Contrastive Brain Network Learning via Hierarchical Signed Graph Pooling Model

Haoteng Tang, *Student Member, IEEE*, Guixiang Ma, *Member, IEEE*, Lei Guo, *Student Member, IEEE*, Xiyao Fu, *Student Member, IEEE*, Heng Huang, *Member, IEEE*, Liang Zhan†, *Member, IEEE*

***Abstract*—Recently brain networks have been widely adopted to study brain dynamics, brain development and brain diseases. Graph representation learning techniques on brain functional networks can facilitate the discovery of novel biomarkers for clinical phenotypes and neurodegenerative diseases. However, current graph learning techniques have several issues on brain network mining. Firstly, most current graph learning models are designed for unsigned graph, which hinders the analysis of many signed network data (e.g., brain functional networks). Meanwhile, the insufficiency of brain network data limits the model performance on clinical phenotypes predictions. Moreover, few of current graph learning model is interpretable, which may not be capable to provide biological insights for model outcomes. Here, we propose an interpretable hierarchical signed graph representation learning model to extract graph-level representations from brain functional networks, which can be used for different prediction tasks. In order to further improve the model performance, we also propose a new strategy to augment functional brain network data for contrastive learning. We evaluate this framework on different classification and regression tasks using the data from HCP and OASIS. Our results from extensive experiments demonstrate the superiority of the proposed model compared to several state-of-the-art techniques. Additionally, we use graph saliency maps, derived from these prediction tasks, to demonstrate detection and interpretation of phenotypic biomarkers.**

***Index Terms*—Signed Graph Learning, Hierarchical Graph Pooling, Contrastive Learning, Brain Functional Networks, Data Augmentation, Interpretability.**

I. Introduction

UNDERSTANDING brain organizations and their relationship to phenotypes (e.g., clinical outcomes, behavioral or demographical variables, etc.) are of prime importance in the modern neuroscience field. One of important research directions is to use non-invasive neuroimaging data (e.g., functional magnetic resonance imaging or fMRI) to identify potential imaging biomarkers for clinical purposes. Most previous studies focus on voxel-wise and region-of-interests (ROIs) imaging features [1]–[3]. However, evidences show that the brain is a complex system whose function relies on a diverse set of interactions among brain regions. These brain functions will further determine human clinical or behavioral phenotypes [4]–[13]. Therefore, more and more studies have been conducted to predict those phenotypes using the brain network as the delegate of interactions among brain regions [14]–[16]. Additionally, compared to traditional neuroimaging features, brain network has more potential to gain interpretable and system-level insights into phenotype-induced brain dynamics [17]. A brain network is a 3D brain graph model, where graph nodes represent the attributes of brain regions and graph edges represent the connections (or interactions) among these regions.

Many studies have been conducted to analyze brain networks based on the graph theory, however, most of these studies focus on pre-defined network features, such as clustering coefficient, small-worldness [18]–[22]. This may be sub-optimal since these pre-defined network features may not be able to capture the characteristics of the whole brain network. However, the whole brain network is difficult to be analyzed due to the high dimensionality. To tackle this issue, Graph Neural Network (GNN), as one of embedding techniques, has gained increasing attentions to explore biological characteristics of brain network-phenotype associations in recent years [23]–[25]. GNN is a class of deep neural networks that can embed the high-dimensional graph topological structures with graph node features into low dimensional latent space based on the information passing mechanism [26]–[28]. A few studies proposed different GNNs to embed the nodes in brain networks and applied a global readout operation (e.g., global mean or sum) to summarize all latent node features as the whole brain network representation for downstream tasks (e.g., behavioral score regression, clinical disease classification) [4], [24], [25], [29]. However, the message passing of GNNs is inherently ‘flat’ which only propagates information across graph edges and is unable to capture hierarchical structures rooted in graphs which are crucial in brain functional organizations [30]–[33]. To address this issue, many recent studies introduce hierarchical GNNs, including node embedding and hierarchical graph pooling strategies, to embed the whole brain network in a hierarchical manner [30], [34]–[37].

Although GNNs have achieved great progresses on brain network mining, several issues should be addressed. First, most existing GNNs are designed for unsigned graphs in which all graph nodes are connected via non-negative edges (i.e., edge weights are in the range of $[0, \infty)$). However, signed graphs are very common in brain research (e.g., functional MRI-derived brain networks or brain functional networks), which leads to a demand of signed graph embedding models. To tackle this issue, a few recent studies proposed signed graph embedding models based on the balance-theory [38]–

H. Tang, L. Guo, X. Fu, H. Huang and L. Zhan are with the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, 15260, USA. (e-mail: {haoteng.tang, Lei.guo, xy_fu, heng.huang, liang.zhan}@pitt.edu)

G. Ma is with the Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA. (e-mail:guixiang.ma@intel.com). This work was done before Dr. Ma joined Intel.

Liang Zhan is the corresponding author (denoted by †).

[41]. The balance-theory, motivated by human attitudes in social networks, is used to describe the node relationship in signed graphs, where nodes connected by positive edges are considered as ‘friends’, otherwise are considered as ‘opponents’. In the realm of brain functional networks, the positive edge means co-activation and the negative edge indicates anti-activation between those connected nodes. Meanwhile, the balance-theory defines 4 higher-order relationships among graph nodes: (1) the ‘friend’ of ‘friend’ is ‘friend’, (2) the ‘opponent’ of ‘friend’ is ‘opponent’, (3) the ‘friend’ of ‘opponent’ is ‘opponent’, and (4) the ‘opponent’ of ‘opponent’ is ‘friend’. These definitions are accorded with nodal relationships in the functional brain network, which indicates that the balance theory is applicable in brain functional network embedding. In this study, we adopt the balance theory to co-embed the positive and negative edges as well as local brain nodes. Therefore, generated latent node features include balanced and unbalanced feature components. Beyond focusing on local structures, we also consider the hierarchical structure in graphs as one of global graph features. As suggested by literature [30], [42]–[44], graph hierarchical structure can facilitates to yield whole graph representations and to enable the graph-level tasks (i.e., clinical disease classification based on whole brain networks). Particularly, we proposal a new hierarchical pooling module for signed graphs based on the information theory and extend current methods on signed graph from the local embedding to the global embedding.

The second issue is that most of current GNNs on brain network studies are not interpretable, and thus are incapable to provide biological explanations or heuristic insights for model outcomes. This is mainly due to the black-box nature of neural networks. To address this issue, we propose a signed graph learning model with an interpretable graph pooling module. Previous studies indicated that brain networks are hierarchically organized by some regions as neuro-information hubs and peripheral regions, respectively [45]–[48]. In our graph pooling module, we compute an information score to measure the information gain for each brain node and choose top-K nodes with high information gains as information hubs. And the information of other peripheral brain nodes will be aggregated onto these hubs. Hence, the proposed pooling module can be interpreted as a brain information hub generator. Apparently, the outcome of this pooling module is a subgraph of the original brain network without creating any new nodes. Therefore, yielded subgraph nodes can be regarded as potential biomarkers to provide heuristic biological explanations for tasks.

To further boost the proposed model performance on prediction tasks, we introduce graph contrastive learning into our proposed hierarchical signed graph representation learning (HSGRL) model. A data augmentation strategy to generate contrastive brain functional network samples is necessary to achieve graph contrastive learning. The data augmentation for contrastive learning aims at creating reasonable data samples, by applying certain transformations, which are similar to original data samples. For example, image rotation and cropping are common transformations to generate new samples in image classification tasks [49]–[53]. In graph structural data, a few studies proposed to utilize graph perturbations (i.e., add/drop graph nodes, manipulate graph edges) and graph view augmentation (e.g., graph diffusion) to generate contrastive graph samples from different views [54]–[58]. These strategies, although boosting the model performance on large-scale benchmark datasets (e.g., CORA, CITESEER, etc.), may not be suitable to generate contrastive brain network samples. On the one hand, each node in brain networks represents a defined brain region with specific brain activity information so that the brain node can not be arbitrarily removed or added. On the other hand, add/drop operations on brain network may lead to unexpected model outcomes which are difficult to explain and understand from biological views. Motivated by [59], [60], we generate contrastive brain functional network samples directly from fMRI BOLD signals, where the generated contrastive samples are similar to the original ones, and the internal biological structure is therefore maintained. Our main contributions are summarized as follow:

- We propose a hierarchical signed graph representation learning (HSGRL) model to embed brain functional networks and we apply the proposed model on multiple phenotype prediction tasks.
- We propose a contrastive learning architecture with our proposed HSGRL model to boost the model performance on several prediction tasks. A graph augmentation strategy is proposed to generate contrastive samples for fMRI-derived brain network data.
- The proposed HSGPL model is interpretable which yields heuristic biological explanations.
- Extensive experiments are conducted to demonstrate the superiority of our method. Moreover, we draw graph saliency maps for clinical tasks, to enable interpretable identifications of phenotype biomarkers.

II. Related Works

A. Graph Neural Networks and Brain Network Embedding

GNNs are generalized deep learning architectures which are broadly utilized for graph representation learning in many fields (e.g., social network mining [61], [62], molecule studies [63], [64] and brain network analysis [65]). Most existing GNN models (e.g., GCN [26], GAT [27], GraphSage [66]) focus on node-level representation learning and only propagate information across edges of the graph in a flat way. When deploying these models on graph-level tasks (e.g., graph classification, graph similarity learning, [42]–[44], [67]), the whole graph representations are obtained by a naive global readout operation (e.g., sum or average all node feature vectors). However, this may lead to poor performance and low efficiency in graph-level tasks since the hierarchical structure, an important property that existed in graphs, is ignored in these models. To explore and capture hierarchical structures in graphs, a few hierarchical graph pooling strategies are proposed to learn representations for the whole graph in a hierarchical manner [30], [34], [35], [68], [69]. Traditional methods to extract brain network patterns are based on graph theory [18]–[22] or geometric network optimization [70]–[73]. A few recent studies [24], [25], [74] introduce GNNs to discover brain patterns for phenotypes predictions. However, hierarchical structures in brain networks are not considered in these models, which limits the model performance in a way. Recently, a few hierarchical brain network embedding models are proposed [36], [75].

However, all the aforementioned GNNs are designed for unsigned graph representation learning. A few recent studies are proposed to handle the signed graphs, however, they only consider the node-level representation learning [39], [41], [76], [77]. In this work, we design a signed graph hierarchical pooling strategy to extract graph-level representations from brain functional networks.

B. Interpretable Graph Learning Model

Generally, the mechanism about how GNNs embed the graph nodes can be explained as a message passing process, which includes message aggregations from neighbor nodes and message (non-linear) transformations [28], [36], [78]. However, most current hierarchical pooling strategies are not interpretable [30], [34], [35]. A few recent studies try to propose interpretable graph pooling strategies to make the pooling module intelligible to the model users. Most of these pooling strategies down-sample graphs relying on network communities which are one of the important hierarchical structures that can be interpreted [36], [37], [79]. For example, [36] proposed a hierarchical graph pooling neural network relying on brain network community to yield interpretable biomarkers. The hierarchical pooling strategy proposed in this work relies on the network information hub which is another important hierarchical structure in brain networks.

C. Data Augmentation for Graph Contrastive Learning

Most current graph contrastive learning methods augment graph contrastive samples by manipulating graph topological structures. For example, [55], [56] generate the contrastive graph samples by dropping nodes and perturbing edges. Other studies generate contrastive samples by changing the graph local receptive field, which is named as the graph view augmentation [54], [80]. In this work, we introduce the graph contrastive learning into brain functional network analysis and generate contrastive samples from the fMRI BOLD signals.

III. Preliminaries of Brain Functional Networks

We denote a brain functional network with N nodes as $G = \{V, E\} = (A, H)$. V is the graph node set where each node (i.e., $v_i, i = 1, ..., N$) represents a brain region. E is the graph edge set where each edge (i.e., $e_{i,j}$) describes the connection between node v_i and v_j. $A \in \mathbb{R}^{N \times N}$ is the graph adjacency matrix where each element, $a_{i,j} \in A$, is the weight of edge $e_{i,j}$. $H \in \mathbb{R}^{N \times C}$ is the node feature matrix where $H_i \in H$ is the $i - th$ row of H representing the feature vector of v_i. Let $B \in \mathbb{R}^{N \times D}$ be the fMRI BOLD signal matrix, where D is the signal length. Generally, the edge weight in the brain functional network can be computed from the fMRI BOLD signal by $a_{i,j} = corr(b_i, b_j)$, where b_i is the $i - th$ row of B representing the BOLD signal of v_i and $corr(\cdot)$ is the correlation coefficient operator. Note that $a_{i,j}$ can be either positive or negative value so that brain functional network is a signed graph. For each subject, we use $\hat{}$ and $\check{}$ to denote a functional brain network contrastive sample pair (i.e., $\hat{G} = (\hat{A}, \hat{H})$ and $\check{G} = (\check{A}, \check{H})$).

IV. Methodology

In this section, we first propose a data augmentation strategy to generate contrastive samples for brain functional networks. Secondly, we introduce our proposed hierarchical signed graph representation learning (HSGRL) model with node embedding and hierarchical graph pooling modules. Finally, we deploy the contrastive learning framework on our proposed HSGRL model to yield the representations for the whole graph, which can be applied to downstream prediction tasks.

A. Contrastive Samples of Brain Functional Networks

The generation of contrastive samples aims at creating reasonable and similar functional brain network pairs by applying certain transformations. Here we propose a new strategy to generate the brain functional network contrastive samples from fMRI BOLD signals. For each node v_i, we generate two sub-BOLD-signals ($\hat{b}_i$ and $\check{b}_i$) by manipulating its original bold signal b_i. Specifically, we use a window ($size = d$) to clamp the b_i from the signal head and tail, respectively:

$$\begin{aligned} \hat{b}_i &= b_i[d+1, \quad d+2, \quad ..., \quad D] \\ \check{b}_i &= b_i[1, \quad 2, \quad ..., \quad D-d] \end{aligned} \tag{1}$$

Obviously, $b_i \in \mathbb{R}^{1 \times D}$, $\hat{b}_i$ and $\check{b}_i \in \mathbb{R}^{1 \times (D-d)}$. To keep the similarity between $\hat{G}$ and $\check{G}$, we set the window size $d \ll D$. After we generate a pair of sub-bold-signals, we can compute edge weights of the pairwise contrastive brain functional network samples by:

$$\begin{aligned} \hat{a}_{i,j} &= corr(\hat{b}_i, \hat{b}_j) \\ \check{a}_{i,j} &= corr(\check{b}_i, \check{b}_j), \end{aligned} \tag{2}$$

where $\hat{a}_{i,j} \in \hat{A}$ and $\check{a}_{i,j} \in \check{A}$ are the weights of $e_{i,j}$ in two contrastive samples. We do not consider the contrastive node features in this work, therefore $\hat{X} = \check{X} = X$. The generated contrastive sample pairs are similar with same node features and slightly different edge weights. We will show this similarity in section V-C.

B. Hierarchical Signed Graph Representation Learning Model

We present our Hierarchical Signed Graph Representation Learning (HSGRL) model in Figure 1. The HSGRL model includes Balanced and Unbalanced Embedding (BUE) module and Hierarchical Graph Pooling (HGP) module.

1) BUE module: The balance theory is broadly used to analyze the node relationships in signed graphs. The theory states that given a node v_i in a signed graph, any other node (i.e., v_j) can be assigned into either balanced node set or unbalanced node set to v_i regarding to a path between v_i and v_j. Specifically, if the number of negative edges are even in the path between v_i and v_j, then v_j belongs to the balanced

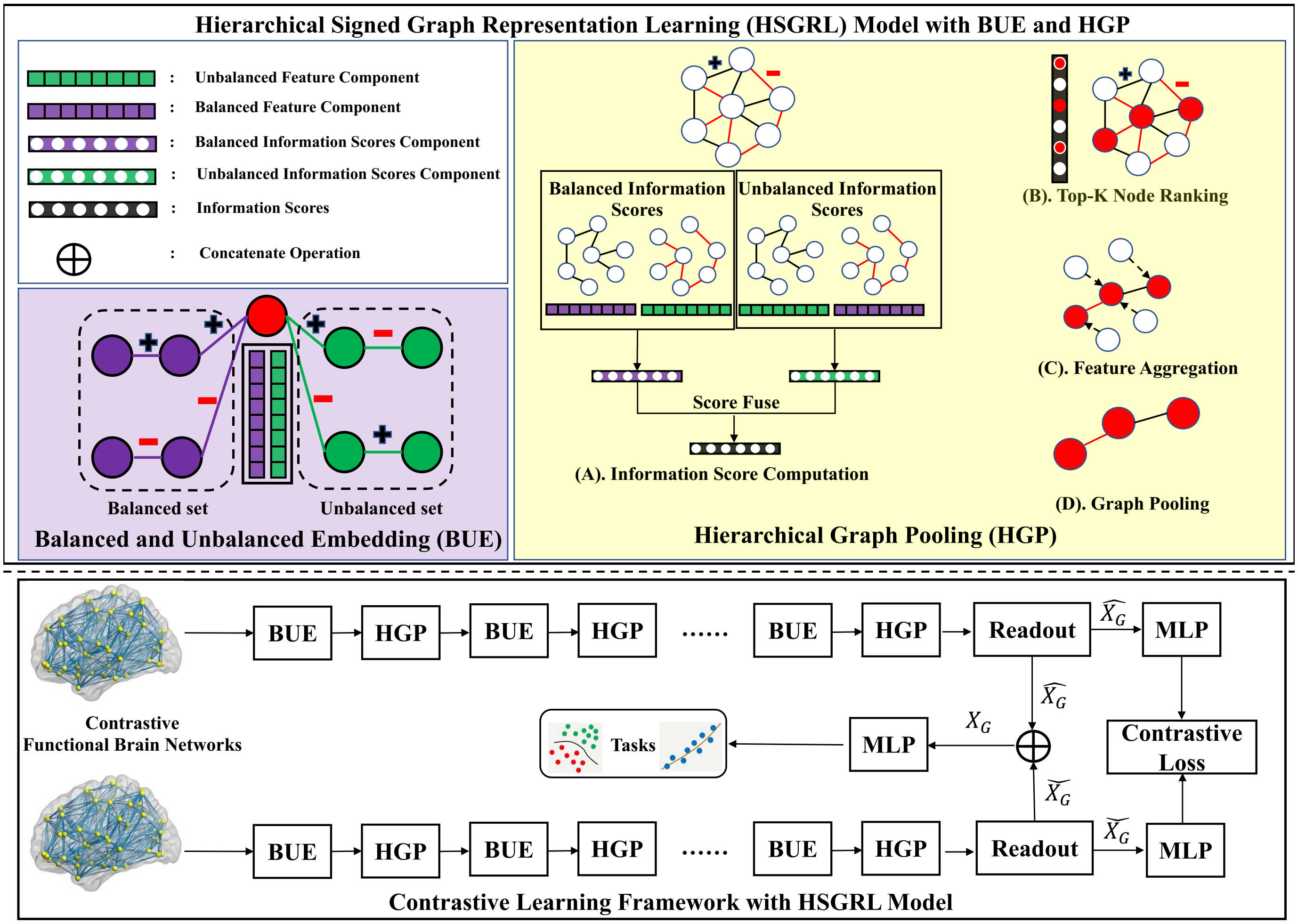


Fig. 1. Diagram of the proposed contrastive graph learning framework (in the bottom black box) with hierarchical signed graph representation learning (HSGRL) model (in the top black box) for functional brain network embedding and downstream tasks (i.e., phenotype classification or regression). The HSGRL model consists of cascaded BUE and HGP modules to extract graph-level representations of contrastive brain functional network pairs (i.e., $\hat{X}_G$ and $\check{X}_G$) in a hierarchical manner. The $\hat{X}_G$ and $\check{X}_G$ participate to build up the contrastive loss for graph constrastive learning. Meanwhile, a concatenate operation is utilized to generate the fused graph feature by $X_G = [\hat{X}_G \| \check{X}_G]$). The fused graph feature X_G is utilized for downstream prediction tasks (i.e., graph classification and regression).

set of v_i. Otherwise, v_j belongs to the unbalanced set of v_i. The balance theory indicates that:

- Each graph node, v_j, can belong to either the balanced or unbalanced node set of a given target node v_i.
- The path between v_i and v_j determines the balance attribute of v_j.

Motivated by this, we adopt the idea of signed graph attention networks from [41] to embed brain functional network nodes to generate latent node features with balanced and unbalanced components:

$$X^B, X^U = F_{sign}(A, H) \tag{3}$$

where $F_{sign}(\cdot)$ is the signed graph attention encoder [41]. X^B and X^U are the node balanced and unbalanced components of node latent features, respectively. We fuse the two feature components as the node latent features by:

$$X = [X^B \| X^U], \tag{4}$$

where $[\|]$ denotes concatenate operation.

2) Hierarchical Signed Graph Pooling: As shown in Figure 1, the proposed Hierarchical Graph Pooling (HGP) module consists of 4 steps including: (A) information scores computation, (B) Top-K informative hubs selection, (C) features aggregation and (D) graph pooling.

Information Score Computation: The information score of each node is also considered to contain balanced and unbalanced components to measure the information quantity that each node gains from balanced node set and unbalanced node set, respectively. We first split the signed graph (i.e., with adjacency matrix as A) into positive sub-graph (with adjacency matrix as A_+) and negative one (with adjacency matrix as A_-). Then we utilize Laplace normalization to normalize these two adjacency matrices as:

$$\begin{aligned} \bar{A}_+ &= D_+^{-\frac{1}{2}} A_+ D_+^{-\frac{1}{2}} \\ \bar{A}_- &= D_-^{-\frac{1}{2}} |A_-| D_-^{-\frac{1}{2}}, \end{aligned} \tag{5}$$

where $\bar{A}$ is the normalized adjacency matrix. D_+ and D_- are degree matrices of A_+ and $|A_-|$, respectively. Note that the i-th line in $\bar{A}$, denoted by $\bar{A}_i$, represents the connectivity probability distribution between v_i and any other nodes. For each node (i.e., v_i), we respectively define the balanced and unbalanced components of information score (IS) by:

$$\begin{aligned} IS_i^B &= \left\|\bar{A}_{+,i:}^\top \otimes X^B\right\|_{\tilde{L}_1} + \left\|\bar{A}_{-,i:}^\top \otimes X^U\right\|_{\tilde{L}_1} \\ IS_i^U &= \left\|\bar{A}_{+,i:}^\top \otimes X^U\right\|_{\tilde{L}_1} + \left\|\bar{A}_{-,i:}^\top \otimes X^B\right\|_{\tilde{L}_1}, \end{aligned} \quad (6)$$

where $\|\cdot\|_{\tilde{L}_1}$ is line-wise L_1 norm, and $\otimes$ is the scalar-multiplication between each line of two matrices. $\top$ represents transpose of vector. Then the IS of v_i can be obtained by:

$$IS_i = IS_i^B + IS_i^U. \quad (7)$$

Top-K Node Selection and Feature Aggregation: After we obtain the information score for each brain node, we rank the IS and select K brain nodes, with top-K IS values, as informative network hubs. For the other nodes, we aggregate their features on the selected K network hubs based on the feature attention. Particularly, the feature attention between v_i and v_j is computed by: $x_i x_j^\top$. We weighted add (i.e., set feature attentions as weights) the feature of each unselected node to one of hub features, where the attention value between these two nodes is the biggest.

Graph Pooling After the feature aggregation, we down-scale the graph node by removing all unselected nodes. In another word, only the selected top-K network hubs as well as the edges among them will be preserved after graph pooling. Since the functional brain network is a fully connected graph so that no isolated node is existed in the down-scaled graph.

C. Contrastive Learning Framework with BUE and HGP

The contrastive learning framework with HSGRL is presented in Figure 1. Assume that we forward a pair of contrastive graph samples into the proposed HSGRL model, we will obtain two node latent features, $\hat{X}$ and $\check{X}$ after the last pooling module. We first generate the graph-level representations of two functional brain networks based on the latent node features by a readout operator:

$$\hat{X}_G = \sum_{i=1}^{N'} \hat{x}_i, \quad \check{X}_G = \sum_{i=1}^{N'} \check{x}_i, \quad (8)$$

where $\hat{x}_i$ and $\check{x}_i$ are $i-th$ row of $\hat{X}$ and $\check{X}$. $N'(< N)$ is the number of nodes in the down-scaled graph generated by the last pooling module.

1) Contrastive Loss: The normalized temperature-scaled cross entropy loss [81]–[83] is utilized to construct the contrastive loss. In the framework training stage, we randomly sample M pairs from the generated contrastive graph samples as a mini-batch and forward them to the proposed HSGRL model to generate contrastive graph representation pairs (i.e., $\hat{X}_G$ and $\check{X}_G$). We use $m \in \{1, ..., M\}$ to denote the ID of the sample pair. The contrastive loss of the $m-th$ sample pair is fomulated as:

$$\ell_m = -log \frac{exp(\Phi(\hat{X}_G^m, \check{X}_G^m)/\alpha)}{\sum_{t=1, t\neq m}^{M} exp(\Phi(\hat{X}_G^m, \check{X}_G^t)/\alpha)}, \quad (9)$$

where α is the temperature parameter. $\Phi(\cdot)$ denotes a similarity function that:

$$\Phi(\hat{X}_G^m, \check{X}_G^m) = \hat{X}_G^{m\top} \check{X}_G^m / \|\hat{X}_G^m\| \|\check{X}_G^m\|. \quad (10)$$

The batch contrastive loss can be computed by:

$$\mathcal{L}_{contrastive} = \frac{1}{M} \sum_{m=1}^{M} \ell_m \quad (11)$$

2) Downstream Task and Loss Functions: We use an MLP to generate the framework prediction for both classification and regression tasks. Specifically, the prediction can be generate by $Y_{pred} = MLP([\hat{X}_G \| \check{X}_G])$. We use $NLLLoss$ and $L_1 Loss$ as supervised loss functions ($\mathcal{L}_{supervised}$) of classification and regression tasks, respectively. The whole framework can be trained in an end-to-end manner by optimizing:

$$\mathcal{L} = \eta_1 \mathcal{L}_{supervised} + \eta_2 \mathcal{L}_{contrastive}, \quad (12)$$

where η_1 and η_2 are the loss weights.

V. Experiments

A. Datasets and Data Preprocessing

Two publicly available datasets were used to evaluate our framework. The first includes 1206 young healthy subjects (mean age 28.19 ± 7.15, 657 women) from the Human Connectome Project (HCP) [84]. The second includes 1326 subjects (mean age $= 70.42 \pm 8.95$, 738 women) from the Open Access Series of Imaging Studies (OASIS) dataset [85]. Details of each dataset can be found on their official websites [1] [2]. CONN [86] were used to preprocess fMRI data and the preprocessing pipeline follows our previous publications [87], [88]. For HCP data, each subject's network has a dimension of 82×82 based on 82 ROIs defined using FreeSurfer (V6.0) [89]. For OASIS data, each subject's network has a dimension of 132×132 based on the Harvard-Oxford Atlas and AAL Atlas. We deliberately chose different network resolutions for HCP and OASIS to evaluate whether the performance of our new framework is affected by the network dimension or atlas.

B. Implementation Details

We randomly split the entire functional brain network dataset into 5 disjoint subsets for 5-fold cross-validations in our experiments. The values in the adjacency matrices ($\hat{A}$ and $\check{A}$) of brain functional networks are within range of $[-1, 1]$. We compute the kurtosis and skewness values of the fMRI BOLD signals as the node feature matrices (H). We use the Adam optimizer [90] to optimize the loss functions in our model with a batch size of 128. The initial learning rate is $1e^{-4}$ and decayed by $(1 - \frac{current\_epoch}{max\_epoch})^{0.9}$. We also regularized the training with an L_2 weight decay of $1e^{-5}$. We set the maximum number of training epochs as 1000 and, following the strategy in [34], [91], stop training if the validation loss does not decrease for 50 epochs. The experiments were deployed on one NVIDIA RTX A6000 GPU.

[1] https://www.oasis-brains.org
[2] https://wiki.humanconnectome.org

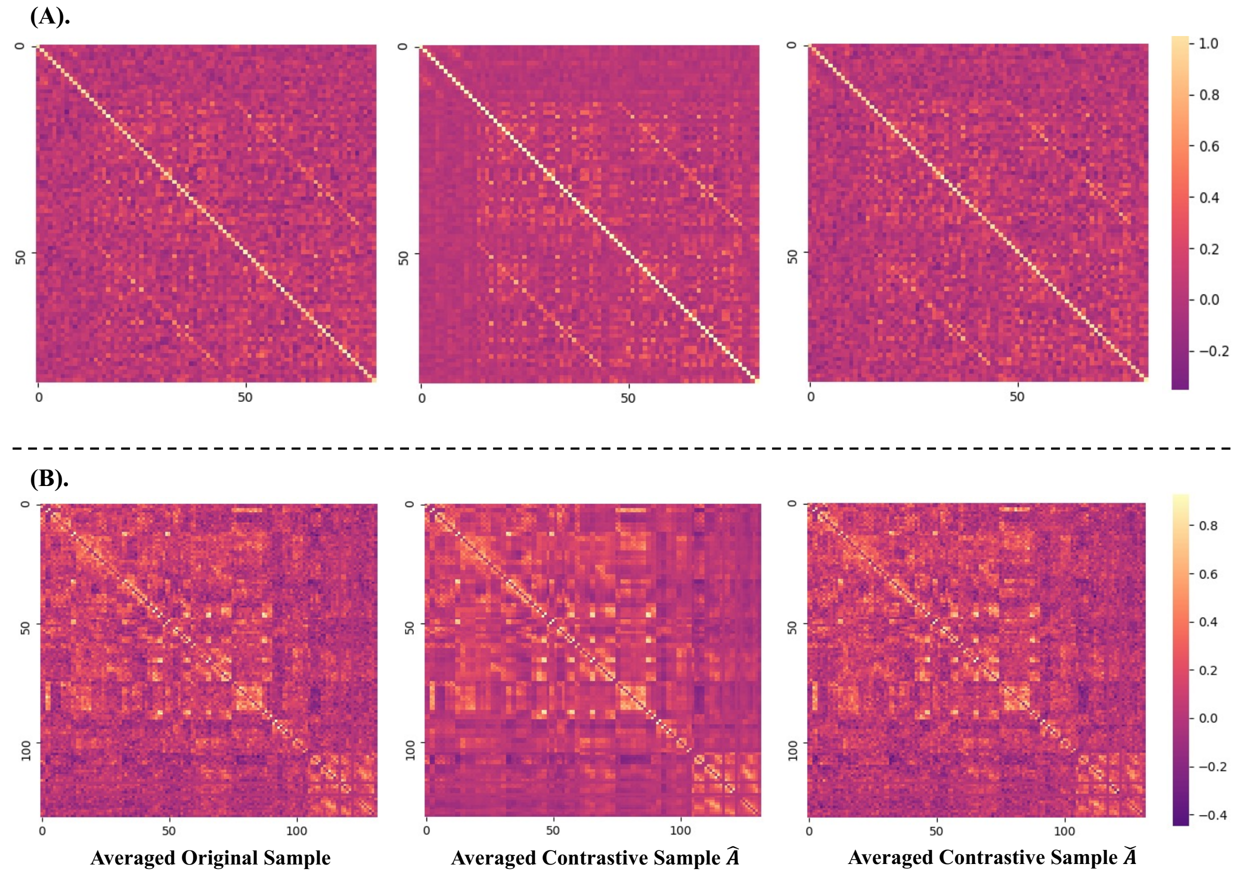


Fig. 2. Visualization of the averaged adjacency matrices for original and contrastive samples on **(A).** HCP dataset and **(B).** OASIS dataset. The averaged contrastive sample pair is generated by using a window size $d = 10$.

C. Similarities of Contrastive Samples

We utilize the L_1 distance and Cosine Similarity to measure the similarities of the adjacency matrices of contrastive brain networks. Here, we set the window size $d = 10$ to generate the contrastive adjacency matrices. The inner-pair similarity is computed by $\frac{1}{M}\sum_{m=1}^{M}\Psi(\hat{A}^m, \breve{A}^m)$, and the inter-pair similarity is computed by $\frac{1}{M^2}\sum_{m=1}^{M}\sum_{t=1}^{M}\Psi(\hat{A}^m, \breve{A}^t)$, where $\Psi(\cdot)$ is the similarity function (i.e., L_1 distance or Cosine Similarity). The inner-pair L_1 distances on HCP and OASIS data are 0.1301 and 0.0915, respectively. The inner-pair Cosine Similarities on HCP and OASIS data are 0.9283 and 0.9466, respectively. The inter-pair L_1 distances on HCP and OASIS data are 0.2925 and 0.3137, respectively. The inter-pair Cosine Similarities on HCP and OASIS data are 0.7311 and 0.7014, respectively. We visualize the averaged adjacency matrics on HCP and OASIS data in Figure 2 (A) and (B), respectively, to show their similarities. The original sample is generated by using the whole fMRI BOLD signal (i.e., $d = 0$).

D. Classification Tasks

1) Experiment Setup: For the comparison, we adopted seven baseline models, which include two traditional graph embedding models (t-BNE [72] and mCCA-ICA [73]), one basic graph neural network (i.e., GCN [26]), two deep graph representation learning models designed for brain network embedding (BrainChey [25] and BrainNet-CNN [24]), and two hierarchical graph neural networks with graph pooling strategies (DIFFPOOL [30] and SAGPOOL [34]). As aforementioned, existing GNN-based models cannot directly take signed graphs as the input, we therefore compute the absolute values of graph adjacency matrices as the input for these baseline models, which is consistent with previous studies [36], [92]. Meanwhile, we compare our model with and without optimizing contrastive loss to demonstrate the effectiveness of contrastive learning in boosting the model performance. The results for gender and Alzheimer Disease (AD) classification are reported in accuracy, precision and F1-score with their standard deviation (*std*). The results for zygosity classification (i.e., 3 classes classification task with class labels as: not twins, monozygotic twins and dizygotic twins) are reported in accuracy and Macro-F1-score with their *std*. The number of cascaded BUE and HGP modules are set to 3 and the number of top-K nodes in the pooling module is 50% of the number of nodes in the current graph. We search the loss weights η_1 and η_2 in range of $[0.1, 1, 5]$ and $[0.01, 0.1, 0.5, 1]$ respectively and determine the loss weights as $\eta_1 = 1$, $\eta_2 = 0.1$. The temperature parameter in contrastive loss is set as 0.2. Details of the hyperparameters analysis are shown in section V-F.

2) Results: Table I shows the results of gender classification, zygosity classification and AD classification. It shows that our model achieves the best performance comparing to all

TABLE I
CLASSIFICATION ACCURACY WITH S.T.D VALUES UNDER 5-FOLD CROSS-VALIDATION ON GENDER CLASSIFICATION, ZYGOSITY CLASSIFICATION AND AD CLASSIFICATION TASKS. THE VALUES IN **BOLD** SHOW THE BEST RESULTS.

Method	HCP					OASIS		
	Gender			Zygosity		AD		
	Acc.	Pre.	F1.	Acc.	Macro-F1.	Acc.	Pre.	F1.
t-BNE	63.84(2.09)	64.17(1.90)	63.264(2.12)	37.19(2.65)	39.67(3.04)	61.26(2.31)	63.58(2.06)	62.05(1.97)
mCCA-ICA	61.21(4.03)	63.11(3.75)	62.20(3.59)	35.51(4.64)	38.71(3.34)	63.37(1.98)	62.06(2.12)	64.37(2.09)
GCN	66.76(2.22)	65.09(3.13)	67.58(2.84)	46.66(2.14)	47.21(2.51)	67.37(2.69)	69.21(2.00)	68.51(4.29)
SAGPOOL	68.12(3.07)	69.96(2.48)	67.51(2.65)	49.91(2.22)	51.07(2.31)	67.23(2.15)	68.83(1.13)	67.51(2.51)
DIFFPOOL	72.06(2.28)	74.05(1.90)	73.07(2.42)	53.37(1.88)	54.28(2.14)	72.79(1.66)	71.55(2.15)	70.83(2.01)
BrainCheby	75.08(1.98)	76.14(2.38)	74.09(1.84)	56.25(2.12)	57.37(2.05)	72.55(2.45)	73.36(1.88)	72.62(1.33)
BrainNet-CNN	74.09(2.49)	73.71(1.96)	73.27(2.21)	54.03(2.20)	55.25(2.46)	68.37(1.71)	69.97(1.30)	68.51(2.02)
Ours w/o Contrastive	**78.86(2.18)**	**80.06(1.33)**	**77.52(1.69)**	**61.05(1.70)**	**63.24(2.51)**	**76.26(2.32)**	**75.42(1.62)**	**76.80(1.72)**
Ours	**81.51(1.14)**	**82.37(1.95)**	**80.69(2.03)**	**63.33(2.06)**	**64.51(1.74)**	**77.51(1.84)**	**78.83(1.78)**	**78.28(1.95)**

TABLE II
REGRESSION MEAN ABSOLUTE ERROR (MAE) WITH S.T.D UNDER 5-FOLD CROSS-VALIDATION. THE VALUES IN **BOLD** SHOW THE BEST RESULTS.

Method	OASIS	HCP				
	MMSE	Flanker	Card-Sort	Aggressive	Intrusive	Rule-Break
t-BNE	2.02(0.36)	1.69(0.19)	1.58(0.22)	1.89(0.10)	1.84(0.22)	1.77(0.41)
mCCA-ICA	2.68(0.19)	1.82(0.21)	1.67(0.17)	1.47(0.26)	1.97(0.13)	1.61(0.29)
GCN	2.05(0.07)	1.67(0.15)	1.46(0.11)	1.59(0.32)	1.66(0.24)	1.69(0.08)
SAGPOOL	1.84(0.33)	1.55(0.06)	1.44(0.13)	1.52(0.18)	1.50(0.24)	1.74(0.23)
DIFFPOOL	1.27(0.20)	1.34(0.14)	1.16(0.30)	1.27(0.41)	1.25(0.07)	1.43(0.15)
Brain-Cheby	1.51(0.67)	1.17(0.26)	1.24(0.31)	0.79(0.06)	1.09(0.21)	1.58(0.41)
BrainNetCNN	1.26(0.19)	1.43(0.24)	**0.91(0.11)**	1.33(0.23)	1.14(0.13)	1.29(0.19)
Ours w/o Contrastive	**1.02(0.11)**	**0.89(0.13)**	0.97(0.20)	**0.74(0.17)**	**0.96(0.15)**	**1.15(0.11)**
Ours	**0.83(0.24)**	**0.66(0.17)**	**0.69(0.14)**	**0.45(0.12)**	**0.73(0.08)**	**1.02(0.16)**

baseline methods on three tasks. For example, in the gender classification, our model outperforms the baselines with at least 8.56%, 8.18% and 8.91% increases in accuracy, precision and F1 scores, respectively. In general, the deep graph neural networks are superior than the traditional graph embedding methods (i.e., t-BNE and mCCA-ICA). When we remove the supervision of the contrastive loss, the performance, though comparable to baselines, decreases in a way. This manifests the effectiveness of the contrastive learning which can substantially boost the model performance.

E. Regression tasks

1) Experiment Setup: In the regression tasks, we use the same baselines for comparisons. The regression tasks include predicting MMSE scores on OASIS data, Flanker scores, Card-Sort scores, and 3 ASR scores (i.e., Aggressive, Intrusive and Rule-Break scores) on HCP data. Particularly, MMSE (Mini-Mental State Exam) test [93], Flanker test [94] and Wisconsin Card-Sort test [95]–[97] are 3 neuropsychological tests designed to measure the status and risks of human neurodegenerative disease and mental illness. The ASR (Achenbach Adult Self-Report) is a life function which is used to measure the emotion and social support of adults. The structure of proposed model remains unchanged. The loss weights are set as $\eta_1 = 0.5$ and $\eta_2 = 1$. The regression results are reported in average Mean Absolute Errors (MAE) with its *std* under 5-fold cross validations.

2) Results: The regression results are presented in Table II. It shows that our model achieves the best MAE values comparing to all baseline methods. Similar to the classification tasks, the deep graph neural networks are superior than traditional graph embedding methods (i.e., t-BNE and mCCA-ICA). Comparing our method with and without the supervision of the contrastive loss, we can hold the conclusion that the contrastive learning can further boost the model performance.

F. Ablation Studies

In this section, we investigate the effect of 4 hyperparameters on our model performance, including (1) the window size (d) which we used to clamp the fMRI BOLD signals when generating contrastive functional brain network samples, (2) temperature parameter (α) within contrastive loss, (3) the number of BUE and HGP modules utilized in HSGRL model, and (4) loss weights η_1 and η_2. First, we set the window size as $[0, 5, 10, 20, 30, 40, 50]$, respectively and generate different contrastive samples as the input of our proposed model. The first column in Figure 3 shows the analysis of the window size parameter. It indicates that the best window size is around $d = 10$. When the window size decreases to 0, the model performance declines since the data is only duplicated without any substantial new samples. It is interesting that the performance when $d = 0$ is even worse than that obtained without contrastive learning but with contrastive samples generated with $d = 10$ (see Ours w/o Contrastive in Table I and II). The reason is that data augmentation is introduced in the latter case but not in the first case. Second, we increase the temperature α from 0.1 to 1.0 with a step of 0.1. The second column in Figure 3 demonstrates the analysis of the temperature parameter. It shows that the best temperature value for our framework is $\alpha = 0.2$. Moreover, we set the number of BUE and HGP modules as $[1, 2, 3, 4, 5]$, respectively for our framework. The third column in Figure 3 shows the analysis of this parameter.

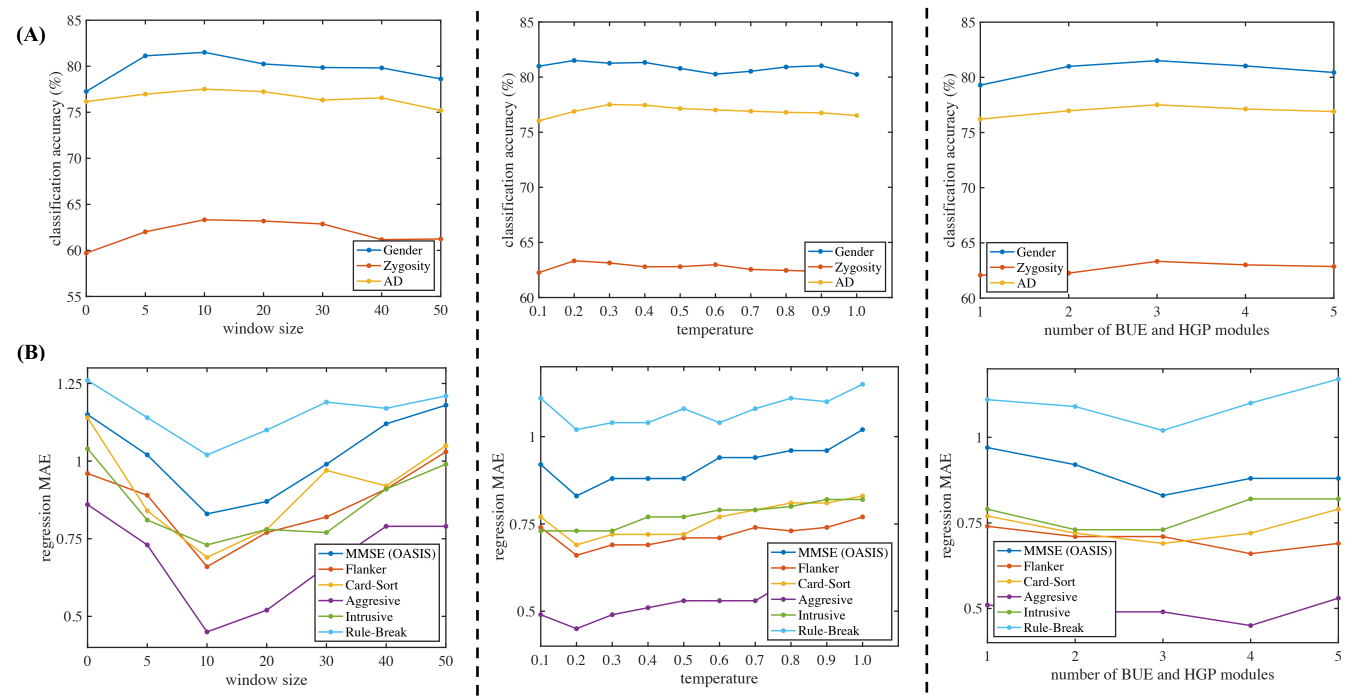


Fig. 3. Parameter analysis. The model performance obtained with: contrastive samples generated by different window sizes (Column 1), different temperature parameters in contrastive loss (Column 2), and different number of BUE and HGP modules (Column 3). **(A)** shows the analysis on classification tasks and **(B)** shows the analysis on regression tasks.

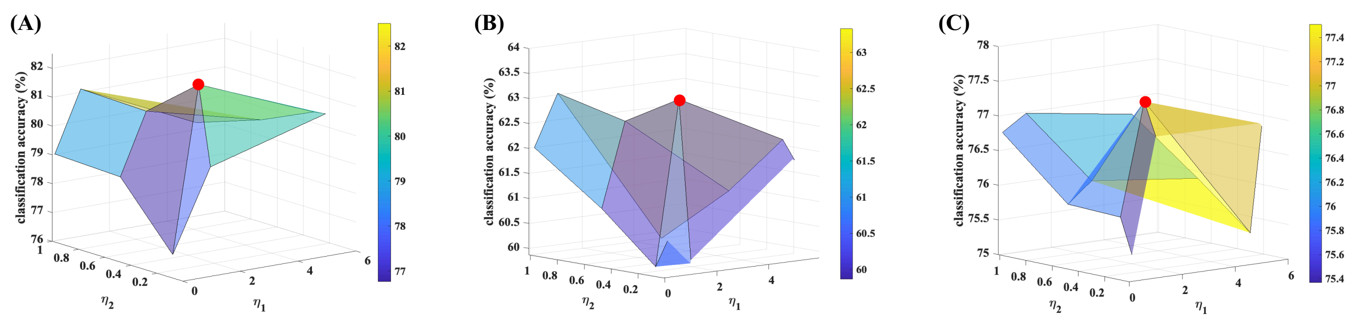


Fig. 4. Loss weights analysis on classification tasks. **(A)** shows the analysis on gender classification, **(B)** shows the analysis on zygosity classification and **(C)** shows the analysis on AD classification. The red points represent the best results, where $\eta_1 = 1$ and $\eta_2 = 0.1$.

It manifests that the framework performance is consistent and steady when different number of BUE and HGP modules are deployed. The best number of the modules for almost all tasks are 3, except for the regression tasks on Flanker and Aggressive. Finally, we present the loss weights analysis (see Figure 4) on the 3 classification tasks and the best results are achieved when $\eta_1 = 1$ and $\eta_2 = 0.1$.

G. Interpretation with Brain Saliency Map

Within our new graph pooling module, an information score is designed to measure the information gain for each brain node and only top-K nodes with high information gains will be preserved as brain information hubs while the information of other peripheral nodes will be aggregated onto these hubs. These hubs, through the final pooling layer, will serve as the delegate of the whole brain network and then be linked to clinical phenotypes (e.g., clinical/behavior scores or diagnosis). Therefore, they can provide hints for further clinical analyses on how this phenotype is associated with brain functional network from the global view. We utilize the Class Activation Mapping (CAM) approach [98]–[100] to generate the brain network saliency map, which indicates the top brain regions associated with each prediction task. Figures 5 and 6 illustrate Brain Saliency Maps for classification and regression tasks, respectively. For example, in the classification task (AD vs. NC), the saliency map for AD highlights multiple regions (such as Planum Polare, Frontal Operculum cortex, Supracalcarine Cortex, etc.) which are conventionally conceived as the biomarkers of AD in medical imaging analysis [101]–[104]. In the meantime, the saliency map for NC highlights many regions in Cerebellum and Frontal lobe. These regions control cognitive thinking, motor control, and social mentalizing as well as emotional self-experiences [105]–[107], in which AD patients typically show problems. Another example is the classification of Male vs. Female on HCP data. Females are more ”emotional” or ”sensitive”, suggested by the regions such as isthmuscingulate and caudalanteriorcingulate while

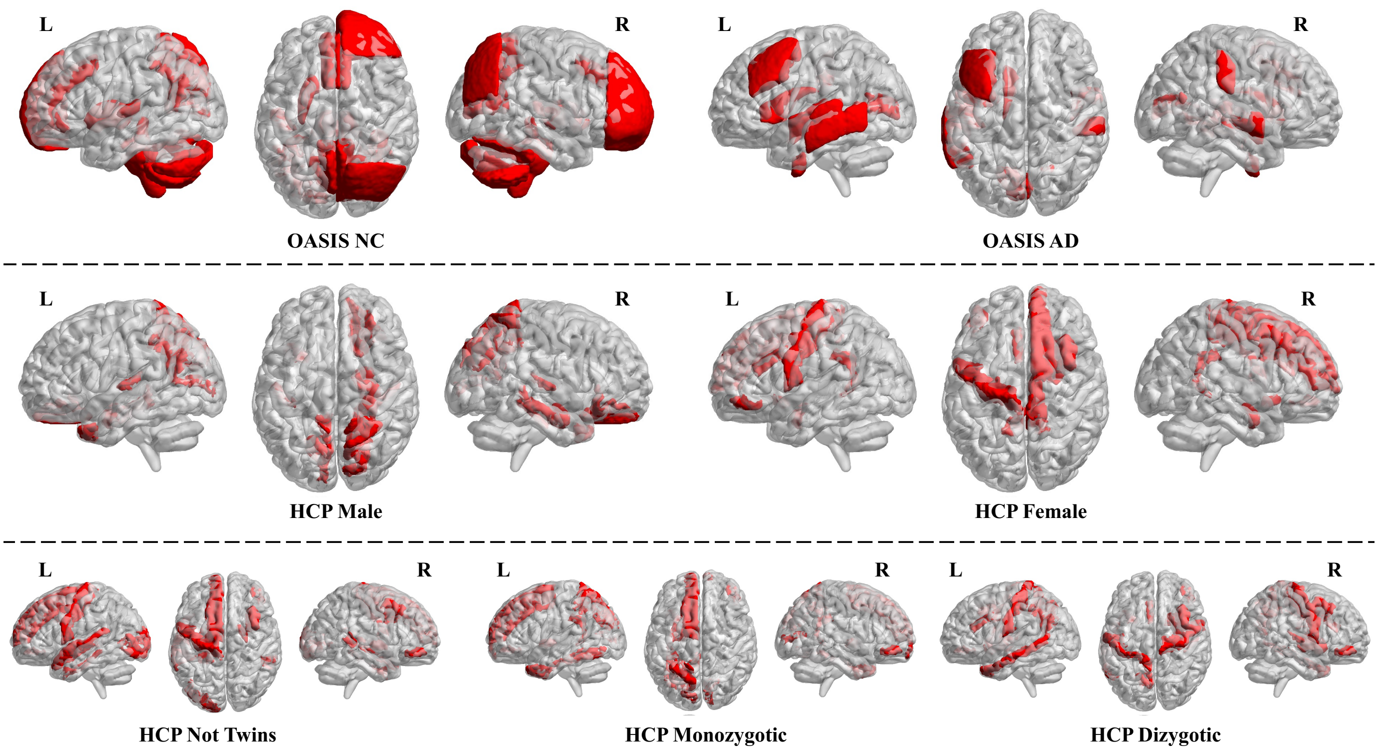


Fig. 5. Brain saliency maps for classification tasks. Here we identify: (1) top 15 regions associated with AD and NC from OASIS, (2) top 10 regions associated with each sex and each zygosity from HCP.

males tend to be more competitive and dominant, manifested in regions such as lateralorbitofrontal and precuneus. These results are consistent with previous findings in the literature [108]–[111]. The details for all highlighted brain regions for each task are summarized in the Table III and Table IV a and b. These highlighted regions can help us locating brain regions associated with any phenotype, which provide clues for future clinical investigations.

VI. Conclusion

We propose a novel contrastive learning framework with an interpretable hierarchical signed graph representation learning model for brain functional network mining. Additionally, a new data augmentation strategy is designed to generate the contrastive samples for brain functional network data. Our new framework is capable of generating more accurate representations for brain functional networks in compared with other state-of-the-art methods and these network representations can be used in various prediction tasks (e.g., classification and regression). Moreover, Brain saliency maps may assist with phenotypic biomarker identification and provide interpretable explanation on framework outcomes.

Acknowledgement

This study is partially supported by The National Institutes of Health (R01AG071243, R01MH125928 and U01AG068057) and National Science Foundation (IIS 2045848 and IIS 1837956).

Data were provided [in part] by the Human Connectome Project, MGH-USC Consortium (Principal Investigators: Bruce R. Rosen, Arthur W. Toga and Van Wedeen; U01MH093765) funded by the NIH Blueprint Initiative for Neuroscience Research grant; the National Institutes of Health grant P41EB015896, and the Instrumentation Grants S10RR023043, 1S10RR023401, 1S10RR019307.

Part of the work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC).

References

[1] H. Rusinek, S. De Santi, D. Frid, W.-H. Tsui, C. Y. Tarshish, A. Convit, and M. J. de Leon, "Regional brain atrophy rate predicts future cognitive decline: 6-year longitudinal mr imaging study of normal aging," *Radiology*, vol. 229, no. 3, pp. 691–696, 2003.

[2] M. R. Sabuncu and E. Konukoglu, "Clinical prediction from structural brain mri scans: a large-scale empirical study," *Neuroinformatics*, vol. 13, no. 1, pp. 31–46, 2015.

[3] S. Seo, J. Mohr, A. Beck, T. Wüstenberg, A. Heinz, and K. Obermayer, "Predicting the future relapse of alcohol-dependent patients from structural and functional brain images," *Addiction biology*, vol. 20, no. 6, pp. 1042–1055, 2015.

[4] Y. Zhang, L. Zhan, W. Cai, P. Thompson, and H. Huang, "Integrating heterogeneous brain networks for predicting brain disease conditions," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 214–222.

[5] S. Genon, A. Reid, R. Langner, K. Amunts, and S. B. Eickhoff, "How to characterize the function of a brain region," *Trends in cognitive sciences*, vol. 22, no. 4, pp. 350–364, 2018.

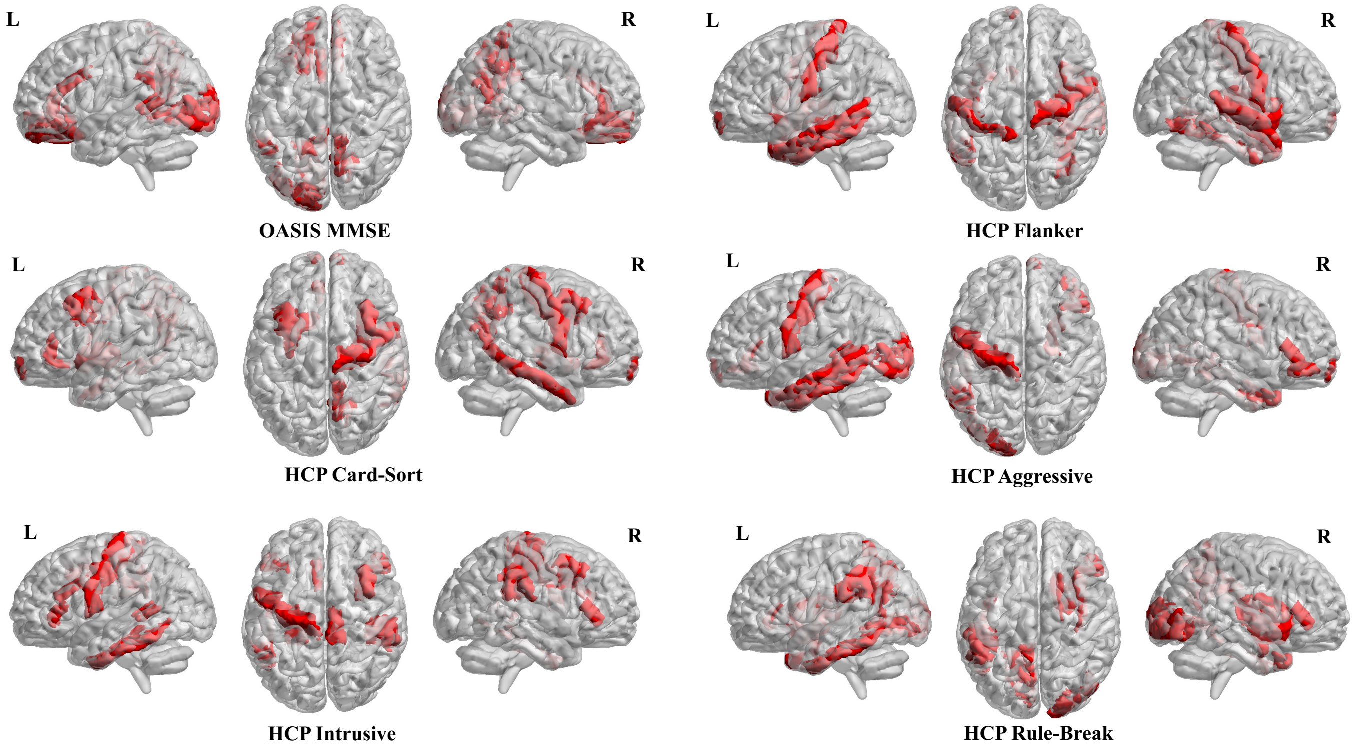


Fig. 6. Brain saliency maps for regression tasks. Here we identify: (1) top 15 regions associated with MMSE from OASIS, (2) top 10 regions associated with Flanker score, Card-Sort score, Aggressive score, Intrusive score and Rule-Break score from HCP.

TABLE III
THE LIST OF HIGHLIGHTED BRAIN REGIONS FOR OASIS DATASET, INCLUDING AD AND NC CLASSIFICATION TASKS AND MMSE REGRESSION TASK.

AD	Planum Polare Left	Frontal Operculum Cortex Left	Supracalcarine Cortex Left	Left-Caudate	Supramarginal Gyrus, anterior division Right	Superior Temporal Gyrus, anterior division Right	Middle Temporal Gyrus, posterior division Left	Superior Temporal Gyrus, posterior division Left
	Heschl's Gyrus Left	Intracalcarine Cortex Left	Middle Frontal Gyrus Left	Planum Polare Right	Temporal Fusiform Cortex, anterior division Left	Middle Temporal Gyrus, temporooccipital part Left	Supracalcarine Cortex Right	—
NC	Paracingulate Gyrus Right	Intracalcarine Cortex Right	Frontal Pole Right	Cerebelum 6 Right	Paracingulate Gyrus Left	Left-Putamen	Cerebelum 8 Left	Cerebelum 7b Right
	Heschl's Gyrus Left	Cuneal Cortex Right	Precuneous Cortex	Cerebelum Crus2 Left	Lateral Occipital Cortex, superior division Right	Brain-Stem	Cerebelum 8 Right	—
MMSE	Right-Caudate	Temporal Pole Right	Planum Temporale Left	Cerebelum Crus1 Right	Middle Temporal Gyrus, posterior division Right	Temporal Occipital Fusiform Cortex Left	Temporal Occipital Fusiform Cortex Right	Middle Temporal Gyrus, temporooccipital part Left
	Planum Temporale Right	Frontal Orbital Cortex Left	Vermis 9	Temporal Pole Left	Middle Temporal Gyrus, temporooccipital part Right	Left-Caudate	Temporal Pole Left	—

[6] N. Kraljević, H. L. Schaare, S. B. Eickhoff, P. Kochunov, B. T. Yeo, S. K. Masouleh, and S. L. Valk, "Behavioral, anatomical and heritable convergence of affect and cognition in superior frontal cortex," *NeuroImage*, vol. 243, p. 118561, 2021.

[7] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in cognitive sciences*, vol. 14, no. 6, pp. 277–290, 2010.

[8] H. Tang, L. Guo, X. Fu, B. Qu, O. Ajilore, Y. Wang, P. M. Thompson, H. Huang, A. D. Leow, and L. Zhan, "A hierarchical graph learning model for brain network regression analysis," *Frontiers in Neuroscience*, vol. 16, 2022.

[9] H. Sun, W. Zhang, H. Cao, H. Sun, J. Dai, S. Li, J. Zeng, X. Wei, B. Tang, Q. Gong *et al.*, "Linked brain connectivity patterns with psychopathological and cognitive phenotypes in drug-naïve first-episode schizophrenia," *Psychoradiology*, vol. 2, no. 2, pp. 43–51, 2022.

[10] R. W. Levenson, V. E. Sturm, and C. M. Haase, "Emotional and behavioral symptoms in neurodegenerative disease: a model for studying the neural bases of psychopathology," *Annual review of clinical psychology*, vol. 10, p. 581, 2014.

[11] J. Yan, Y. Chen, Z. Xiao, S. Zhang, M. Jiang, T. Wang, T. Zhang, J. Lv, B. Becker, R. Zhang *et al.*, "Modeling spatio-temporal patterns of holistic functional brain networks via multi-head guided attention graph neural networks (multi-head gagnns)," *Medical Image Analysis*, vol. 80, p. 102518, 2022.

[12] P. L. Baniqued, C. L. Gallen, M. W. Voss, A. Z. Burzynska, C. N. Wong, G. E. Cooke, K. Duffy, J. Fanning, D. K. Ehlers, E. A. Salerno *et al.*, "Brain network modularity predicts exercise-related executive function gains in older adults," *Frontiers in aging neuroscience*, vol. 9, p. 426, 2018.

[13] U. Braun, S. F. Muldoon, and D. S. Bassett, "On human brain networks in health and disease," *eLS*, pp. 1–9, 2015.

[14] M. P. Van Den Heuvel, R. S. Kahn, J. Goñi, and O. Sporns, "High-cost, high-capacity backbone for global brain communication," *Proceedings of the National Academy of Sciences*, vol. 109, no. 28, pp. 11 372–11 377, 2012.

[15] O. Sporns, "The human connectome: origins and challenges," *Neuroimage*, vol. 80, pp. 53–61, 2013.

[16] M. G. Mattar and D. S. Bassett, "Brain network architecture," *Network science in cognitive psychology*, p. 30, 2019.

[17] Y. Zhang, L. Zhan, S. Wu, P. Thompson, and H. Huang, "Disentangled and proportional representation learning for multi-view brain connectomes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 508–518.

[18] R. E. Beaty, Y. N. Kenett, A. P. Christensen, M. D. Rosenberg, M. Benedek, Q. Chen, A. Fink, J. Qiu, T. R. Kwapil, M. J. Kane *et al.*, "Robust prediction of individual creative ability from brain functional

TABLE IV
THE LIST OF HIGHLIGHTED BRAIN REGIONS FOR HCP DATASET, WHERE (A) SHOWS THE RESULTS ON CLASSIFICATION TASKS AND (B) SHOWS THE RESULTS ON REGRESSION TASKS.

Male	Female	Not Twins	Monozygotic	Dizygotic
ctx-lh-precuneus	ctx-rh-superiorfrontal	ctx-lh- lateraloccipital	ctx-lh- isthmuscingulate	ctx-lh-postcentral
ctx-rh- superiorparietal	Right-Accumbens-area	ctx-rh-bankssts	ctx-rh-pericalcarine	ctx-rh- transversetemporal
Right-Hippocampus	ctx-rh- caudalmiddlefrontal	ctx-lh-precentral	ctx-rh-frontalpole	ctx-rh- transversetemporal
ctx-rh- parahippocampal	ctx-lh-parsorbitalis	ctx-lh- parahippocampal	ctx-lh-fusiform	Paracingulate Gyrus Right Paracingulate
Right-Amygdala	Right-Amygdala	ctx-lh-entorhinal	ctx-lh-entorhinal	ctx-lh- caudalanteriorcingulate
ctx-lh-pericalcarine	ctx-rh-paracentral	Right-Pallidum	ctx-lh- superiorfrontal	ctx-rh-parsorbitalis Right-Putamen
ctx-lh- transversetemporal	ctx-lh-precentral	ctx-lh- superiortemporal	ctx-lh- temporalpole	ctx-rh-precentral
ctx-rh- transversetemporal	ctx-lh- isthmuscingulate	ctx-rh-parsorbitalis	ctx-lh- superiorparietal	ctx-rh- caudalmiddlefrontal
ctx-rh- lateralorbitofrontal	ctx-rh- isthmuscingulate	ctx-lh-superiorfrontal	Left-Pallidum	ctx-lh-precuneus
ctx-lh- temporalpole	ctx-lh- caudalanteriorcingulate	ctx-rh- caudalmiddlefrontal	ctx-rh-parsorbitalis	ctx-lh-temporalpole

(a)

Flanker	Card-Sort	Aggressive	Intrusive	Rule-Break
Left-Accumbens-area	Left-Accumbens-area	ctx-lh-bankssts	ctx-lh-bankssts	ctx-lh-precuneus
ctx-lh-inferiortemporal	ctx-lh- caudalmiddlefrontal	ctx-lh- inferiortemporal	ctx-lh- inferiortemporal	ctx-lh- inferiortemporal
ctx-rh-insula	ctx-rh-frontalpole	ctx-lh- lateraloccipital	ctx-lh- parahippocampal	Right-Caudate
ctx-lh- middletemporal	ctx-lh- rostralanteriorcingulate	ctx-lh-precentral	ctx-rh- supramarginal	ctx-rh- lateraloccipital
ctx-lh-postcentral	ctx-rh- middletemporal	ctx-rh-frontalpole	ctx-rh-paracentral	ctx-lh- supramarginal
ctx-lh-temporalpole	ctx-lh-frontalpole	ctx-rh-parsorbitalis	ctx-rh- parstriangularis	ctx-rh-insula
ctx-rh- superiortemporal	ctx-rh-precentral	ctx-rh- parstriangularis	ctx-lh- caudalanteriorcingulate	ctx-rh- parstriangularis
ctx-lh-frontalpole	ctx-rh- caudalmiddlefrontal	ctx-lh- middletemporal	ctx-lh-precentral	ctx-lh-lingual
ctx-rh-precentral	ctx-rh-precuneus	ctx-rh-entorhinal	ctx-rh- caudalmiddlefrontal	ctx-rh- temporalpole
ctx-rh-fusiform	Left-Putamen	ctx-rh-temporalpole	ctx-lh-parsorbitalis	Right-Amygdala

(b)

connectivity," *Proceedings of the National Academy of Sciences*, vol. 115, no. 5, pp. 1087–1092, 2018.

[19] C. J. Brown, K. P. Moriarty, S. P. Miller, B. G. Booth, J. G. Zwicker, R. E. Grunau, A. R. Synnes, V. Chau, and G. Hamarneh, "Prediction of brain network age and factors of delayed maturation in very preterm infants," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 84–91.

[20] T. Eichele, S. Debener, V. D. Calhoun, K. Specht, A. K. Engel, K. Hugdahl, D. Y. Von Cramon, and M. Ullsperger, "Prediction of human errors by maladaptive changes in event-related brain networks," *Proceedings of the National Academy of Sciences*, vol. 105, no. 16, pp. 6173–6178, 2008.

[21] X. Li, Y. Li, and X. Li, "Predicting clinical outcomes of alzheimer's disease from complex brain networks," in *International Conference on Advanced Data Mining and Applications*. Springer, 2017, pp. 519–525.

[22] D. E. Warren, N. L. Denburg, J. D. Power, J. Bruss, E. J. Waldron, H. Sun, S. E. Petersen, and D. Tranel, "Brain network theory can predict whether neuropsychological outcomes will differ from clinical expectations," *Archives of Clinical Neuropsychology*, vol. 32, no. 1, pp. 40–52, 2017.

[23] C. Hu, R. Ju, Y. Shen, P. Zhou, and Q. Li, "Clinical decision support for alzheimer's disease based on deep learning and brain network," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.

[24] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

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

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

[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[28] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," *Advances in neural information processing systems*, vol. 32, 2019.

[29] R. Bao, B. Gu, and H. Huang, "Fast oscar and owl regression via safe screening rules," in *International Conference on Machine Learning*. PMLR, 2020, pp. 653–663.

[30] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *Advances in neural information processing systems*, vol. 31, 2018.

[31] C. C. Hilgetag and A. Goulas, "'hierarchy'in the organization of brain networks," *Philosophical Transactions of the Royal Society B*, vol. 375, no. 1796, p. 20190319, 2020.

[32] R. Mastrandrea, A. Gabrielli, F. Piras, G. Spalletta, G. Caldarelli, and T. Gili, "Organization and hierarchy of the human functional brain network lead to a chain-like core," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017.

[33] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. T. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in neuroinformatics*, vol. 3, p. 37, 2009.

[34] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *International conference on machine learning*. PMLR, 2019, pp. 3734–3743.

[35] Z. Zhang, J. Bu, M. Ester, J. Zhang, C. Yao, Z. Yu, and C. Wang, "Hierarchical graph pooling with structure learning," *arXiv preprint arXiv:1911.05954*, 2019.

[36] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, vol. 74, p. 102233, 2021.

[37] H. Tang, G. Ma, L. He, H. Huang, and L. Zhan, "Commpool: An interpretable graph pooling framework for hierarchical graph representation learning," *Neural Networks*, vol. 143, pp. 669–677, 2021.

[38] D. Cartwright and F. Harary, "Structural balance: a generalization of heider's theory." *Psychological review*, vol. 63, no. 5, p. 277, 1956.

[39] T. Derr, Y. Ma, and J. Tang, "Signed graph convolutional networks," in

2018 IEEE International Conference on Data Mining (ICDM). IEEE, 2018, pp. 929–934.

[40] F. Heider, "Attitudes and cognitive organization," *The Journal of psychology*, vol. 21, no. 1, pp. 107–112, 1946.

[41] Y. Li, Y. Tian, J. Zhang, and Y. Chang, "Learning signed network embedding via graph attention," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4772–4779.

[42] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," *arXiv preprint arXiv:1511.05493*, 2015.

[43] O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: Sequence to sequence for sets," *arXiv preprint arXiv:1511.06391*, 2015.

[44] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An end-to-end deep learning architecture for graph classification," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[45] M. P. van den Heuvel and O. Sporns, "Network hubs in the human brain," *Trends in cognitive sciences*, vol. 17, no. 12, pp. 683–696, 2013.

[46] M. U. Ilyas, M. Z. Shafiq, A. X. Liu, and H. Radha, "A distributed and privacy preserving algorithm for identifying information hubs in social networks," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 561–565.

[47] K. Hwang, M. N. Hallquist, and B. Luna, "The development of hub architecture in the human functional brain network," *Cerebral Cortex*, vol. 23, no. 10, pp. 2380–2393, 2013.

[48] L. Zhan, L. M. Jenkins, O. E. Wolfson, J. J. GadElkarim, K. Nocito, P. M. Thompson, O. A. Ajilore, M. K. Chung, and A. D. Leow, "The significance of negative correlations in brain connectivity," *Journal of Comparative Neurology*, vol. 525, no. 15, pp. 3251–3265, 2017.

[49] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 661–18 673, 2020.

[50] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[51] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6256–6268, 2020.

[52] Y. Wu, Z. Wang, D. Zeng, M. Li, Y. Shi, and J. Hu, "Decentralized unsupervised learning of visual representations," pp. 2326–2333, 2022.

[53] Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu, "Federated contrastive learning for volumetric medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 367–377.

[54] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *International Conference on Machine Learning*. PMLR, 2020, pp. 4116–4126.

[55] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 5812–5823, 2020.

[56] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Graph contrastive learning with adaptive augmentation," in *Proceedings of the Web Conference 2021*, 2021, pp. 2069–2080.

[57] T. Zhao, Y. Liu, L. Neves, O. Woodford, M. Jiang, and N. Shah, "Data augmentation for graph neural networks," *arXiv preprint arXiv:2006.06830*, 2020.

[58] Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu, "Distributed contrastive learning for medical image segmentation," *Medical Image Analysis*, vol. 81, p. 102564, 2022.

[59] L. Zhang, A. Zaman, L. Wang, J. Yan, and D. Zhu, "A cascaded multi-modality analysis in mild cognitive impairment," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 557–565.

[60] L. Zhang, L. Wang, and D. Zhu, "Jointly analyzing alzheimer's disease related structure-function using deep cross-model attention network," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 563–567.

[61] J. Chen, T. Ma, and C. Xiao, "Fastgcn: fast learning with graph convolutional networks via importance sampling," *arXiv preprint arXiv:1801.10247*, 2018.

[62] W. Huang, T. Zhang, Y. Rong, and J. Huang, "Adaptive sampling towards fast graph representation learning," in *Advances in neural information processing systems*, 2018, pp. 4558–4567.

[63] H. Dai, B. Dai, and L. Song, "Discriminative embeddings of latent variable models for structured data," in *International conference on machine learning*, 2016, pp. 2702–2711.

[64] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Advances in neural information processing systems*, 2015, pp. 2224–2232.

[65] J. Liu, G. Ma, F. Jiang, C.-T. Lu, S. Y. Philip, and A. B. Ragin, "Community-preserving graph convolutions for structural and functional joint embedding of brain networks," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 1163–1168.

[66] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in neural information processing systems*, 2017, pp. 1024–1034.

[67] G. Ma, N. K. Ahmed, T. L. Willke, and P. S. Yu, "Deep graph similarity learning: A survey," *Data Mining and Knowledge Discovery*, vol. 35, no. 3, pp. 688–725, 2021.

[68] H. Gao and S. Ji, "Graph u-nets," in *international conference on machine learning*. PMLR, 2019, pp. 2083–2092.

[69] H. Yuan and S. Ji, "Structpool: Structured graph pooling via conditional random fields," in *Proceedings of the 8th International Conference on Learning Representations*, 2020.

[70] L. E. Korthauer, L. Zhan, O. Ajilore, A. Leow, and I. Driscoll, "Disrupted topology of the resting state structural connectome in middle-aged apoe $\varepsilon 4$ carriers," *Neuroimage*, vol. 178, pp. 295–305, 2018.

[71] L. Zhan, Y. Liu, J. Zhou, J. Ye, and P. M. Thompson, "Boosting classification accuracy of diffusion mri derived brain networks for the subtypes of mild cognitive impairment using higher order singular value decomposition," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 131–135.

[72] B. Cao, L. He, X. Wei, M. Xing, P. S. Yu, H. Klumpp, and A. D. Leow, "t-bne: Tensor-based brain network embedding," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 2017, pp. 189–197.

[73] J. Sui, G. Pearlson, A. Caprihan, T. Adali, K. A. Kiehl, J. Liu, J. Yamamoto, and V. D. Calhoun, "Discriminating schizophrenia and bipolar disorder by fusing fmri and dti in a multimodal cca+ joint ica model," *Neuroimage*, vol. 57, no. 3, pp. 839–855, 2011.

[74] Y. Zhang and H. Huang, "New graph-blind convolutional network for brain connectome data analysis," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 669–681.

[75] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, "Hi-gcn: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction," *Computers in Biology and Medicine*, vol. 127, p. 104096, 2020.

[76] J. Jung, J. Yoo, and U. Kang, "Signed graph diffusion network," *arXiv preprint arXiv:2012.14191*, 2020.

[77] X. Shen and F.-L. Chung, "Deep network embedding for graph representation learning in signed networks," *IEEE transactions on cybernetics*, vol. 50, no. 4, pp. 1556–1568, 2018.

[78] Q. Huang, M. Yamada, Y. Tian, D. Singh, D. Yin, and Y. Chang, "Graphlime: Local interpretable model explanations for graph neural networks," *arXiv preprint arXiv:2001.06216*, 2020.

[79] H. Cui, W. Dai, Y. Zhu, X. Li, L. He, and C. Yang, "Brainnnexplainer: An interpretable graph neural network framework for brain network based disease analysis," *arXiv preprint arXiv:2107.05097*, 2021.

[80] D. Xu, W. Cheng, D. Luo, H. Chen, and X. Zhang, "Infogcl: Information-aware graph contrastive learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[81] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3733–3742.

[82] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," *Advances in neural information processing systems*, vol. 29, 2016.

[83] A. Van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv e-prints*, pp. arXiv–1807, 2018.

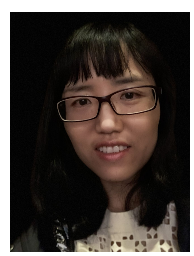
[84] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[85] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko *et al.*, "Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease," *MedRxiv*, 2019.

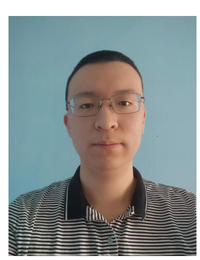
[86] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain connectivity*, vol. 2, no. 3, pp. 125–141, 2012.
[87] I. Fortel, L. E. Korthauer, Z. Morrissey, L. Zhan, O. Ajilore, O. Wolfson, I. Driscoll, D. Schonfeld, and A. Leow, "Connectome signatures of hyperexcitation in cognitively intact middle-aged female apoe-ε4 carriers," *Cerebral Cortex*, vol. 30, no. 12, pp. 6350–6362, 2020.
[88] O. Ajilore, L. Zhan, J. GadElkarim, A. Zhang, J. Feusner, S. Yang, P. M. Thompson, A. Kumar, and A. Leow, "Constructing the resting state structural connectome," *Frontiers in neuroinformatics*, vol. 7, p. 30, 2013.
[89] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
[90] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[91] O. Shchur, M. Mumme, A. Bojchevski, and S. Günnemann, "Pitfalls of graph neural network evaluation," *arXiv preprint arXiv:1811.05868*, 2018.
[92] L. Zhang, L. Wang, D. Zhu, A. D. N. Initiative *et al.*, "Predicting brain structural network using functional connectivity," *Medical Image Analysis*, vol. 79, p. 102463, 2022.
[93] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: a comprehensive review," *Journal of the American Geriatrics Society*, vol. 40, no. 9, pp. 922–935, 1992.
[94] B. A. Eriksen and C. W. Eriksen, "Effects of noise letters upon the identification of a target letter in a nonsearch task," *Perception & psychophysics*, vol. 16, no. 1, pp. 143–149, 1974.
[95] V. C. Pangman, J. Sloan, and L. Guse, "An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice," *Applied Nursing Research*, vol. 13, no. 4, pp. 209–213, 2000.
[96] O. Monchi, M. Petrides, V. Petre, K. Worsley, and A. Dagher, "Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging," *Journal of Neuroscience*, vol. 21, no. 19, pp. 7733–7741, 2001.
[97] E. A. Berg, "A simple objective technique for measuring flexibility in thinking," *The Journal of general psychology*, vol. 39, no. 1, pp. 15–22, 1948.
[98] W. Zhang, L. Zhan, P. Thompson, and Y. Wang, "Deep representation learning for multimodal brain networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 613–624.
[99] S. Arslan, S. I. Ktena, B. Glocker, and D. Rueckert, "Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity," in *Graphs in biomedical image analysis and integrating medical imaging and non-imaging modalities*. Springer, 2018, pp. 3–13.
[100] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, "Explainability methods for graph convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 772–10 781.
[101] J. Rasero, N. Amoroso, M. La Rocca, S. Tangaro, R. Bellotti, S. Stramaglia, and A. D. N. Initiative, "Multivariate regression analysis of structural mri connectivity matrices in alzheimer's disease," *PLoS One*, vol. 12, no. 11, p. e0187281, 2017.
[102] M. Kutova, J. Mrzilkova, J. Riedlova, and P. Zach, "Asymmetric changes in limbic cortex and planum temporale in patients with alzheimer disease," *Current Alzheimer Research*, vol. 15, no. 14, pp. 1361–1368, 2018.
[103] L. V. Hiscox, C. L. Johnson, M. D. McGarry, H. Marshall, C. W. Ritchie, E. J. Van Beek, N. Roberts, and J. M. Starr, "Mechanical property alterations across the cerebral cortex due to alzheimer's disease," *Brain communications*, vol. 2, no. 1, p. fcz049, 2020.
[104] A. Hafkemeijer, C. Möller, E. G. Dopper, L. C. Jiskoot, T. M. Schouten, J. C. Van Swieten, W. M. Van der Flier, H. Vrenken, Y. A. Pijnenburg, F. Barkhof *et al.*, "Resting state functional connectivity differences between behavioral variant frontotemporal dementia and alzheimer's disease," *Frontiers in human neuroscience*, vol. 9, p. 474, 2015.
[105] C. J. Stoodley, E. M. Valera, and J. D. Schmahmann, "Functional topography of the cerebellum for motor and cognitive tasks: an fmri study," *Neuroimage*, vol. 59, no. 2, pp. 1560–1570, 2012.
[106] F. Van Overwalle, Q. Ma, and E. Heleven, "The posterior crus ii cerebellum is specialized for social mentalizing and emotional self-experiences: a meta-analysis," *Social Cognitive and Affective Neuroscience*, vol. 15, no. 9, pp. 905–928, 2020.
[107] R. P. Sawyer, F. Rodriguez-Porcel, M. Hagen, R. Shatz, and A. J. Espay, "Diagnosing the frontal variant of alzheimer's disease: a clinician's yellow brick road," *Journal of clinical movement disorders*, vol. 4, no. 1, pp. 1–9, 2017.
[108] R. Calati, J. J. Maller, C. Meslin, J. Lopez-Castroman, K. Ritchie, P. Courtet, and S. Artero, "Repatriation is associated with isthmus cingulate cortex reduction in community-dwelling elderly," *The World Journal of Biological Psychiatry*, vol. 19, no. 6, pp. 421–430, 2018.
[109] J. Hornung, E. Smith, J. Junger, K. Pauly, U. Habel, and B. Derntl, "Exploring sex differences in the neural correlates of self-and other-referential gender stereotyping," *Frontiers in behavioral neuroscience*, vol. 13, p. 31, 2019.
[110] T. Chen, Z. Li, J.-f. Cui, J. Huang, M. Irish, Y. Wang, and R. C. Chan, "The neural substrates of sex differences in balanced time perspective: A unique role for the precuneus," *Brain Imaging and Behavior*, pp. 1–9, 2022.
[111] B. Adinoff, M. J. Williams, S. E. Best, T. S. Harris, P. Chandler, and M. D. Devous Sr, "Sex differences in medial and lateral orbitofrontal cortex hypoperfusion in cocaine-dependent men and women," *Gender medicine*, vol. 3, no. 3, pp. 206–222, 2006.

Haoteng Tang received his B.E. degree in Biomedical Engineering from Xi'an Jiaotong University (XJTU) in 2016 and received his M.S. degree in Biomedical Engineering from University of Southern California (USC) in 2018. He is currently working toward the Ph.D. degree in the Electrical and Computer Engineering Department at the University of Pittsburgh. His research interests include graph mining, brain network representation learning and medical image analysis.

Dr. Guixiang Ma received her Ph.D. in Computer Science from University of Illinois at Chicago (UIC). She is currently an AI Research Scientist at Intel Labs. Her research interests include machine learning, data mining, graph representation learning and brain network analysis.



Lei Guo received her B.E. degree in Electrical Engineering from Xi'an Jiaotong University (XJTU), China and M.S. degree in Industrial and System Engineering from National University of Singapore, Singapore. She is now working toward the Ph.D. degree in the Electrical and Computer Engineering at the University of Pittsburgh. Her research interests include brain network mining and bioinformatics.



Xiyao Fu received his B.E. degree in Computer Science and Engineering from University of Electronic Science and Technology of China (UESTC) in 2017. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Pittsburgh. His research interests include medical image analysis and graph mining.

Prof. Heng Huang received the B.S. and M.S. degrees from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 1997 and 2001, respectively, and the Ph.D. degree in computer science from the Dartmouth College, Hanover, NH, USA, in 2006. He is currently the John A. Jurenko Endowed Professor of computer engineering with the Electrical and Computer Engineering Department, University of Pittsburgh, Pittsburgh, PA, USA. His research interests include machine learning, data mining, computer vision, pattern recognition, and biomedical data science.

Prof. Liang Zhan received his Ph.D. degree in Biomedical Engineering from University of California, Los Angeles (UCLA). He is currently an Associate Professor in the Departments of Electrical and Computer Engineering and Bioengineering at the University of Pittsburgh, PA, USA. His research interests include computational neuroimaging, brain connectomics, machine learning and bioinformatics.