Contrastive Brain Network Learning via Hierarchical Signed Graph Pooling Model

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

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. First, 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 the current graph learning models are interpretable, which may not be capable of providing biological insights for model outcomes. Here, we propose an interpretable hierarchical signed graph representation learning (HSGPL) model to extract graph-level representations from brain functional networks, which can be used for different prediction tasks. 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 data from human connectome project (HCP) and open access series of imaging studies (OASIS). Our results from extensive experiments demonstrate the superiority of the proposed model compared with several state-of-the-art techniques. In addition, we use graph saliency maps, derived from these prediction tasks, to demonstrate detection and interpretation of phenotypic biomarkers.

Index Terms—Brain functional networks, contrastive learning, data augmentation, hierarchical graph pooling (HGP), interpretability, signed graph learning.

I. INTRODUCTION

Understanding brain organizations and their relationship with phenotypes (e.g., clinical outcomes, behavioral, or demographical variables) are of prime importance in the modern neuroscience field. One of the important research directions is to use noninvasive neuroimaging data (e.g., functional magnetic resonance imaging or fMRI) to identify potential imaging biomarkers for clinical purposes. Most previous studies focus on voxelwise and region-of-interests (ROIs) imaging features [1], [2], [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], [5], [6], [7], [8], [9], [10], [11], [12], [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], [15], [16]. In addition, compared with the 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 3-D 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 predefined network features, such as clustering coefficient and small-worldness [18], [19], [20], [21], [22]. This may be suboptimal since these predefined 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 the biological characteristics of brain network–phenotype associations in recent years [23], [24], [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], [27], [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 the 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], [31], [32], [33]. To address this issue, many recent studies...
introduce hierarchical GNNs, including node embedding and hierarchical graph pooling (HGP) strategies, to embed the whole brain network in a hierarchical manner [30], [34], [35], [36], [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 the graph nodes are connected via nonnegative 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], [39], [40], [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 the brain functional networks, the positive edge means coactivation and the negative edge indicates antiactivation between those connected nodes. Meanwhile, the balance theory defines four 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 the 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 coembed 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 the global graph features. As suggested by literature [30], [42], [43], [44], the graph hierarchical structure can facilitate to yield whole graph representations and to enable the graph-level tasks (i.e., clinical disease classification based on the whole brain networks). Particularly, we propose a new hierarchical pooling module for signed graphs based on the information theory and extend the current methods on signed graph from local embedding to global embedding.

The second issue is that most of the current GNNs on brain network studies are not interpretable, and thus are incapable of providing 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], [46], [47], [48].

In our graph pooling module, we compute an information score (IS) to measure the information gain for each brain node and choose top-$K$ nodes with high information gains as information hubs. 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 the original data samples. For example, image rotation and cropping are common transformations to generate new samples in the image classification tasks [49], [50], [51], [52], [53]. In graph structural data, a few studies proposed to use 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], [55], [56], [57], [58]. These strategies, although boosting the model performance on large-scale benchmark datasets (e.g., CORA, CITESEER), may not be suitable to generate contrastive brain network samples. On one hand, each node in brain networks represents a defined brain region with specific brain activity information so that the brain node cannot be arbitrarily removed or added. On the other hand, add/drop operations on the 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 the fMRI blood-oxygen-level-dependent (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 follows.

1) We propose an HSGRL model to embed the brain functional networks and we apply the proposed model on multiple phenotype prediction tasks.
2) 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 the fMRI-derived brain network data.
3) The proposed HSGRL model is interpretable which yields heuristic biological explanations.
4) 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. GNNs and Brain Network Embedding

GNNs are generalized deep learning architectures which are broadly used 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., graph convolutional network (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], [43], [44], [67]), the whole graph representations are obtained by a naive global readout operation (e.g., sum or average all the 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 HGP strategies are proposed to learn representations for the whole graph in a hierarchical manner [30], [34], [35], [68], [69]. The traditional methods to extract brain network patterns are based on the graph theory [18], [19], [20], [21], [22], [70] or geometric network optimization [71], [72], [73], [74]. A few recent studies [24], [25], [75] introduce GNNs to discover brain patterns for phenotypes’ predictions. However, hierarchical structures in the brain networks are not considered in these models, which limit the model performance in a way. Recently, a few hierarchical brain network embedding models are proposed [36], [65], [76].

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 ode-level representation learning [39], [41], [77], [78]. In this work, we design a signed graph hierarchical pooling strategy to extract graph-level representations from the 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 (nonlinear) transformations [28], [36], [79]. 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 downsample graphs relying on network communities which are one of the important hierarchical structures that can be interpreted [36], [37], [80], [81]. For example, [36] proposed an HGP neural network relying on the 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 the 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 graph view augmentation [54], [82]. In this work, we introduce graph contrastive learning into the brain functional network analysis and generate contrastive samples from the fMRI BOLD signals.
Fig. 1. Diagram of the proposed contrastive graph learning framework (in the bottom black box) with the HSGRL model (in the top black box) for the functional brain network embedding and downstream tasks (i.e., phenotype classification or regression). The HSGRL model consists of the cascaded balanced and unbalanced embedding (BUE) and HGP modules to extract graph-level representations of contrastive brain functional network pairs (i.e., $\hat{X}_G$ and $\hat{X}_G$) in a hierarchical manner. $\hat{X}_G$ and $\hat{X}_G$ participate to build up the contrastive loss for graph contrastive learning. Meanwhile, a concatenate operation is used to generate the fused graph feature by $X_G = \left[ \hat{X}_G \ | \ \hat{X}_G \right]$. The fused graph feature $X_G$ is used for downstream prediction tasks (i.e., graph classification and regression).

B. HSGRL Model

We present our HSGRL model in Fig. 1. The HSGRL model includes the BUE module and HGP module.

1) BUE Module: The balance theory is broadly used to analyze the node relationships in the 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 a path between $v_i$ and $v_j$. Specifically, if the number of negative edges is even in the path between $v_i$ and $v_j$, then $v_j$ belongs to the balanced set of $v_i$. Otherwise, $v_j$ belongs to the unbalanced set of $v_i$. The balance theory indicates that the following.

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

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

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

where $F_{\text{sign}}(\cdot)$ 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

$$X = \left[ X^B \ | \ X^U \right]$$

where $\left[ \left| \right| \right]$ denotes the concatenate operation.

2) Hierarchical Signed Graph Pooling: As shown in Fig. 1, the proposed HGP module consists of four steps including: 1) ISs computation; 2) top-K informative hubs selection; 3) features' aggregation; and 4) graph pooling.

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

$$
\tilde{A}_+ = D_+^{-\frac{1}{2}} A_+ D_+^{-\frac{1}{2}} \\
\tilde{A}_- = D_-^{-\frac{1}{2}} A_- D_-^{-\frac{1}{2}}
$$

(5)

where $\tilde{A}$ is the normalized adjacency matrix. $D_+$ and $D_-$ are degree matrices of $A_+$ and $|A_-|$, respectively. Note that the $i$th line in $\tilde{A}$, denoted by $\tilde{A}_i$, 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 IS by

$$
\text{IS}_i^B = \|\tilde{A}_+^{T,i} \otimes X^B \|_L_1 + \|\tilde{A}_-^{T,i} \otimes X^U \|_L_1 \\
\text{IS}_i^U = \|\tilde{A}_+^{T,i} \otimes X^U \|_L_1 + \|\tilde{A}_-^{T,i} \otimes X^B \|_L_1
$$

(6)

where $\|\cdot\|_L_1$ is the lineewise $L_1$ norm, and $\otimes$ is the scalar multiplication between each line of two matrices. $^T$ represents the transpose of vector. Then the IS of $v_i$ can be obtained by

$$
\text{IS}_i = \text{IS}_i^B + \text{IS}_i^U.
$$

(7)

b) Top-K node selection and feature aggregation: After we obtain the IS 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 $\hat{x}_i \hat{x}_j^T$. We weighted add (i.e., set feature attentions as weights) the feature of each unselected node to one of the hub features, where the attention value between these two nodes is the biggest.

c) Graph pooling: After feature aggregation, we downscale the graph node by removing all the unselected nodes. In another word, only the selected top-$K$ network hubs and the edges among them will be preserved after graph pooling. Since the functional brain network is a fully connected graph, no isolated node existed in the down-scaled graph.

C. Contrastive Learning Framework With BUE and HGP

The contrastive learning framework with HSGRL is presented in Fig. 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 $\hat{X}$, 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

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

(8)

where $\hat{x}_i$ and $\hat{x}_i$ are the $i$th row of $\hat{X}$ and $\hat{X}$, respectively. $N'(\leq N)$ 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 [83], [84], [85] is used 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 $\hat{X}_G$). We use $m \in \{1, \ldots, M\}$ to denote the identity number (ID) of the sample pair. The contrastive loss of the $mth$ sample pair is formulated as

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

(9)

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

$$
\Phi(\hat{x}_G^m, \hat{x}_G^m) = \frac{\hat{x}_G^m \hat{x}_G^m^T}{\|\hat{x}_G^m\| \|\hat{x}_G^m\|}.
$$

(10)

The batch contrastive loss can be computed by

$$
\mathcal{L}_{\text{contrastive}} = \frac{1}{M} \sum_{m=1}^M \ell_m.
$$

(11)

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

$$
\mathcal{L} = \eta_1 \mathcal{L}_{\text{supervised}} + \eta_2 \mathcal{L}_{\text{contrastive}}
$$

(12)

where $\eta_1$ and $\eta_2$ 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) [86]. The second includes 1326 subjects (mean age = 70.42±8.95, 738 women) from the Open Access Series of Imaging Studies (OASIS) dataset [87]. Details of each dataset can be found on their official websites1,2 CONN [88] was used to preprocess fMRI data, and the preprocessing pipeline follows our previous publications [89], [90]. For HCP data, each subject’s network has a dimension of $82 \times 82$ based on $82$ ROIs defined using FreeSurfer (V6.0) [91]. For OASIS data, each subject’s network has a dimension of $132 \times 132$ based on the Harvard-Oxford Atlas and automated anatomical labeling (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.

1https://www.oasis-brains.org
2https://wiki.humanconnectome.org
Fig. 2. Visualization of the averaged adjacency matrices for the original and contrastive samples on (a) HCP dataset and (b) OASIS dataset. The averaged contrastive sample pair is generated using a window size \( d = 10 \).

B. Implementation Details

We randomly split the entire functional brain network dataset into five disjoint subsets for five-fold cross-validations in our experiments. The values in the adjacency matrices \((\hat{A} \text{ and } \tilde{A})\) of the brain functional networks are within the 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 [92] to optimize the loss functions in our model with a batch size of 128. The initial learning rate is \(1e^{-4}\) and decayed by \((1 - (\text{current_epoch}/\text{max_epoch}))^{0.9}\). 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] and [93], stop training if the validation loss does not decrease for 50 epochs. The experiments were deployed on one NVIDIA RTX A6000 graphics processing unit (GPU).

C. Similarities of Contrastive Samples

We use 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 \((1/M) \sum_{m=1}^{M} \Psi(\hat{A}^m, \hat{A}^m)\), and the interpair similarity is computed by \((1/M^2) \sum_{m=1}^{M} \sum_{t=1}^{M} \Psi(\hat{A}^m, \tilde{A}^t)\), 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 interpair \(L_1\) distances on HCP and OASIS data are 0.2925 and 0.3137, respectively. The interpair Cosine similarities on HCP and OASIS data are 0.7311 and 0.7014, respectively. We visualize the averaged adjacency matrices on HCP and OASIS data in Fig. 2(a) and (b), respectively, to show their similarities. The original sample is generated using the whole fMRI BOLD signal (i.e., \(d = 0\)).

D. Classification Tasks

1) Experiment Setup: For comparison, we adopted seven baseline models, which include two traditional graph embedding models (tensor-based brain network embedding (t-BNE) [73] and multimodal CCA+ joint ICA (mCCA-ICA) [74]), one basic GNN (i.e., GCN [26]), two deep graph representation learning models designed for brain network embedding (BrainChey [25] and BrainNet-convolutional neural networks (CNNs) [24]), and two hierarchical GNNs with graph pooling strategies (hierarchical graph representation learning with differentiable pooling (DIFFPOOL) [30] and self-attention graph pooling (SAGPOOL) [34]). As aforementioned, the existing GNN-based models cannot directly take signed graphs as the input, and we therefore compute the absolute values of graph adjacency matrices as the input for these baseline models, which is consistent with previous studies [36], [94]. 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., three 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 the cascaded BUE and HGP modules is set to three, 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 $\eta_1$ and $\eta_2$ in the range of $[0, 1]$, and determine the loss weights as $\eta_1 = 1$ and $\eta_2 = 0.1$. The temperature parameter in contrastive loss is set as 0.2. Details of the hyperparameters analysis are shown in Section V-F.

2) Results: The regression results are presented in Table II. It shows that our model achieves the best MAE values compared with all the baseline methods. Similar to the classification tasks, the deep GNNs are superior than the 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 contrastive learning can further boost the model performance.

F. Ablation Studies

In this section, we investigate the effect of four 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 ($\alpha$) within contrastive loss; 3) the number of the BUE and HGP modules used in the HSGRL model; and 4) loss weights $\eta_1$ and $\eta_2$. 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.
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 numbers of the BUE and HGP modules (Column 3). (a) Analysis on the classification tasks. (b) Analysis on the regression tasks.

The first column in Fig. 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 are 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 Tables 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 $\alpha$ from 0.1 to 1.0 with a step of 0.1. The second column in Fig. 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 the BUE and HGP modules as $[1, 2, 3, 4, 5]$, respectively, for our framework. The third column in Fig. 3 shows the analysis of this parameter. It manifests that the framework performance is consistent and steady when different numbers of the BUE and HGP modules are deployed. The best number of the modules for almost all the tasks is three, except for the regression tasks on Flanker and Aggressive. Finally, we present the loss weights analysis (see Fig. 4) on the three 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 IS 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).
Fig. 5. Brain saliency maps for the classification tasks. Here we identify: 1) top 15 regions associated with AD and NC from OASIS and 2) top 10 regions associated with each sex and each zygosity from HCP.

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

Therefore, they can provide hints for further clinical analyses on how this phenotype is associated with the brain functional network from the global view. We use the class activation mapping (CAM) approach [100], [101], [102] to generate
the brain network saliency map, which indicates the top brain regions associated with each prediction task. Figs. 5 and 6 illustrate brain saliency maps for the classification and regression tasks, respectively. For example, in the classification task [AD versus normal control (NC)], the saliency map for AD highlights multiple regions (such as planum polare, frontal operculum cortex, supracalcarine cortex) which are conventionally conceived as the biomarkers of AD in medical imaging analysis [103], [104], [105], [106]. 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 [107], [108], [109], in which patients with AD typically show problems. Another example is the classification of male versus female on HCP data. Females are more “emotional” or “sensitive,” suggested by regions such as isthmuscingulate and caudalanteriorcingulate, while males tend to be more competitive and dominant, manifested in regions such as lateraborbitofrontal and precuneus. These results are consistent with previous findings in the literature [110], [111], [112], [113]. The details of all the highlighted brain regions in each task are summarized in Table III for the OASIS dataset, and in Table IV(a) and (b) for the HCP dataset. These highlighted regions can help us locating the 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 HSGrL model for brain functional network mining. In addition, a new data augmentation strategy is designed to generate the contrastive samples for the brain functional network data. Our new framework is capable of generating more accurate representations for the brain functional networks 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.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
ACKNOWLEDGMENT

Data were provided in part by the Human Connectome Project, MHG-USC Consortium (Principal Investigators: Bruce R. Rosen, Arthur W. Toga, and Van Wedeen; U01MH093765) funded by the NIH Blueprint Initiative for Neuroscience Research Grant; in part by the National Institutes of Health under Grant P41EB015896; and in part by the Instrumentation under Grant S10RR023043; Grant 1S10RR023401, and Grant 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 et al., “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 Biol., vol. 20, no. 6, pp. 1042–1055, Nov. 2015.
[4] Y. Zhang, L. Zhan, W. Cai, P. Thompson, and H. Huang, “Integrating heterogeneous brain networks for predicting brain disease conditions,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: 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 Cognit. Sci., vol. 22, no. 4, pp. 350–364, Apr. 2018.
[6] N. Kraljević et al., “Behavioral, anatomical and heritable convergence of affect and cognition in superior frontal cortex,” Neuroimage, vol. 243, Nov. 2021, Art. no. 118561.
[7] S. L. Bressler and V. Menon, “Large-scale brain networks in cognition: Emerging methods and principles,” Trends Cogn. Sci., vol. 14, no. 6, pp. 277–290, Jun. 2010.
[8] H. Tang et al., “A hierarchical graph learning model for brain network regression analysis,” Frontiers Neurosci., vol. 16, pp. 1–12, Nov. 2022.
[9] H. Sun 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,” Annu. Rev. Clin. Psychol., vol. 10, p. 581, Mar. 2014.
[11] J. Yan et al., “Modeling spatio-temporal patterns of holistic functional brain networks via multi-head guided attention graph neural networks (multi-head GAGNNs),” Med. Image Anal., vol. 80, Aug. 2022, Art. no. 102518.
[12] P. L. Banique et al., “Brain network modularity predicts exercise-related executive function gains in older adults,” Frontiers Aging Neurosci., vol. 9, p. 426, Jan. 2018.
[13] U. Braun, S. F. Muldoon, and D. S. Bassett, “On human brain networks in health and disease,” eLS, pp. 1–9, Feb. 2015.
[14] M. P. van den Heuvel, R. S. Kahn, J. Gotlib, and O. Sporns, “High-cost, high-capacity backbone for global brain communication,” Proc. Nat. Acad. Sci. USA, vol. 109, no. 28, pp. 11372–11377, Jul. 2012.
[15] O. Sporns, “The human connectome: Origins and challenges,” Neuron, vol. 80, pp. 53–61, Oct. 2013.
[16] M. G. Mattar and D. S. Bassett, “Brain network architecture,” in Network Science in Cognitive Psychology. Oxfordshire, U.K.: Routledge, 2019, p. 30.
[17] Y. Zhang, L. Zhan, S. Wu, P. Thompson, and H. Huang, “Disentangled and proportional representation learning for multi-view brain connectomes,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2021, pp. 508–518.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
[98] 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,” J. Neurosci., vol. 21, no. 19, pp. 7733–7741, Oct. 2001.

[99] E. A. Berg, “A simple objective technique for measuring flexibility in thinking,” J. Gen. Psychol., vol. 39, no. 1, pp. 15–22, Jul. 1948.

[100] W. Zhang, L. Zhan, P. Thompson, and Y. Wang, “Deep representation learning for multimodal brain networks,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2020, pp. 613–624.

[101] S. Arslan, S. I. Ktena, B. Glocker, and D. Rueckert, “Graph saliency analysis and integrating medical imaging and non-imaging modalities,” Cham, Switzerland: Springer, 2018, pp. 3–13.

[102] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, “Explainability methods for graph convolutional neural networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 10772–10781.

[103] J. Rasero et al., “Multivariate regression analysis of structural MRI connectivity matrices in Alzheimer’s disease,” PLoS ONE, vol. 12, no. 11, 2017, Art. no. e0187281.

[104] M. Kutová, J. Mrzílková, J. Riedlová, and P. Zach, “Asymmetric changes in limbic cortex and planum temporale in patients with Alzheimer disease,” Current Alzheimer Res., vol. 15, no. 14, pp. 1361–1368, Nov. 2018.

[105] L. V. Hiscox et al., “Mechanical property alterations across the cerebral cortex due to Alzheimer’s disease,” Brain Commun., vol. 2, no. 1, 2020, Art. no. fc9249.

[106] A. Hafkemeijer et al., “Resting state functional connectivity differences between behavioral variant frontotemporal dementia and Alzheimer’s disease,” Frontiers Hum. Neurosci., vol. 9, p. 474, Sep. 2015.

[107] 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, Jan. 2012.

[108] 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 Cognit. Affect. Neurosci., vol. 15, no. 9, pp. 905–928, Nov. 2020.

[109] 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,” J. Clin. Movement Disorders, vol. 4, no. 1, pp. 1–9, 2017.

[110] R. Calati et al., “Repatriation is associated with isthmus cingulate cortex reorganization in community-dwelling elderly,” World J. Biol. Psychiatry, vol. 19, no. 6, pp. 421–430, 2018.

[111] J. Hornung, E. Smith, J. Junger, K. Pauly, U. Habel, and B. Demtl, “Exploring sex differences in the neural correlates of self-and other-referential gender stereotyping,” Frontiers Behav. Neurosci., vol. 13, p. 31, Feb. 2019.

[112] T. Chen et al., “The neural substrates of sex differences in balanced time perspective: A unique role for the precuneus,” Brain Imag. Behav., vol. 16, pp. 2239–2247, Jun. 2022.

[113] B. Adinoff, M. J. Williams, S. E. Best, T. S. Harris, P. Chandler, and M. D. Devous, “Sex differences in medial and lateral orbitofrontal cortex hyperfusion in cocaine-dependent men and women,” Gender Med., vol. 3, no. 3, pp. 206–222, Sep. 2006.

Haoteng Tang (Graduate Student Member, IEEE) received the B.E. degree in biomedical engineering from Xi’an Jiaotong University (XJTU), Xi’an, China, in 2016, and the M.S. degree in biomedical engineering from the University of Southern California (USC), Los Angeles, CA, USA, in 2018. He is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, University of Pittsburgh, Pittsburgh, PA, USA. His research interests include graph mining, brain network representation learning, and medical image analysis.

Lei Guo received the B.E. degree in electrical engineering from Xi’an Jiaotong University (XJTU), Xi’an, China, in 2003, and the M.S. degree in industrial and system engineering from the National University of Singapore, Singapore, in 2007. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Pittsburgh, Pittsburgh, PA, USA. Her research interests include brain network mining and bioinformatics.

Guixiang Ma (Member, IEEE) received the Ph.D. degree in computer science from the University of Illinois at Chicago (UIC), Chicago, IL, USA, in 2019. She is currently an artificial intelligence (AI) Research Scientist with the Intel Laboratory, Hillsboro, OR, USA. Her research interests include machine learning, data mining, graph representation learning, and their applications to various domains.

Liang Zhan received the Ph.D. degree in biomedical engineering from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 2011. He is currently an Associate Professor with the Departments of Electrical and Computer Engineering and Bioengineering, University of Pittsburgh, PA, USA. His research interests include computational neuroimaging, brain connectomics, machine learning, and bioinformatics.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.