (window size)

1: Obtain the mean k-nn graph $\bar{G}(V, E, \bar{\mathbf{A}})$;
2: Initialize a frequency matrix $\mathbf{F} \in \mathbb{R}^{m \times m}$ with zeros;
3: **for** $i = 0$ to γ **do**
4:      $V' = \text{Shuffle}(V)$;
5:      **for each** $v_i \in V'$ **do**
6:          $W_{v_i} = RandomWalk(\bar{G}, v_i, l)$;
7:          Update $\mathbf{F}$;
8:      **end for**
9: **end for**
10: Obtain a k-nn graph G' based on $\mathbf{F}$;
11: Merge the edges of G' into $\bar{G}$;
12: Obtain the updated adjacency matrix $\mathbf{A}$;
13: Prepare pairs of training samples from $\mathcal{G}$;
14: Initialize the parameters Θ of GCNs in Siamese network;
15: **while** not converge **do**
16:      Perform spectral filterings according to Equation (3);
17:      Compute the similarity estimate s_{ij} for the input pair (G_i, G_j);
18:      Compute the loss L^{hinge} according to Equation (4) ;
19:      Apply stochastic gradient descent with ADAM optimizer to update Θ;
20: **end while**

---

with different classes. We also initialize the neural network parameters Θ of the Siamese GCNs. Then we input the pairs into the twin networks and perform spectral convolutions according to Equation (3). Specifically, for each input brain network, we use the functional correlation value of node v_s with v_i as the signal $x_{s,i}$ in Equation (3). After the convolutions, the outputs of the twin GCNs are then combined by a dot product layer, followed by a fully connected (FC) layer that integrates the information learned from the preceding filters and output a similarity estimate s_{ij} for the input pair G_i and G_j. Then we compute the loss for the Siamese network. We use the Hinge loss in Equation (4), by minimizing which the mean similarity between input pairs belonging to the same class will be maximized while the mean similarity between pairs belonging to different classes will be minimized. We use stochastic gradient descent with Adaptive Moment Estimation (ADAM) optimizer to train the model.

$$L^{hinge} = \frac{1}{N_p} \sum_{i=1}^{N} \sum_{j=i+1}^{N} max(0, 1 - Y_{ij}s_{ij}), \qquad (4)$$

where N is the total number of subjects in the training set, and $N_p = N(N-1)/2$ is the total number of pairs from the training set.

3 EXPERIMENTS & RESULTS

We evaluate the performance of the proposed framework on four real resting-state fMRI brain datasets, and we compare with the state-of-the-art baselines. There are mainly two categories of neuroimaging applications involved in the evaluation: (1) medical imaging analysis, which aims to distinguish brain disordered subjects and healthy controls; (2) the cognitive analysis using Human Connectome Project data [49], where the goal is to learn similarity with respect to cognitive abilities from the brain networks. It is worth

mentioning that the empirical similarity learning on fMRI brain networks for cognitive analysis is not well-studied [12] and our work provides new insights and capabilities into the problems in this area.

3.1 Datasets and Preprocessing

- *Autism Brain Imaging Data Exchange (ABIDE)*: This dataset is provided by the ABIDE initiative [20]. It has the resting-state fMRI images of 70 patients with autism spectrum disorder (ASD) and 102 healthy controls, acquired from the largest data acquisition site involved in that project. The preprocessing of the fMRI data includes slice timing correction, motion correction, band-pass filtering, and registering to standard anatomical space. After the preprocessing, a brain network with 264 nodes was constructed for each subject by computing the Pearson correlation between the fMRI time series of the 264 putative regions.

- *Human Connectome Project (HCP)*: This dataset consists of resting-state fMRI imaging data and behavioral data of 100 healthy volunteers from the publicly available Washington University - Minnesota (WU-Min) Human Connectome Project (HCP) [49]. The preprocessing of the fMRI data consists of intensity normalization, phase-encoding direction unwarping, motion, correction, spatial normalization to standard template, and artifact removing[46]. After preprocessing, for each subject, BOLD time series were extracted from the 360 parcels, and functional connectivity network with 360 nodes was constructed for each individual. In this work, we are interested in solving the pair classification problem with respect to cognitive traits. Since this dataset does not have class labels of cognitive traits, we use three key cognitive features from the participants' behavioral data, including Executive Function (Flanker Task), Fluid Intelligence (Penn Progressive Matrices) and Working Memory (List Sorting) [1], and apply K-means clustering with the three features to cluster the subjects into 2 groups. Fig. 2 shows the plot of the 2 clusters in the 3-dimension feature space, where there is a clear boundary between the two clusters. We label the subjects in the two clusters as Class 1 and Class 2, respectively, and prepare pairs of subjects based on this class label for the pair classification evaluation.

- *Bipolar*: This dataset consists of the fMRI data of 52 bipolar I subjects who are in euthymia and 45 healthy controls with matched demographic characteristics [16]. The brain networks were constructed with the CONN[2] toolbox [52]. The raw images were realigned and co-registered, followed by the normalization and smoothing steps. Then the confound effects from motion artifact, white matter, and CSF were regressed out of the signal. Finally, the brain networks were created using the signal correlations between each pair of regions among the 82 labeled Freesurfer-generated cortical/subcortical gray matter regions.

- *Human Immunodeficiency Virus Infection (HIV)*: This dataset is collected from the Chicago Early HIV Infection Study at Northwestern University [38]. It contains the resting-state fMRI data of 77 subjects, 56 of which are early HIV patients and the other 21 subjects are seronegative controls. We use the DPARSF toolbox[3] to process the fMRI data. The images were realigned to the first volume, followed by the slice timing correction and normalization. We focus on the 116 anatomical volumes of interest (AVOI) and construct a brain network with the 90 cerebral regions, where each node in the graph represents a cerebral region, and links are created based on the correlations between different regions.

3.2 Baselines and Metrics

We compare our Higher-order Siamese GCN framework with three other baseline methods for two classification tasks based on the similarity learning on brain networks, and we use AUC and accuracy as the evaluation metrics.

- **PCA** is the Principal Component Analysis approach that is widely used for dimension reduction and feature extraction [45]. We apply PCA on the correlation matrices of the brain networks and perform similarity learning based on the PCA results.

- **SE** is the Spectral Embedding approach, which finds a low dimensional representation of the data using a spectral decomposition of the graph Laplacian [9].

- **Siamese GCN (S-GCN)** is the method proposed in [29], which employs the traditional GCN to capture localized structural information of graphs and learn similarity scores between graphs based on the outputs of GCNs in a Siamese network. This was the first work of applying graph convolutional neural network on brain connectivity networks.

- **Higher-order Siamese GCN (HS-GCN)** is the proposed approach in this paper.

Experimental Setup. We use 60% of the data for training and the other 40% for testing. We use the experiment settings for the S-GCN following the instructions provided in [29] and use 5-fold cross validation for optimizing the parameters in both S-GCN and HS-GCN. For HS-GCN, we use 2 GCN layers with $f = 32$ features for each. We use the stochastic gradient descent with ADAM algorithm [27] for the optimization, where we set learning rate to be 0.001 and use $K = 3$ for the polynomial filters in the spectral filtering. We set the dropout rate at the fully connected layer as 0.8 and use 0.0005 for the regularization parameter. For the constrained variance loss in Equation (6) used in the subject classification task, we set $a = m/2$ for both datasets. For the parameters in random walk, we employ the grid search in a range of values to find the optimal parameter values. We fix the number of walks γ to be 10, and search the value for the walk length l from $[30, 40, \ldots, 100]$ and the value for window size w from $[1, 2, \ldots, 10]$. For the k-nn graph construction stage, we use 10% of the number of nodes in the brain networks as the value for k. For PCA and SE, the only parameter is the number of components to be preserved in the output lower-dimensional representation. We employ the grid search in the range

---

[1]https://www.humanconnectome.org/study/hcp-young-adult/document/quick-reference-open-access-vs-restricted-data
[2]http://www.nitrc.org/projects/conn

[3]http://rfmri.org/DPARSF.

of $[5, 10, 15, \ldots, 60]$ to find the optimal value for that parameter. After we obtain the output representations from PCA and SE, we calculate the similarity score for each pair according to Equation (5) [21].

$$s_{ij} = 1 - \sqrt{\text{Tr}\left(\left(\mathbf{F}_i - \mathbf{F}_j\right)^T \left(\mathbf{F}_i - \mathbf{F}_j\right)\right)} \quad (5)$$

where $\mathbf{F}_i$ and $\mathbf{F}_j$ are the PCA results of subject i and j, respectively. For each experiment, we run for 5 times and report the average results.

3.3 Evaluations

We evaluate the performance of the proposed framework in similarity learning of brain networks by applying it in two classification tasks: (1) Pair classification, and (2) Subject classification.
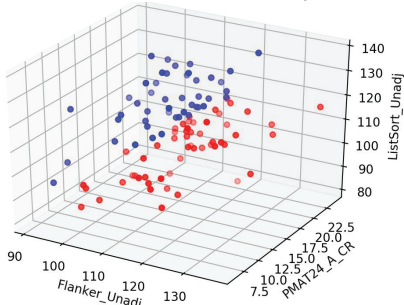

Figure 2: Clustering the HCP subjects using the three major cognitive features.

Table 1: AUC Scores of Pair Classification (mean ± std).

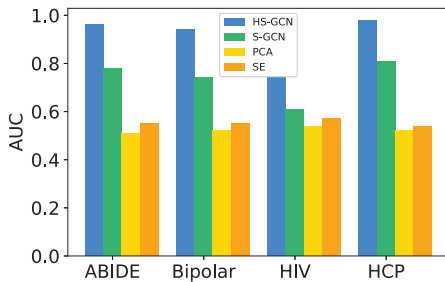| Methods	ABIDE	HCP	HIV	Bipolar
PCA	0.51 ± 0.01	0.52 ± 0.01	0.54 ± 0.07	0.52 ± 0.01
SE	0.55 ± 0.02	0.54 ± 0.01	0.57 ± 0.02	0.55 ± 0.01
S-GCN	0.78 ± 0.29	0.81 ± 0.36	0.61 ± 0.25	0.74 ± 0.19
HS-GCN	$\mathbf{0.96 \pm 0.02}$	$\mathbf{0.98 \pm 0.03}$	$\mathbf{0.77 \pm 0.20}$	$\mathbf{0.94 \pm 0.07}$

Figure 3: AUC Scores of Pair Classification.

Pair Classification. Pair classification refers to the classification of similar pairs (brain networks from the same class) versus dissimilar pairs (brain networks from different classes) based on the similarity learned by the model. This is a very important task in brain connectivity analysis, especially for the brain disorder identification problem when there is a very limited number of labeled

samples, which is a common scenario in biomedical data mining. If we could build a powerful model that can distinguish similar pairs and dissimilar pairs well, then we could use it for predicting unseen samples by looking into their similarity with the labeled samples. With the Siamese architecture, the proposed model has inherent advantages for this task. For instance, given a training set with 100 samples, we could generate 4950 unique pairs of brain networks that could be used as inputs for the Higher-order Siamese GCN model, greatly improving the training effectiveness and robustness of the model for predicting unseen samples. Table 1 and Fig. 3 shows the classification AUC of the Higher-order Siamese GCN and that of the baseline methods on pair classification on the four datasets.

As shown in Table 1, the classification AUC of HS-GCN is significantly higher than that of the baseline methods. More specifically, our proposed HS-GCN achieves an average AUC gain of 75% compared to PCA, an average AUC gain of 65.5% compared to Spectral Embedding, and an average AUC gain of 24.3% compared to S-GCN across all datasets. Our proposed HS-GCN is more accurate and has a lower variance compared to S-GCN. The PCA based approach achieved the lowest AUC scores. This is probably due to the fact that PCA learns lower-dimensional feature representations directly from the correlation matrix while not considering the structural information of the graph. The Spectral Embedding based approach achieved higher AUC than PCA. Since SE is designed for spectral clustering, which means the embedding results of SE encodes the community structure of graphs, this property would help discriminate the brain networks of different classes. But since SE is not able to capture the complex local structure of brain networks, and it is not an end-to-end approach for similarity learning, the AUC scores of SE is much lower than the proposed approach. The S-GCN model achieved fairly good results on ABIDE and HCP datasets, but the AUC scores are still not as good as that of HS-GCN. The superior performance of our proposed HS-GCN framework indicates that the localized structural information and community structure learned by higher-order graph convolutions did benefit the similarity learning of brain networks. Moreover, the HS-GCN achieved the best results on all the four datasets, demonstrating its generalizing ability in similarity learning of brain networks.
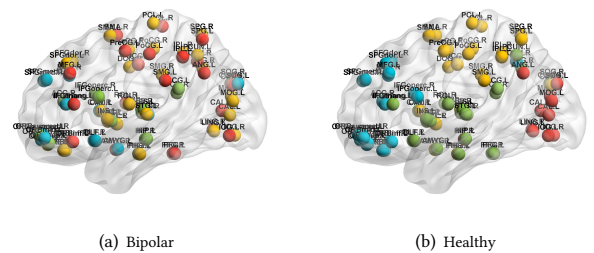

(a) Bipolar

(b) Healthy

Figure 4: Visualization of the community structure captured by HS-GCN in healthy and bipolar disease networks. Notably, this figure highlights the reduced functional connectivity, as shown by decreased clustering in the bipolar network.

Evaluation of the Community Structure in health and disease. To evaluate the effectiveness of HS-GCN for capturing the
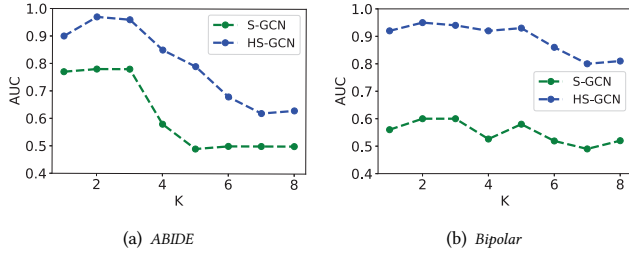
(a) *ABIDE*

(b) *Bipolar*

Figure 5: Pair classification AUC of S-GCN and HS-GCN with different values for K

(a) *ABIDE*

(b) *Bipolar*
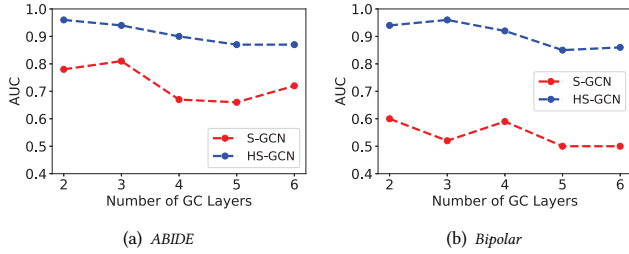
Figure 6: Pair classification AUC of S-GCN and HS-GCN with different numbers of GC layers

(a) *Hinge Loss*

(b) *Constrained Variance Loss*

Figure 7: Accuracy for Subject classification on ABIDE and Bipolar with two different loss functions

organization and community structure of brain networks, we investigate the resulting brain network embedding by the higher-order GCN. For each brain network, we cluster the brain regions (nodes) using K-means clustering with $k = 4$, and using their embedding feature vectors in K-means. Fig. 4 shows a visualization [4] of the community structure of a Bipolar subject versus a healthy control, and nodes are colored by their corresponding community. From Fig. 4, we observe that the nodes of the healthy brain network are well-grouped into the four clusters, while the cluster boundaries in the Bipolar subject's brain network are blurry and the nodes widely spread out over different clusters. We conjecture this observation is due to the reduced functional connectivity as shown by decreased clustering in bipolar subjects versus healthy controls (consistent with prior observations in [7, 48]).

Impact of Higher-order Proximity on GCNs. In the GCN models, two important parameters are the number of graph convolutional (GC) layers and the value K which represents the K-localized neighborhood considered in the convolutions. In order to evaluate the impact of these parameters on the performance of S-GCN and HS-GCN in pair classification, we apply the two models with various values for K and the number of GC layers. Fig. 5 and Fig. 6 show the evaluation results. For the experiments in Fig. 5, we use two GC layers for both models, and evaluate the models with K values ranging from 1 to 8. For the experiments in Fig. 6, we fix K as 3 and vary the number of GC layers from 2 to 6. We set the random walk parameters with the optimal values we found in the experiments above. We can see from Fig. 5 that when the K value increases, the performance of S-GCN first goes up, and when K goes beyond 3,
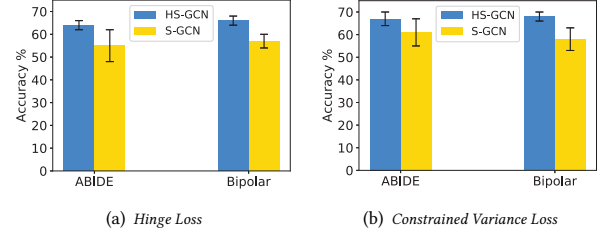
---

[4]We used Brainnet Viewer toolbox [53]

the AUC scores start to decline. The performance of HS-GCN has a similar trend, but it is always better than S-GCN. This indicates that incorporating a larger range of localized neighborhood information in the graph convolutions cannot improve the performance of S-GCN as the higher-order property did in HS-GCN. Meanwhile, as we observe in Fig. 6, adding more GC layers did not improve the learning ability of S-GCN either. This is mainly because adding more layers also introduces more hyper-parameters into the model, which could cause overfitting.

Subject Classification. In this experiment, we use the pairwise similarity learned by the model to further classify the subjects with brain disorder versus healthy controls. We evaluate the proposed HS-GCN model and the baseline S-GCN model on the ABIDE and Bipolar datasets. We apply the weighted k-nearest neighbor (kNN) classifier [25] with the similarity scores we learned for the classification task. The class label of one subject is determined based on a weighted combination of the labels of its neighbors. Here we consider all the neighbors with positive similarity scores in the weighted calculation. Meanwhile, we explore the influence of different loss functions in the subject classification performance. Besides evaluating the two models with the Hinge loss in Equation (4), we also evaluate them with the following constrained variance loss:

$$L^{convar} = max(0, \sigma^{2+} - a) + max(0, \sigma^{2-} - a) + max(0, m - (\mu^+ - \mu^-)),$$
(6)

where μ^+ represents the mean similarity between embeddings belonging to the same class, and μ^- represents the mean similarity between embeddings belonging to different classes, while σ^{2+} and σ^{2-} refer to the variance of pairwise similarity for the same class and different classes, respectively. m is the margin between the means of the same-class and different-class similarity distributions, and a is the variance threshold. This loss function is proposed by [29]. By this formulation, the variance is only penalized when it exceeds the threshold a, which allows the similarity estimates to vary around the means, thus could be used to accommodate the diversity that usually exists in fMRI data due to the varied factors in the acquisition process.

Fig. 7 shows the evaluation results of subject classification by the two models with different loss functions. As shown in the figure, the proposed HS-GCN model achieves a higher accuracy with both loss functions on both datasets compared to the baseline S-GCN model. This implies that the similarity scores learned by
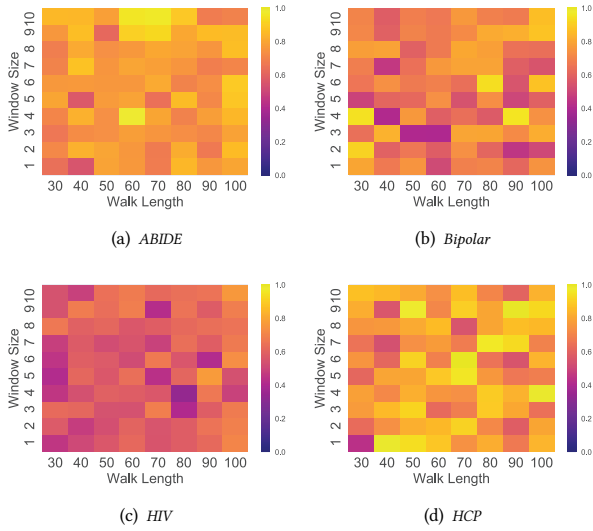
(a) *ABIDE*

(b) *Bipolar*

(c) *HIV*

(d) *HCP*

Figure 8: AUC scores for Pair classification with different values for Random Walk parameters

the proposed model are more accurate, and thus more reliable to be used for further multi-subject brain connectivity analysis. By comparing Fig. 7(a) and Fig. 7(b), we find that both models get higher classification accuracy with the similarity scores learned by the constrained variance loss. This could be the benefit of allowing for more diversity across the samples by the constrained variance loss.

Parameter Analysis. To analyze the influence of the parameters in the random walk process on the similarity learning performance of HS-GCN, we perform a parameter sensitivity evaluation for the two key parameters in random walk: the walk/path length l and the window size w. The bar plots in Fig. 8 show the AUC scores of the proposed model with different values for l and w. We observe that the AUC scores vary across different parameter settings. The highest AUC scores on all datasets tend to come from the cases with longer walk length and larger window size. We conjecture this is due to the importance of the long-range functional connectivity for multi-subject brain analysis. For instance, the best result on ABIDE is achieved when $l = 60$ and $w = 4$, and the best result on HIV is achieved when $l = 90$ and $w = 9$. The selection of the parameter values also relates to the scale of the brain networks involved. Among the four datasets, HCP has the largest number of nodes. The optimal walk length value and window size are also the largest among the four datasets. This is reasonable, as the random walks generated on larger graph tend to be longer than those on smaller graphs assuming the two graphs have the same density. To capture the high-order information from larger graphs, we should slide a relatively large window on the walks as well. Therefore, it is important to consider the scale and other relevant properties of the brain networks when selecting parameter values.

4 RELATED WORK

Similarity learning is at the heart and foundation of various machine learning tasks, e.g., classification, prediction, etc. Most of the existing works in this area focus on images [13, 18, 28, 40], while the similarity learning on graph data is seldom studied [2, 43]. Existing methods for similarity estimation among graphs are mainly based on graph embedding, graph kernels or motifs [15, 32]. These methods are designed for specific scenarios and have their limitations. For example, the graph embedding learned in [1] may discard structural information that could be important for similarity estimation. In [47], the graph kernels used for brain network comparison focus on features of small subgraphs [3, 39], which ignore global structures of graphs. In these works, the graph feature extraction and similarity estimation are done in completely separate stages, where the features extracted may not be suitable for similarity estimation. In this paper, we build an end-to-end similarity learning framework to solve these problems.

Brain network analysis has been an emerging research area[31]. Existing work in brain networks mainly focuses on discovering brain networks from spatio-temporal voxel-level data or mining brain networks for neurological analysis [6, 34, 35, 51]. For example, in [6], an unsupervised matrix tri-factorization method is developed to simultaneously discover nodes and edges of the underlying brain networks in fMRI data. In [51], a CNN-based model is proposed to learn non-linear structures from brain networks for brain disorder diagnosis. Most of these works aim to learn discriminative features from brain networks for the classification or clustering of subjects. However, how to measure the similarity in the graph domain is seldom studied for multi-subject brain network analysis. In [29], the metric learning of brain connectivity networks is first studied, and they use the GCNs to learn features from brain networks. However, they use the traditional GCN model, which cannot capture the higher-order community structure in brain networks important for the analysis of neurological disorders. In our work, we solve this problem by proposing the higher-order GCNs to characterize the community structure of brain networks, which leads to much higher performance in similarity learning compared to the method used in [29].

5 CONCLUSION

We proposed an end-to-end framework called *Higher-order Siamese GCN* (HS-GCN) for learning similarity among fMRI brain networks using higher-order GCNs as the twin networks. The proposed higher-order graph convolutions leverage the higher-order community structure of brain networks. We use this approach to characterize the community and organizational structure in brain networks. To the best of our knowledge, this is the first community-preserving similarity learning framework for multi-subject brain network analysis. Experimental results on four real fMRI datasets demonstrate the potential use cases of the proposed framework for multi-subject brain analysis in health and neuropsychiatric disorders.

ACKNOWLEDGMENTS

REFERENCES

[1] Alexandre Abraham, Michael P Milham, Adriana Di Martino, R Cameron Crad-dock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux. 2017. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage* 147 (2017), 736–745.

[2] Nesreen K Ahmed, Nick Duffield, and Liangzhen Xia. 2018. Sampling for approx-imate bipartite network projection. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence.* AAAI Press, 3286–3292.

[3] Nesreen K Ahmed, Jennifer Neville, Ryan A Rossi, and Nick Duffield. 2015. Effi-cient graphlet counting for large networks. In *ICDM.* IEEE, 1–10.

[4] Nesreen K Ahmed, Jennifer Neville, Ryan A Rossi, Nick G Duffield, and Theodore L Willke. 2017. Graphlet decomposition: Framework, algorithms, and applications. *Knowledge and Information Systems* 50, 3 (2017), 689–722.

[5] Nesreen K Ahmed, Ryan Rossi, John Boaz Lee, Xiangnan Kong, Theodore L Willke, Rong Zhou, and Hoda Eldardiry. 2018. Learning Role-based Graph Embeddings. *ICML workshop on StarAI* (2018).

[6] Zilong Bai, Peter Walker, Anna Tschiffely, Fei Wang, and Ian Davidson. 2017. Unsupervised Network Discovery for Brain Imaging Data. In *SIGKDD.* ACM, 55–64.

[7] Danielle S Bassett and Edward T Bullmore. 2009. Human brain networks in health and disease. *Current opinion in neurology* 22, 4 (2009), 340.

[8] Miguel Ángel Bautista, Artsiom Sanakoyeu, and Björn Ommer. 2017. Deep Unsupervised Similarity Learning using Partially Ordered Sets.. In *CVPR.* 1923–1932.

[9] Mikhail Belkin and Partha Niyogi. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation* 15, 6 (2003), 1373–1396.

[10] Austin R Benson, David F Gleich, and Jure Leskovec. 2016. Higher-order organi-zation of complex networks. *Science* 353, 6295 (2016), 163–166.

[11] Maxwell A Bertolero, BT Thomas Yeo, and Mark D'Esposito. 2015. The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences* 112, 49 (2015), E6798–E6807.

[12] Susan Y Bookheimer et al. 2018. The lifespan Human Connectome Project in aging: An overview. *NeuroImage* (2018).

[13] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a" siamese" time delay neural network. In *NeurIPS.* 737–744.

[14] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2013. Spectral net-works and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203* (2013).

[15] James P Canning, Emma E Ingram, Sammantha Nowak-Wolff, Adriana M Ortiz, Nesreen K Ahmed, Ryan A Rossi, Karl RB Schmitt, and Sucheta Soundarajan. 2018. Predicting graph categories from structural properties. *arXiv preprint arXiv:1805.02682* (2018).

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

[17] Gal Chechik, Uri Shalit, Varun Sharma, and Samy Bengio. 2009. An online algorithm for large scale image similarity learning. In *NeurIPS.* 306–314.

[18] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *CVPR,* Vol. 1. IEEE, 539–546.

[19] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolu-tional neural networks on graphs with fast localized spectral filtering. In *NeurIPS.* 3844–3852.

[20] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castel-lanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry* 19, 6 (2014), 659.

[21] Brendan J Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *science* 315, 5814 (2007), 972–976.

[22] Michael Greicius. 2008. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology* 21, 4 (2008), 424–430.

[23] Guo-Dong Guo, Anil K Jain, Wei-Ying Ma, and Hong-Jiang Zhang. 2001. Learning similarity measure for natural image retrieval with relevance feedback. In *CVPR,* Vol. 1. IEEE, I–I.

[24] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. 2008. Mapping the structural core of human cerebral cortex. *PLoS biology* 6, 7 (2008), e159.

[25] Klaus Hechenbichler and Klaus Schliep. 2004. Weighted k-nearest-neighbor techniques and ordinal classification. (2004).

[26] CJ Honey, O Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106, 6 (2009), 2035–2040.

[27] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic opti-mization. *arXiv preprint arXiv:1412.6980* (2014).

[28] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop,* Vol. 2.

[29] Sofia Ira Ktena, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, and Daniel Rueckert. 2018. Metric learning with spectral graph convo-lutions on brain connectivity networks. *NeuroImage* 169 (2018), 431–442.

[30] Qimai Li, Zhichao Han, and Xiao-Ming Wu. 2018. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. *arXiv preprint arXiv:1801.07606* (2018).

[31] Jin Liu, Min Li, Yi Pan, Wei Lan, Ruiqing Zheng, Fang-Xiang Wu, and Jianxin Wang. 2017. Complex brain network analysis and its applications to brain disorders: a survey. *Complexity* 2017 (2017).

[32] Lorenzo Livi and Antonello Rizzi. 2013. The graph matching problem. *Pattern Analysis and Applications* 16, 3 (2013), 253–283.

[33] Guixiang Ma, Lifang He, Bokai Cao, Jiawei Zhang, S Yu Philip, and Ann B Ragin. 2016. Multi-graph clustering based on interior-node topology with applications to brain networks. In *ECML PKDD.* Springer, 476–492.

[34] Guixiang Ma, Lifang He, Chun-Ta Lu, Weixiang Shao, Philip S Yu, Alex D Leow, and Ann B Ragin. 2017. Multi-view clustering with graph embedding for connec-tome analysis. In *CIKM.* ACM, 127–136.

[35] Guixiang Ma, Chun-Ta Lu, Lifang He, S Yu Philip, and Ann B Ragin. 2017. Multi-view graph embedding with hub detection for brain network analysis. In *ICDM.* IEEE, 967–972.

[36] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *SIGKDD.* ACM, 701–710.

[37] Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. 2011. Functional network organization of the human brain. *Neuron* 72, 4 (2011), 665–678.

[38] Ann B. Ragin, Hongyan Du, Renee Ochs, Ying Wu, Christina L Sammet, Alfred Shoukry, and Leon G Epstein. 2012. Structural brain alterations can be detected early in HIV infection. *Neurology* 79, 24 (2012), 2328–2334.

[39] Ryan A Rossi, Nesreen K Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunyee Koh. 2019. Heterogeneous network motifs. *arXiv preprint arXiv:1901.10026* (2019).

[40] Ryan A Rossi, Nesreen K Ahmed, Hoda Eldardiry, and Rong Zhou. 2018. Similarity-based Multi-label Learning. In *2018 International Joint Conference on Neural Networks (IJCNN).* IEEE, 1–8.

[41] Ryan A Rossi, Nesreen K Ahmed, and Eunyee Koh. 2018. Higher-order Network Representation Learning. In *WWW.* International World Wide Web Conferences Steering Committee, 3–4.

[42] Ryan A Rossi, Rong Zhou, and Nesreen K Ahmed. 2017. Deep feature learning for graphs. *arXiv preprint arXiv:1704.08829* (2017).

[43] Ryan A Rossi, Rong Zhou, Nesreen K Ahmed, and Hoda Eldardiry. 2018. Relational Similarity Machines (RSM): A Similarity-based Learning Framework for Graphs. In *2018 IEEE International Conference on Big Data (Big Data).* IEEE, 1807–1816.

[44] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. 2013. The emerging field of signal processing on graphs: Ex-tending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine* 30, 3 (2013), 83–98.

[45] Lindsay I Smith. 2002. *A tutorial on principal components analysis.* Technical Report.

[46] Marjolein Spronk, Jie Lisa Ji, Kaustubh Kulkarni, Grega Repovs, Alan Anticevic, and Michael W Cole. 2018. Mapping the human brain's cortical-subcortical functional network organization. *bioRxiv* (2018), 206292.

[47] Sylvain Takerkart, Guillaume Auzias, Bertrand Thirion, and Liva Ralaivola. 2014. Graph-based inter-subject pattern analysis of fMRI data. *PloS one* 9, 8 (2014), e104586.

[48] Shin Teng, Chia-Feng Lu, Po-Shan Wang, Cheng-Ta Li, Pei-Chi Tu, Chih-I Hung, Tung-Ping Su, and Yu-Te Wu. 2014. Altered resting-state functional connectivity of striatal-thalamic circuit in bipolar disorder. *PloS one* 9, 5 (2014), e96422.

[49] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. 2013. The WU-Minn human connectome project: an overview. *Neuroimage* 80 (2013), 62–79.

[50] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. 2014. Learning fine-grained image similarity with deep ranking. In *CVPR.* 1386–1393.

[51] Shen Wang, Lifang He, Bokai Cao, Chun-Ta Lu, Philip S Yu, and Ann B Ragin. 2017. Structural deep brain network mining. In *SIGKDD.* ACM, 475–484.

[52] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity* 2, 3 (2012), 125–141.

[53] Mingrui Xia, Jinhui Wang, and Yong He. 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one* 8, 7 (2013), e68910.