Similarity Learning with Higher-Order Proximity for Brain Network Analysis

Guixiang Ma1, Nesreen K. Ahmed2, Ted Willke2, Dipanjan Sengupta3, Michael W. Cole3, Nick Turk-Browne4, Philip S. Yu1

1University of Illinois at Chicago, 2Intel Labs, 3Rutgers University, 4Yale University
gma4@uic.edu, nesreen.k.ahmed@intel.com, ted.willke@intel.com, dipanjan.sengupta@intel.com
mwcole@mwcole.net, nicholas.turk-browne@yale.edu, psyu@uic.edu

Abstract
In recent years, the similarity learning problem has been widely studied. Most of the existing works focus on images and few of these works could be applied to learn similarity between neuroimages, such as fMRI images and DTI images, which are important data sources for human brain analysis. In this paper, we focus on the similarity learning for fMRI brain network analysis. We propose a general framework called Multi-hop Siamese GCN for similarity learning on graphs. This framework provides options for refining the graph representations with high-order structure information, thus can be used for graph similarity learning on various brain network data sets. We apply the proposed Multi-hop Siamese GCN approach on four real fMRI brain network datasets for similarity learning with respect to brain health status and cognitive abilities. Our proposed method achieves an average AUC gain of 82.6% compared to PCA, and an average AUC gain of 42% compared to S-GCN across a variety of datasets, indicating its promising learning ability for clinical investigation and brain disease diagnosis.

Introduction
In many applications, the ability to compute similarity scores between objects is crucial to a variety of machine learning tasks such as classification, clustering and ranking. For example, finding images that are similar to a query image is an indispensable problem in search engines (Wang et al. 2014), and an effective image similarity metric is the key for finding similar images. In the past decade, quite a few of works have been done on similarity learning. In (Bautista, Sanakoyeu, and Ommer 2017), similarity learning is formulated as a partial ordering task with soft correspondences of all samples to classes. Adopting a strategy of self-supervision, a CNN is trained to optimally represent samples in a mutually consistent manner while updating the classes. The similarity learning and grouping procedure are integrated in a single model and optimized jointly. Another deep learning model for ranking is proposed in (Wang et al. 2014) for learning fine-grained image similarity. An efficient triplet sampling algorithm is proposed to learn the model with distributed asynchronous stochastic gradient. The proposed approach is shown to outperform models based on hand-crafted visual features and deep classification models.

Although the similarity learning problem has been widely studied, most existing works focus on images and few of these works could be applied to learn similarity in neuroimaging, such as fMRI images and DTI images, which are important data sources for human brain analysis. In this paper, we focus on the similarity learning for fMRI brain data analysis. Instead of looking at the original fMRI images, we look at the fMRI brain connectivity networks derived from the original fMRI images, as the brain connectivity networks could reflect the overall region-by-region interactions in human brain and have been shown to be an important view to investigate human brains (Bullmore and Sporns 2009). Existing works in this area have shown that the structure information in human brain networks could reflect their brain activity patterns, and people with brain disorders tend to have different patterns with healthy people. The graph representation has privileges in capturing structure and high-order information of human brain. Therefore, we study the similarity learning problem in the graph domain and aim to apply it for fMRI brain connectivity (network) analysis.

Recently, graph convolutional neural networks (GCNs) have drawn much attention to learn useful representations from graph data, and they have been shown to be effective compared to other relational learning methods (Defferrard, Bresson, and Vandergheynst 2016; Li, Han, and Wu 2018). However, these works mainly focus on social and information networks, where the goal is the node-level similarity or relationship analysis. In this paper, however, we focus on learning useful representations from fMRI brain connectivity networks. Our goal is to model the structural similarity for multi-subject neurological or cognitive status analysis. Since there are various structure information in graphs and the brain networks might differ in global structure or local structure, the main challenge of this similarity learning task is how to build a general learning framework that can learn discriminating graph structure features from graphs using GCNs and leverage the graph structure for the similarity learning task. Recently, (Klaen et al. 2018) studied the similarity learning problem on graphs with GCNs and applied it to fMRI brain networks, which is the most relevant work to our work in this paper. However, they did not consider high-order information of graphs when applying GCNs, thus may ignore important graph features in the similarity learning process. In this paper, we propose to incorporate the high-order structural information of graphs into the GCNs, and build a general framework with Siamese network for similarity learning of...
graphs. Specifically, we apply a random walk strategy with sliding windows to capture high-order information on graphs and use it to refine the graph representations, which allows for multi-hop convolutions on graphs. Our contributions can be summarized as follows:

- We propose a general framework called "Multi-hop Siamese GCN" for similarity learning on graphs. This framework provides multiple options for refining the graph representations with high-order structure information, thus can be used for graph similarity learning on various data sets.
- The proposed framework employs random walks with sliding windows to obtain high-order information on graphs and leverage the high-order information for the similarity learning task.
- We apply the Multi-hop Siamese GCN approach on four real fMRI brain network datasets for similarity learning with respect to brain health status and cognitive abilities. The experiment results demonstrate the effectiveness of the proposed framework for similarity learning in brain network analysis.
- Our proposed approach achieves an average AUC gain of 82.6% compared to PCA, and an average AUC gain of 42% compared to S-GCN across a variety of datasets, indicating its promising learning ability for clinical investigation and applications.

Preliminaries

In many machine learning problems where data comes in graphs, a key application is how to measure the similarity between graphs. In the field of brain network analysis, measuring similarity between brain networks is especially important for further analysis such as brain disorder diagnosis. Existing methods for similarity estimation between graphs are mainly based on graph embedding, graph kernels or motifs (Livni and Rizzi 2013). These methods are designed for specific scenarios and have their limitations. For example, the graph embedding learned in (Abraham et al. 2017) may discard structural information that could be important for similarity estimation. In (Takerkart et al. 2014), the graph kernels used for brain network comparison focus on features of small subgraphs, which ignored global structures of graph. Another problem in these works is that the graph feature extraction and similarity estimation are done in completely separate stages, while the features extracted may not be suitable for similarity estimation.

Recently, a metric learning method with spectral graph convolutions is proposed in (Ktena et al. 2018), where a Siamese network with graph convolutional neural networks is used to get similarity estimate of two brain connectivity networks. This method consists of two components: (a) Graph Convolutional Network (GCN) and (b) Siamese Network. This method shows promising results in the experiment of similarity estimate between brain connectivity networks. However, the way that GCN was used in that work focuses on local structure in the graph representation, while ignoring high-order structural information, which makes the method less generic. In this paper, we go beyond of that method and propose a Multi-hop Siamese GCN approach, which is a general framework with different options in each component, and the Multi-hop property of the approach allows us to incorporate high-order structural information from the graph representations into the learning process. In this section, we will first introduce GCN and Siamese network individually. Then we will introduce the proposed Multi-hop Siamese GCN approach, which is a general framework with different options in each component for similarity learning of graphs.

Graph Convolutional Networks

Graph convolutional network (GCN), as a generalized convolutional neural network from grid-structure domain to graph-structure domain, has been emerging as a powerful approach for graph mining (Bruna et al. 2013, Defferrard, Bresson, and Vandergheynst 2016). In GCNs, filters are defined in the graph spectral domain. Given a graph $G = (V, E, A)$, where $V$ is the set of vertices, $E \subset V \times V$ is the set of edges, and $A \in \mathbb{R}^{m \times m}$ is the adjacency matrix, the diagonal degree matrix $D$ will have elements $D_{ii} = \sum_j A_{ij}$. The graph Laplacian matrix is $L = D - A$, which can be normalized as $L = I_m - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, where $I_m$ is the identity matrix. As $L$ is a real symmetric positive semidefinite matrix, it has a set of orthonormal eigenvectors $\{u_i\}_{i=0}^{m-1} \in \mathbb{R}^{m \times m}$, and their associated eigenvalues $\{\lambda_i\}_{i=0}^{m-1}$. The Laplacian is diagonalized by the Fourier basis $[u_0, \ldots, u_{m-1}] \in \mathbb{R}^{m \times m}$ and $L = U \Lambda U^T$, where $\Lambda = \text{diag}(\lambda_0, \ldots, \lambda_{m-1}) \in \mathbb{R}^{m \times m}$. The graph Fourier transform of a signal $x \in \mathbb{R}^m$ can then be defined as $\hat{x} = U^T x \in \mathbb{R}^m$ (Shuman et al. 2013). The transform enables the convolution operation on graph in the Fourier domain. Suppose a signal vector $x: V \rightarrow \mathbb{R}$ is defined on the nodes of graph $G$, where $x_i$ is the value of $x$ at the $i$th node. Then the signal $x$ can be filtered by $g_0$ as

$$y = g_0 \ast x = g_0(L)x = g_0(U \Lambda U)x = U g_0(\Lambda)U^T x \quad (1)$$

where the filter $g_0(\Lambda)$ can be defined as $g_0(\Lambda) = \text{diag}(\theta)$ and the parameter $\theta \in \mathbb{R}^n$ is a vector of Fourier coefficients (Shuman et al. 2013).

In (Bruna et al. 2013), GCN was formulated for the first time, which was parameterised on the eigenvectors of Laplacian. However, their computations of the eigendecomposition are very expensive, and the filters represented in the spectral domain may not be localized in the graph spatial domain. To overcome these issues, (Defferrard, Bresson, and Vandergheynst 2016) proposed to use a polynomial filter by

$$g_0(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k \quad (2)$$

where the parameter $\theta \in \mathbb{R}^K$ is a vector of polynomial coefficients. According to (Defferrard, Bresson, and Vandergheynst 2016), 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, (Defferrard, Bresson, and Vandergheynst 2016) proposed to use the Chebyshev polynomials which can be computed recursively by $T_k(x) = 2x T_{k-1}(x) - T_{k-2}(x)$
with $T_0 = 1$ and $T_1 = x$, and a filter of order $K - 1$ is parameterized as the truncated expansion

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

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

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

where $x_{s,i}$ are the input feature maps, and $F_{in}$ represents the number of input filters. The $F_{in} \times F_{out}$ vectors of Chebyshev coefficients $	heta_{i,j} \in \mathbb{R}^{K}$ are the layer’s trainable parameters.

A graph convolutional neural network can therefore be built by stacking multiple convolutional layers in the form of Eq. [4] with a non-linearity following each layer.

### Siamese Network

Siamese networks were first introduced in (Bromley et al. 1994) to solve image matching problem for signature verification. A Siamese network consists of two twin neural networks who share parameters with each other. The inputs of the twin networks are distinct, but their highest-level feature representations are joined by a function at the top. As the parameters between the twin networks are tied, each input will be processed in the same way respectively in the twin networks, which can guarantee that similar input samples not be mapped to very different locations by the respective networks. Therefore, the Siamese network tend to be good for differentiating the two inputs or measuring the similarity between them. Some existing works on image similarity learning use CNNs in the twin networks of Siamese, as CNNs works well in learning 2D grid features from images (Chopra, Hadsell, and LeCun 2005). In this work, as we aim to learn similarity metric between graphs, we explore the capability of GCNs with Siamese architecture. Figure 1 shows a simple illustration of the Siamese architecture with GCNs in the twin networks.

### Multi-hop Siamese GCN: A General Framework

As introduced above, GCN uses spectral filterings, which consider localized convolutions while ignoring the nodes with shortest path length beyond a threshold. However, high-order structural information are very important for learning from graphs (Rossi, Ahmed, and Koh 2018; Benson, Gleich, and Leskovec 2016; Rossi, Zhou, and Ahmed 2017). In this section, we introduce a general framework for similarity learning of graphs, which is able to capture high-order structure of graphs and incorporate it into the similarity learning process. Figure 1 shows the framework we propose. The detailed illustration of the proposed framework is as follows.

**Problem Definition.** Given a pair of graphs $G_i = (V_i, E_i, A_i)$ and $G_j = (V_j, E_j, A_j)$, where $V_i$ and $V_j$ contain the same number of vertices with fully aligned physical meanings, the goal of similarity learning on the pair of graphs is to learn a similarity score $s$ between $G_i$ and $G_j$.

**Random Walk Sampling.** Random walk, as a way for sampling, has been used for sampling vertexes or edges for graphs, and the sampling on graphs tend to capture community structure information (Perozzi, Al-Rfou, and Skiena 2014; Ahmed et al. 2018). In this paper, we employ a random walk sampling process on graphs and aim to refine the graph representations with the high-order structural information obtained by the random walks, which will be further incorporated into the similarity learning process in the framework. We denote a random walk rooted at vertex $v_i$ as $W_{v_i}$, which is a stochastic process with random variables $W^{1}_{v_i}, W^{2}_{v_i}, \ldots, W^{K}_{v_i}$, and $W^{j+1}_{v_i}$ is a vertex chosen randomly from the neighbors of vertex $W^{j}_{v_i}$. 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 path length is reached. There could be multiple walks starting from each vertex, depending on the number of walks specified (Perozzi, Al-Rfou, and Skiena 2014). Line 2-14 in Algorithm 1 illustrates the random walk sampling process we employ for capturing high-order structure information. Note that we slide a window with size $w$ on each walk generated and record the frequency of nodes that coccur within a window in $F$, and $w$ decides the number of hops considered for refining the graph representation.
We also initialize the neural network parameters. In the walk sampling stage, we obtain a k-walk to get the cooccurrence frequency between two nodes.

To evaluate the performance of the proposed model for similarity learning of brain networks, we test our framework on four real resting-state fMRI brain datasets and compare with state-of-the-art baselines.

Then we compute the loss for the Siamese network. We use the Hinge loss in Equation (5). To optimize the model, we apply the stochastic gradient descent with ADAM optimizer to update $\Theta$.

### Algorithm 1 Multi-hop Siamese GCN

**Input:** $D = G_1, G_2, \ldots, G_n$ (training graph samples); $y$ (class labels); random walk parameters: $\gamma$ (number of walks), $l$ (walk length), $w$ (window size).

1. Compute $\bar{A}$ for $\bar{G}(V, E)$;
2. Initialize a frequency matrix $F \in \mathbb{R}^{m \times m}$ with 0s;
3. for $i = 0$ to $\gamma$ do
4. $V' = \text{Shuffle}(V)$;
5. for each $v_i \in V'$ do
6. $W_{v_i} = \text{RandomWalk}(\bar{G}, v_i, l)$;
7. Update $F$;
8. end for
9. end for
10. Obtain a $k$-nn graph based on $F$;
11. Add the edges of the $k$-nn graph into $\bar{G}$;
12. Obtain the refined adjacency matrix $\bar{A}$;
13. Prepare pairs of training samples from $D$;
14. Initialize the parameters $\Theta$ of GCNs in Siamese network;
15. while not converge do
16. Perform spectral filterings according to Equation (4);
17. Compute the similarity estimate $s_i$ for the $i^{th}$ pair;
18. Compute the loss $L_{\text{hinge}}$ according to Equation (5);
19. Apply stochastic gradient descent with ADAM optimizer to update $\Theta$;
20. end while

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) (Van Essen et al. 2013). The preprocessing of the fMRI data consists of intensity normalization, phase-encoding direction unwarping, motion correction, spatial normalization to standard template and artifact removing (Spronk et al. 2018). 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 based on the label from cognitive traits.

- **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. The brain networks were constructed with the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon 2012). The raw images were realigned and co-registered, followed by the normalization and smoothing steps. Then the confound

---

### Experiments & Results

To evaluate the performance of the proposed model for similarity learning of brain networks, we test our framework on four real resting-state fMRI brain datasets and compare with state-of-the-art baselines.

### Datasets and Preprocessing

- **Autism Brain imaging Data Exchange (ABIDE):** This dataset is provided by the ABIDE initiative (Di Martino et al. 2014). The fMRI data includes subject demographic information, disease status, and other clinical information. 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) (Van Essen et al. 2013). The preprocessing of the fMRI data consists of intensity normalization, phase-encoding direction unwarping, motion correction, spatial normalization to standard template and artifact removing (Spronk et al. 2018). 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 based on the label from cognitive traits.

- **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. The brain networks were constructed with the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon 2012). 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 (Ragin et al. 2012). 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 toolbox³ 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 (AVOIs), corresponding to 116 brain regions. Then we 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.

### Baselines and Metrics

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

- **PCA** is the Principal Component Analysis approach that is widely used for dimension reduction and feature extraction (Smith 2002). We apply PCA on the correlation matrices of the brain networks and perform similarity learning based on the PCA results.

- **Siamese GCN (S-GCN)** is the method proposed in (Ktena et al. 2018), which learns similarity scores between graphs based on the outputs of graph convolutional neural networks in a Siamese framework. This was the first work of applying graph convolutional neural network on brain connectivity networks for similarity learning.

- **Multi-hop Siamese GCN (S-MGCN)** is the proposed approach in this paper, which is a general framework that provides options for incorporating high-order information of graphs into the similarity learning process.

In the evaluations of the Siamese GCN and Multi-hop Siamese GCN, we use 60% of the data for training and the other 40% for testing. We choose the optimal parameter settings for the Siamese GCN model following the instructions provided in (Ktena et al. 2018). For the PCA model, as it is an unsupervised method, we directly apply it on the testing portion of the data. The only parameter in the PCA model is the number of components to be preserved in the output lower dimensional representation. We use the number of atlas regions (i.e., 110) as the number for the ABIDE dataset and HCP dataset, as the brain networks in two datasets involves over 200 brain regions and we hope to keep the brain network aligned with each other while doing PCA by latently mapping them to the 110 atlas regions. For the HIV and HCP datasets, we use 55 for the number of components, which is equal to the number of the ICA non-artefactual components introduced in (Ktena et al. 2018). After we obtain the output representations from PCA, we calculate the similarity score for each pair according to Equation (6) (Frey and Dueck 2007).

\[
S_{ij} = 1 - \sqrt{\text{Tr} \left( (F_i - F_j)^T (F_i - F_j) \right)}
\]

where \(F_i\) and \(F_j\) are the PCA results of subject \(i\) and \(j\), respectively. For each experiment, we run for 5 times and report the average results.

### Evaluation

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.

#### 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 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 its similarity with the labeled samples. With the Siamese architecture, the proposed model inherently has 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 Multi-hop Siamese GCN model, which would greatly guarantee the training effectiveness and robustness of the model when applying it for predicting unseen samples. Table 1 and Figure 3 shows the classification AUC of the Multi-hop Siamese GCN and that of the baseline methods on pair classification on the four datasets.

As shown in Table 1 we observe the classification AUC of Multi-hop Siamese GCN is significantly higher than that

| Methods              | ABIDE  | HCP    | HIV    | Bipolar |
|----------------------|--------|--------|--------|---------|
| PCA                  | 0.50 ± 0.01 | 0.48 ± 0.004 | 0.54 ± 0.07 | 0.49 ± 0.01 |
| Siamese GCN          | 0.78 ± 0.20  | 0.81 ± 0.36  | 0.46 ± 0.35  | 0.60 ± 0.32  |
| Multi-hop Siamese GCN| 0.96 ± 0.02 | 0.98 ± 0.03 | 0.77 ± 0.20 | 0.94 ± 0.07 |

Figure 3: AUC Scores of Pair Classification.

³http://rfmri.org/DPARSF.
of the baseline methods. More specifically, our proposed approach achieves an average AUC gain of 82.6% compared to PCA, and an average AUC gain of 42% compared to S-GCN across all datasets. Thus, our proposed method is more accurate and has a lower variance compared to S-GCN. Among the three methods, 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. However, in brain functional connectivity networks, the inner structure usually reflects the collaborative patterns or interactions between different brain regions, which could serve as important features for discriminating brain health status. The Siamese GCN model instead employs the neighborhood structural information of graphs and performs graph convolutions with spectral filtering, which tends to capture localized structural information of graphs. It achieved fairly good results on ABIDE and HCP datasets although not as good as our proposed framework. The superior performance of our proposed Multi-hop Siamese GCN approach indicates that the multi-hop graph convolutions enabled by the high-order structural representation introduced by the random walk stage did help the similarity learning of brain networks. Moreover, the Multi-hop Siamese GCN achieved the best results on all the four datasets, demonstrating its generalizing ability in similarity learning of brain networks.

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 Multi-hop Siamese GCN model and the baseline Siamese GCN model on the ABIDE and Bipolar datasets. We apply the weighted k-nearest neighbour (kNN) classifier (Hechenbichler and Schliep 2004) 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 class labels of its k-nearest neighbors. Here we consider all the neighbors with positive similarity scores in the weighted calculation. Meanwhile, we hope to explore the influence of different loss functions in the subject classification performance. Besides evaluating the two models with the Hinge loss in Equation (5), we also evaluate them with the following constrained variance loss:

\[
  L_{\text{convar}} = \max(0, \delta^2 + a) + \max(0, \delta^2 - a) \\
  + \max(0, m - (\sigma^+ - \sigma^-)),
\]

where \(\mu^+\) represents the mean similarity between embeddings belonging to the same class, and \(\mu^-\) represents the mean similarity between embeddings belonging to different classes, while \(\sigma^+\) and \(\sigma^-\) 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 Kienzle et al. (2018). By this formulation, the variance is only penalised 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. Figure 4 shows the evaluation results of subject classification by the two models with different loss functions. As shown in the Figure, the proposed Multi-hop Siamese GCN model achieves a higher accuracy with both loss functions on both datasets compared to the baseline Siamese GCN model. This implies that the similarity scores learned by the proposed model are more accurate, thus more reliable to be used for further multi-subject brain connectivity analysis. By comparing Figure 4(a) and Figure 4(b) we can find that both models get higher classification accuracy with the similarity scores learned by the constrained variance loss. This could be the benefit from allowing for more diversity across the samples by the constrained variance loss.

Parameter Analysis

In the proposed Multi-hop Siamese GCN model, there are two sets of parameters. One is the set of parameters for the convolutional networks, and the other set are the parameters for the random walk algorithm. In the experiment, we use 2 GCN layers with \(f = 32\) features for each. We use the stochastic gradient descent with ADAM algorithm (Kingma and Ba 2014) 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 (7) 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 \(\gamma\) 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 for refining graph representation, we use 10\% of the number of nodes in the brain networks as the value for \(k\).

To analyze the influence of the parameters in the random walk process on the similarity learning performance of the proposed model, we perform a parameter sensitivity evaluation for the two key parameters in random walk: the walk/path length \(l\) and the window size \(w\) which is also the number of hops considered in the graph convolutions. The bar plots in Figure 5 show the AUC scores of the proposed model with different parameter values for path length and number of hops. From the figure, we observe that the AUC scores
Wang et al. 2017; Ma et al. 2017b). For example, in (Bai et al. 2017), which learns a unified network embedding from functional voxel-level data or mining from brain networks for neurology and number of hops are also the largest among the four value and number of hops is 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 the larger graph, we should perform a relatively large number of hops as well. Therefore, it is important to consider the scale and other relevant properties of the brain networks when selecting the parameter values.

Related Work

Our work relates to several branches of studies, which include brain network analysis, Siamese networks and graph convolutional networks. Brain network analysis has been an emerging research area, as it yields new insights concerning the understanding of brain function and many neurological disorders (Liu et al. 2017). Existing works in brain networks mainly focus on discovering brain network from spatio-temporal voxel-level data or mining from brain networks for neurological analysis (Ma et al. 2016; Bai et al. 2017; Ma et al. 2017a; Wang et al. 2017; Ma et al. 2017b). For example, in (Bai et al. 2017), an unsupervised matrix tri-factorization method is developed to simultaneously discover nodes and edges of the underlying brain networks in fMRI data. (Ma et al. 2017b) propose a multi-view graph embedding approach which learns a unified network embedding from functional and structural brain networks as well as hubs for brain disorder analysis. In (Wang et al. 2017), a deep model with CNN is proposed to learn non-linear and modular-preserving 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. A general framework for similarity learning on networks is highly demanded for group-contrasting brain network analysis.

Siamese networks were first introduced in (Bromley et al. [1994]) to solve image matching problem for signature verification. In recent years, Siamese architecture has drawn much attention from researchers in image recognition (Chopra, Hadsell, and LeCun 2005; Koch, Zemel, and Salakhutdinov 2015). In (Chopra, Hadsell, and LeCun 2005), a Siamese framework with CNN is applied to learn similarity metric for face verification, while (Koch, Zemel, and Salakhutdinov 2015) proposes an approach with Siamese networks to generalize the predictive power of the model to unseen classes and solve one-shot classification problems. Although the Siamese architecture has been widely used in image classification, it is seldom explored in the domain of graph mining. We aim to take advantage of this architecture to learn similarity metric of graphs with Multi-hop GCN, which incorporates high-order graph representation into the learning process.

Graph convolutional network (GCN), first proposed in (Defferrard, Bresson, and Vandergheynst 2016), is a general-ized architecture of convolutional neural network on graphs, where localized spectral filters are defined by Chebyshev polynomials and a recursive formulation is employed for fast filtering operations. In (Kipf and Welling 2016), a renormalization trick is introduced to simplify and speedup computations when applying GCNs for semi-supervised classification. (Li, Han, and Wu 2018) proposes co-training and self-training approaches to improve the training of GCNs in learning with very few labels. In this work, we apply GCNs in Siamese architecture to obtain important structural features in graphs for similarity learning. As the spectral filters in GCNs are closely related to the graph Laplacian, it is very important to refine the graph representation used in the framework. That motivates us to design the multi-hop GCNs that can capture refined high-order structural information of graphs for a better learning performance. In future work, we aim to explore the impact of attention networks on both pair and subject classification (Lee et al. 2018; Veličković et al. 2017; Lee, Rossi, and Kong 2018).

Conclusion

We present a general framework called Multi-hop Siamese GCN for learning similarity between brain networks. We employ random walks with sliding windows to obtain high-order proximity in graphs and leverage the high-order information in graph convolutional networks for the similarity learning. The multi-hop property allows for multiple options of refining the graph representation used in the framework. That motivates us to design the multi-hop GCNs that can capture refined high-order structural information of graphs for a better learning performance. In future work, we aim to refine the graph representation used in the framework.
References

[Abraham et al. 2017] Abraham, A.; Milham, M. P.; Di Martino, A.; Craddock, R. C.; Samaras, D.; Thirion, B.; and Varoquaux, G. 2017. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage* 147:736–745.

[Ahmed et al. 2018] Ahmed, N. K.; Rossi, R.; Lee, J. B.; Kong, X.; Willke, T. L.; Zhou, R.; and Eldardiry, H. 2018. Learning role-based graph embeddings. *arXiv preprint arXiv:1802.02896*.

[Bai et al. 2017] Bai, Z.; Walker, P.; Tschiffely, A.; Wang, F.; and Davidson, I. 2017. Unsupervised network discovery for brain imaging data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 55–64. ACM.

[Bautista, Sanakoyeu, and Ommer 2017] Bautista, M. Á.; Sanakoyeu, A.; and Ommer, B. 2017. Deep unsupervised similarity learning using partially ordered sets. In *CVPR*, 1923–1932.

[Benson, Gleich, and Leskovec 2016] Benson, A. R.; Gleich, D. F.; and Leskovec, J. 2016. Higher-order organization of complex networks. *Science* 353(6295):163–166.

[Bromley et al. 1994] Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; and Shah, R. 1994. Signature verification using *a* siamese *time* delay neural network. In *Advances in Neural Information Processing Systems*, 737–744.

[Bruna et al. 2013] Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.2603*.

[Bullmore and Sporns 2009] Bullmore, E., and Sporns, O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3):186.

[Chopra, Hadsell, and LeCun 2005] Chopra, S.; Hadsell, R.; and LeCun, Y. 2005. Learning a similarity metric discriminatively, with application to face verification. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, 539–546. IEEE.

[Defferrard, Bresson, and Vandergheynst 2016] Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, 3844–3852.

[Di Martino et al. 2014] Di Martino, A.; Yan, C.-G.; Li, Q.; Denio, E.; Castellanos, F. X.; Alaerts, K.; Anderson, J. S.; Assaf, M.; Bookheimer, S. Y.; Dapretto, M.; 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):659.

[Frey and Dueck 2007] Frey, B. J., and Dueck, D. 2007. Clustering by passing messages between data points. *Science* 315(5814):972–976.

[Hechenbichler and Schliep 2004] Hechenbichler, K., and Schliep, K. 2004. Weighted k-nearest-neighbor techniques and ordinal classification.

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

[Kipf and Welling 2016] Kipf, T. N., and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

[Koch, Zemel, and Salakhutdinov 2015] Koch, G.; Zemel, R.; and Salakhutdinov, R. 2015. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2.

[Ktena et al. 2018] Ktena, S. I.; Parisot, S.; Ferrante, E.; Rajchl, M.; Lee, M.; Glocker, B.; and Rueckert, D. 2018. Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage* 169:431–442.

[Lee et al. 2018] Lee, J. B.; Rossi, R. A.; Kim, S.; Ahmed, N. K.; and Koh, E. 2018. Attention models in graphs: A survey. *arXiv preprint arXiv:1807.07984*.

[Lee, Rossi, and Kong 2018] Lee, J. B.; Rossi, R.; and Kong, X. 2018. Graph classification using structural attention. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1666–1674. ACM.

[Li, Han, and Wu 2018] Li, Q.; Han, Z.; and Wu, X.-M. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. *arXiv preprint arXiv:1801.07606*.

[Liu et al. 2017] Liu, J.; Li, M.; Pan, Y.; Lan, W.; Zheng, R.; Wu, F.-X.; and Wang, J. 2017. Complex brain network analysis and its applications to brain disorders: a survey. *Complexity* 2017.

[Livi and Rizzi 2013] Livi, L., and Rizzi, A. 2013. The graph matching problem. *Pattern Analysis and Applications* 16(3):253–283.

[Ma et al. 2016] Ma, G.; He, L.; Cao, B.; Zhang, J.; Philip, S. Y.; and Ragan, A. B. 2016. 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*, 476–492. Springer.

[Ma et al. 2017a] Ma, G.; He, L.; Lu, C.-T.; Shao, W.; Yu, P. S.; Leow, A. D.; and Ragan, A. B. 2017a. Multi-view clustering with graph embedding for connectome analysis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 127–136. ACM.

[Ma et al. 2017b] Ma, G.; Lu, C.-T.; He, L.; Philip, S. Y.; and Ragan, A. B. 2017b. Multi-view graph embedding with hub detection for brain network analysis. In *Data Mining (ICDM), 2017 IEEE International Conference on*, 967–972. IEEE.

[Perozzi, Al-Rfou, and Skiena 2014] Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710. ACM.

[Ragan et al. 2012] Ragan, A. B.; Du, H.; Ochs, R.; Wu, Y.; Sammet, C. L.; Shoukry, A.; and Epstein, L. G. 2012. Structural brain alterations can be detected early in hiv infection. *Neurology* 79(24):2328–2334.
[Rossi, Ahmed, and Koh 2018] Rossi, R. A.; Ahmed, N. K.; and Koh, E. 2018. Higher-order network representation learning. In Companion of the The Web Conference 2018 on The Web Conference 2018, 3–4. International World Wide Web Conferences Steering Committee.

[Rossi, Zhou, and Ahmed 2017] Rossi, R. A.; Zhou, R.; and Ahmed, N. K. 2017. Deep feature learning for graphs. arXiv preprint arXiv:1704.08829

[Shuman et al. 2013] Shuman, D. I.; Narang, S. K.; Frossard, P.; Ortega, A.; and Vandergheynst, P. 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine 30(3):83–98.

[Smith 2002] Smith, L. I. 2002. A tutorial on principal components analysis. Technical report.

[Spronk et al. 2018] Spronk, M.; Ji, J. L.; Kulkarni, K.; Repovs, G.; Anticevic, A.; and Cole, M. W. 2018. Mapping the human brain’s cortical-subcortical functional network organization. bioRxiv 206292.

[Takerkart et al. 2014] Takerkart, S.; Auzias, G.; Thirion, B.; and Ralaivola, L. 2014. Graph-based inter-subject pattern analysis of fmri data. PloS one 9(8):e104586.

[Van Essen et al. 2013] Van Essen, D. C.; Smith, S. M.; Barch, D. M.; Behrens, T. E.; Yacoub, E.; Ugurbil, K.; Consortium, W.-M. H.; et al. 2013. The wu-minn human connectome project: an overview. Neuroimage 80:62–79.

[Veličković et al. 2017] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2017. Graph attention networks. arXiv preprint arXiv:1710.10903

[Wang et al. 2014] Wang, J.; Song, Y.; Leung, T.; Rosenberg, C.; Wang, J.; Philbin, J.; Chen, B.; and Wu, Y. 2014. Learning fine-grained image similarity with deep ranking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1386–1393.

[Wang et al. 2017] Wang, S.; He, L.; Cao, B.; Lu, C.-T.; Yu, P. S.; and Ragin, A. B. 2017. Structural deep brain network mining. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 475–484. ACM.

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