Customized Relationship Graph Neural Network for Brain Disorder Identification

Zhengwang Xia[1], Huan Wang[2], Tao Zhou[1], Zhuqing Jiao[2], and Jianfeng Lu[1(✉)]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China lujf@njust.edu.cn
[2] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

Abstract. The connectivity structure of brain networks/graphs provides insights into the segregation and integration patterns among diverse brain regions. Numerous studies have demonstrated that specific brain disorders are associated with abnormal connectivity patterns within distinct regions. Consequently, several Graph Neural Network (GNN) models have been developed to automatically identify irregular integration patterns in brain graphs. However, the inputs for these GNN-based models, namely brain networks/graphs, are typically constructed using statistical-specific metrics and cannot be trained. This limitation might render them ineffective for downstream tasks, potentially leading to suboptimal outcomes. To address this issue, we propose a Customized Relationship Graph Neural Network (CRGNN) that can bridge the gap between the graph structure and the downstream task. The proposed method can dynamically learn the optimal brain networks/graphs for each task. Specifically, we design a block that contains multiple parameterized gates to preserve causal relationships among different brain regions. In addition, we devise a novel node aggregation rule and an appropriate constraint to improve the robustness of the model. The proposed method is evaluated on two publicly available datasets, demonstrating superior performance compared to existing methods. The implementation code is available at https://github.com/NJUSTxiazw/CRGNN.

Keywords: Graph Neural Network · Brain Disorder · Graph Structure Learning.

1 Introduction

The connectivity structure within functional brain networks provides profound insights into the patterns of segregation and integration across diverse brain regions [6]. These patterns, in turn, reflect the brain's efficiency and flexibility in processing information [13]. Numerous studies have demonstrated that abnormal integration patterns among brain regions are associated with neurological disorders [15]. For example, Aggarwal et al. [1] reported significant abnormalities in the interactions among the frontal, temporal, and occipital lobes in individuals

with autism spectrum disorder (ASD). Zhou et al. [19] also observed some impaired functional connections between the right frontal lobe and the left parietal lobe in patients with mild cognitive impairment (MCI). These findings imply that abnormal functional connectivity could be the underlying cause of various neurological disorders.

Recently, graph neural networks (GNNs) have shown significant advantages in analyzing graph-structured data [17] and have been successfully applied to various domains [16]. Within the domain of brain disorder identification, multiple GNNs have been developed to efficiently extract features from brain networks/graphs. For instance, Li et al. [9] developed a new model called BrainGNN to enhance the interpretability of GNN-based methods, which incorporates an innovative node pooling operation to identify the most discriminative subgraph structures within brain networks. Chen et al. [3] proposed a novel network named LSGNN to address the heterogeneity among brain networks, which integrates a trainable module to encode brain networks into multiple latent subspaces in a learnable manner. However, the inputs of these methods, such as brain networks/graphs, are typically constructed using specific statistical metrics (e.g., Pearson correlation) rather than being trainable. The separation of graph construction and feature extraction may lead to a discrepancy between the input graphs and the subsequent task, potentially impeding the achievement of optimal results.

To address this issue, we propose a Customized Relationship Graph Neural Network (CRGNN) that integrates graph structure learning and subsequent tasks within a unified framework. The main contributions of this paper are as follows: 1) We develop a novel CRGNN framework to address the discrepancy between graph structure and downstream tasks, which adaptively learns the most suitable brain networks/graphs for various downstream tasks in a customizable manner. 2) Our approach offers a significant advantage over traditional statistical-specific methods (e.g., Pearson correlation) by capturing nonlinear interactions among brain regions, which provides a more comprehensive understanding of brain dysfunction. 3) The proposed CRGNN presents a robust and flexible solution to the challenge of existing GNN-based methods for brain disorder identification.

2 Method

2.1 Overview

Fig. 1 provides a detailed schematic of the proposed CRGNN framework, which comprises two key components: the Customized Relationship Block (CRB) and the Relational Aggregation Block (RAB). The CRB is designed to generate a learnable matrix, denoted as $\boldsymbol{G} \in \mathbb{R}^{v \times v}$, which captures the causal relationships among various brain regions. It takes the fMRI signal $\boldsymbol{X} \in \mathbb{R}^{t \times v}$ directly as input. Here, t represents the length of the time series, while v corresponds to the number of brain regions defined by a specific brain atlas [12]. The CRB block consists of v gated structures, all of which are learnable parameters. Each gate

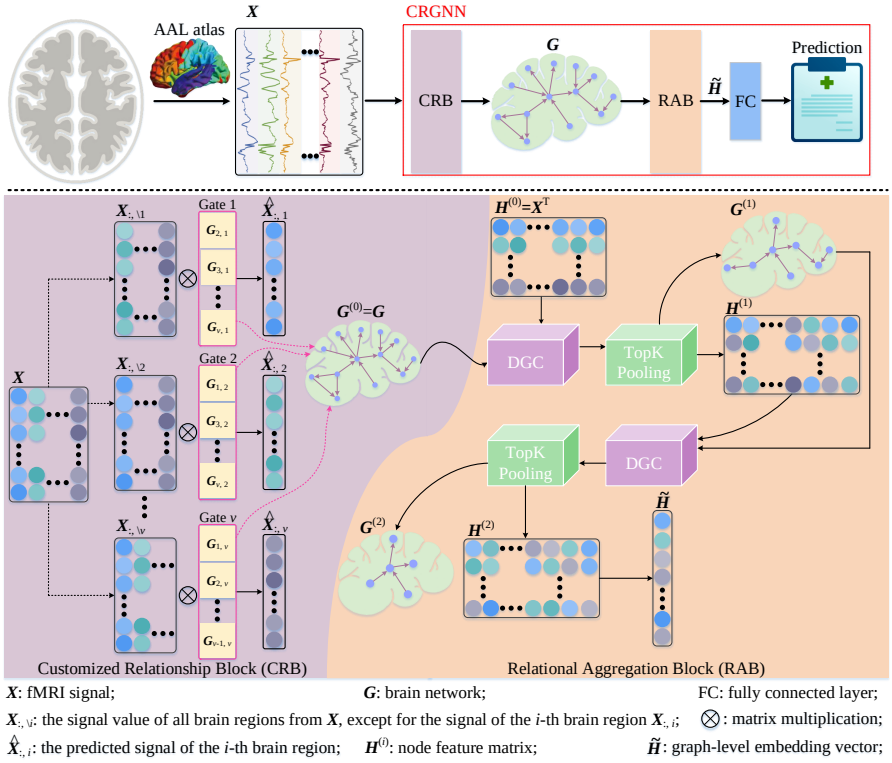

Fig. 1. An overview of the CRGNN framework, which comprises two main blocks: the CRB for graph structure learning and the RAB for comprehensive brain network representation.

preserves the causal effects of other brain regions on the specific brain region it corresponds to. Subsequently, these v gates are integrated to formulate the brain network/graph $\boldsymbol{G}$. Furthermore, the RAB is proposed to aggregate the node features into a joint latent space based on the learned graph $\boldsymbol{G}$, facilitating the generation of a graph-level representation. Ultimately, the model generates the predicted outcome $\hat{y}$ for each individual based on the graph-level representation.

2.2 Customized Relationship Block

Given the variability of factors contributing to different brain disorders, it is evident that the patterns of abnormalities in brain functional integration also vary across diseases. Traditional approaches that rely on predefined metrics such as Pearson's correlation coefficient for the construction of brain networks may not yield optimal results for every neurological disorder. The primary challenge stems

from the non-learnable nature of the brain connectivity structure, preventing simultaneous optimization with the subsequent feature extraction step.

To address this challenge, we propose a novel block referred to as CRB, which takes time series data as input and generates causal effects among all brain regions as output. The brain connectivity structure is incorporated as trainable parameters within the block, enabling it to be co-optimized with downstream tasks. This allows for the identification of the most suitable brain networks for specific neurological disorders, or customized graph learning that meets the specific requirements of different tasks. Specifically, we incorporated v gates into the CRB, where each gate serves to evaluate the causal effect of other brain regions on the corresponding brain region. In the domain of causal discovery, it is widely acknowledged that causes generate effects. However, to determine the potential causes of a variable, a limited search space is essential. It is commonly assumed that all other variables are potential causes of that variable [10]. Essentially, this means that the signals from each brain region can be reconstructed using signals from all other brain regions, which can be expressed as:

$$\hat{\boldsymbol{X}}_{:,\,i} = f_i(\boldsymbol{X}_{:,\setminus i}, \boldsymbol{N}_i) = tanh\left(\sum_{m=1, m\neq i}^{v} \boldsymbol{G}_{m,i}\boldsymbol{X}_{:,m} + \boldsymbol{N}_i\right), \tag{1}$$

where $\boldsymbol{X}_{:,\setminus i} \in \mathbb{R}^{t\times(v-1)}$ contains the signal values of all brain regions except for the i-th brain region $\boldsymbol{X}_{:,i}$. $f_i(\cdot)$ is associated with a parameter vector, i.e., $Gate_i = [\boldsymbol{G}_{1,i}, \boldsymbol{G}_{2,i}, \ldots, \boldsymbol{G}_{i-1,i}, \boldsymbol{G}_{i+1,i}, \ldots, \boldsymbol{G}_{v,i}]$, which is used to preserve the causal effects of other brain regions on the i-th brain region. $\boldsymbol{N}_i \in \mathbb{R}^t$ is the noise term, and $\hat{\boldsymbol{X}}_{:,\,i}$ represents the predicted signal value of the i-th brain region. The symbol $tanh(\cdot)$ is a nonlinear activation function. By concatenating the parameters of v gates and appending a diagonal vector with all elements equal to 0, a graph with a size of $v \times v$ can be obtained. This is the optimal brain network $\boldsymbol{G}$ that we have been looking for, i.e., $\boldsymbol{G} = [Gate_1, Gate_2, \ldots, Gate_v]$.

2.3 Relational Aggregation Block

With the derived brain network/graph $\boldsymbol{G}$, the next step is to obtain a comprehensive graph-level representation for the diagnosis of brain disorders. To accomplish this, we employ multiple Directed Graph Convolution (DGC) layers and graph pooling layers. The DGC layers are utilized to aggregate node features, while the graph pooling layers are employed to identify significant substructures within the graph. Notably, the brain network $\boldsymbol{G}$ we obtained is an asymmetric graph because it preserves the causal relationships among brain regions. Therefore, we have developed a DGC layer designed to aggregate node features from two perspectives. The propagation rule for the l-th DGC layer is formulated as follows:

$$\boldsymbol{H}^{(l)} = mean(\widetilde{\boldsymbol{G}^{(l-1)}}\boldsymbol{H}^{(l-1)}\theta_1^{(l)}, tran(\widetilde{\boldsymbol{G}^{(l-1)}})\boldsymbol{H}^{(l-1)}\theta_2^{(l)}), \tag{2}$$

where $\boldsymbol{G}^{(l)}$ is the updated graph after l steps of graph pooling layers. Here, $\widetilde{\boldsymbol{G}^{(l)}} = \boldsymbol{G}^{(l)} + \boldsymbol{I}_{v^{(l)}}$ is a matrix with added self-connections. $\boldsymbol{I}_{v^{(l)}}$ is an identity matrix, and $v^{(l)}$ equals to the number of nodes in the graph $\boldsymbol{G}^{(l)}$. $\boldsymbol{H}^{(l)} \in \mathbb{R}^{v^{(l)} \times d^{(l)}}$ represents the computed embedding following l steps of the DGC layer, and $d^{(l)}$ is the dimension of the embedding. $\theta_1^{(l)}$ and $\theta_2^{(l)} \in \mathbb{R}^{d^{(l-1)} \times d^{(l)}}$ are two trainable parameters of the same dimension. The function $tran(\cdot)$ represents the matrix transpose operation.

In addition to updating node features through the DGC layer, the graph pooling layer also plays a crucial role in improving the robustness of the GNN model. This layer is tasked with identifying critical substructures relevant to downstream tasks, facilitating the integration of essential information. In this study, we utilize the widely used TopK pooling strategy [2] to preserve essential substructures, which can be defined as follows:

$$sort\_index = TopK\left(\boldsymbol{H}^{(l)}; w^{(l)}\right), \tag{3}$$

where $w^{(l)} \in \mathbb{R}^{d^{(l)}}$ is a parameter vector and it is used to evaluate the importance of all the nodes within the graph $\boldsymbol{G}^{(l)}$. $TopK(\cdot)$ is a sorting function that returns the indices of nodes in graph $\boldsymbol{G}^{(l)}$ based on their importance. After calculating the importance indices of the nodes, the next step is to extract the structure of the critical subgraph and update the node embeddings of the identified key nodes. The updated graph and node embeddings will be used as input in the subsequent layers.

As illustrated in Fig. 1, the initial node features $\boldsymbol{H}^{(0)}$ will be updated to $\boldsymbol{H}^{(2)}$ by several iterative operations, including the DGC layer and the TopK pooling layer. Finally, the 2D matrix $\boldsymbol{H}^{(2)}$ is transformed into a graph-level representation, denoted as the vector $\widetilde{\boldsymbol{H}}$. The vector is subsequently fed into a fully connected layer to yield the predicted outcome $\hat{y}$.

2.4 Objective Function

In this study, our objective function comprises three components. Firstly, we employ the conventional cross-entropy loss to minimize the discrepancy between the ground-truth label y and the predicted outcome $\hat{y}$, which can be defined by

$$\mathcal{L}_{CE} = ylog(\hat{y}) + (1-y)log(1-\hat{y}), \tag{4}$$

Secondly, each gate in the CRB block will yield a prediction for its corresponding brain region. By combining the predictions of these v gates, we can obtain a 2D matrix that has the same size as the input $\boldsymbol{X}$. To minimize the error between the predicted result $\hat{\boldsymbol{X}}$ and the input $\boldsymbol{X}$, we employ the mean squared error (MSE) as the loss function, which is defined as:

$$\mathcal{L}_{MSE} = ||\boldsymbol{X} - \hat{\boldsymbol{X}}||_2^2. \tag{5}$$

Finally, to enhance the model's generalization ability and reduce over-fitting, we incorporate an L_1 regularization term into the computed brain network/graph,

denoted as $\mathcal{L}_G = ||\boldsymbol{G}||_1$. In summary, the total loss for the proposed model can be expressed by

$$\mathcal{L} = \mathcal{L}_{CE} + \alpha * \mathcal{L}_{MSE} + \beta * \mathcal{L}_G, \tag{6}$$

where α and β are two trade-off parameters.

3 Experiments

3.1 Dataset and Experimental Settings

We evaluated our framework on two publicly available datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI)[3] and the Autism Brain Imaging Data Exchange (ABIDE)[4]. The ADNI includes 170 normal controls (NC) and 283 individuals with mild cognitive impairment (MCI), while the ABIDE comprises 571 typically developing (TD) subjects and 531 individuals with Autism Spectrum Disorder (ASD). The resting-state functional magnetic resonance imaging (rs-fMRI) data is preprocessed using a standardized protocol that involves several steps: slice time correction, motion correction, spatial filtering, and covariates regression [4]. The time-series data for 90 brain regions were extracted using the Automated Anatomical Labeling (AAL) atlas [14], which is commonly employed in various frameworks to identify brain disorders.

We compare the proposed model with five GNN-based methods that are specifically designed for brain network analysis, including: 1) Hi-GCN [7], 2) BrainGNN, [9], 3) PSCR-GNN [18], 4) IBGNN [5], and 5) LSGNN [3]. All experiments are conducted using NVIDIA GeForce GTX 1080Ti GPUs. The CRGNN is implemented in the PyTorch framework [11] and trained with an Adam optimizer (with a learning rate of 0.0001, training epochs of 50, and a batch size of 8). We conduct a grid search to determine the optimal values for the hyper-parameters: $\alpha = 10^{-3}$ and $\beta = 10^{-2}$. For the impact of hyper-parameter configurations on the results, please refer to the Supplementary Material. For the TopK pooling layer, 1/3 of the original number of nodes is retained as significant nodes to be kept each time. All reported results are the average of 5 rounds of ten-fold cross-validation. Finally, four metrics are utilized in our experiments, namely accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1 score.

3.2 Result Analysis

Table 1 presents the classification results of all methods across two publicly available datasets. Based on the experimental results, we can make the following observations: First, IBGNN and LSGNN consistently outperform previously developed models in identifying brain disorders, such as Hi-GCN, BrainGNN, and PSCR-GNN. Their superior performance is primarily attributed to their ability to mitigate the effects of noise factors within predefined brain networks/graphs.

[3] http://adni.loni.usc.edu/
[4] http://fcon_1000.projects.nitrc.org/indi/abide/

Table 1. Classification results (mean±std) of all methods on two tasks (%).

Tasks	Metrics	Hi-GCN	BrainGNN	PSCR-GNN	IBGNN	LSGNN	CRGNN
NC vs. MCI	ACC	71.31±2.32	73.72±2.56	69.33±2.59	74.18±2.29	78.36±2.40	**82.78**±1.94
	SEN	73.94±6.38	77.29±4.18	74.71±6.79	78.42±4.76	83.47±3.39	**86.10**±2.98
	SPE	70.16±5.28	71.81±3.75	66.58±5.17	71.82±5.39	75.29±5.06	**80.92**±2.64
	F1	65.73±2.99	68.70±3.18	64.49±2.11	69.39±2.54	74.20±3.20	**78.90**±2.09
TD vs. ASD	ACC	62.43±1.37	65.25±1.08	58.26±2.17	67.06±1.23	68.33±0.83	**72.23**±1.01
	SEN	58.16±3.01	62.07±3.76	55.41±2.15	63.67±1.77	65.53±2.09	**69.68**±2.37
	SPE	67.20±4.82	69.32±5.97	61.78±5.47	70.95±3.81	71.48±2.49	**75.35**±3.79
	F1	61.49±2.36	64.83±0.88	57.83±2.78	66.63±1.37	68.10±1.83	**72.15**±1.26

As a result, they achieve a more accurate graph-level representation of the brain network, leading to a significant improvement in the recognition performance of subsequent tasks. Second, our proposed approach exhibits the best performance compared to other GNN-based methods. This is mainly due to the flexibility of our approach, which allows for the customization of the brain network/graph to accommodate various downstream tasks. Unlike other methods that rely on predefined statistical metrics for brain network/graph construction, our approach dynamically learns the optimal graph structure tailored to the subsequent task. Moreover, our approach is statistically superior to other GNN-based methods (with $p < 0.05$) according to pairwise t-test, thereby providing robust evidence of its effectiveness.

3.3 Ablation Study

We conduct ablation studies to evaluate the effectiveness of the novel components, including: 1) the customized graph within the CRB module, 2) the rule for aggregating node information in the RAB module, and 3) the regularized loss $\mathcal{L}_G$ applied to the learned graph structure. In the CRB module, we adopt a widely recognized approach of replacing the brain networks/graphs learned with those derived from the Pearson correlation coefficient. In the RAB module, we adopt a commonly used node information aggregation rule, specifically the Graph Convolution Network (GCN) [8], to update node features, instead of the DGC layer. For the loss function, we conduct comparative experiments on whether to impose regularization loss on the underlying graph structure. The experimental results of the ablation study are listed in Table 2. The research findings reveal that both the proposed customized graph learning and the innovative node feature updating rules play a crucial role in enhancing the recognition performance for brain disorders. Furthermore, the ablation experiment conducted on the loss term $\mathcal{L}_G$ reveals that incorporating this term not only enhances the robustness of the model but also mitigates the risk of overfitting.

Table 2. Ablation study of CRGNN with different components on two tasks.

CRB		RAB	Loss	NC vs. MCI		TD vs. ASD	
Learnable	Pearson	DGC	GCN $\mathcal{L}_G$ Without	ACC	F1	ACC	F1
✓		✓	✓	**82.78**±1.94	**78.90**±2.09	**72.23**±1.01	**72.15**±1.26
✓		✓	✓	80.78±2.72	75.94±3.67	70.78±1.26	68.62±1.88
✓		✓	✓	79.90±2.14	75.26±2.81	70.60±1.41	68.80±1.86
✓		✓	✓	76.58±2.77	69.75±3.90	68.05±1.59	65.78±1.99
✓		✓	✓	72.39±2.79	63.93±3.62	64.79±1.48	62.37±1.90

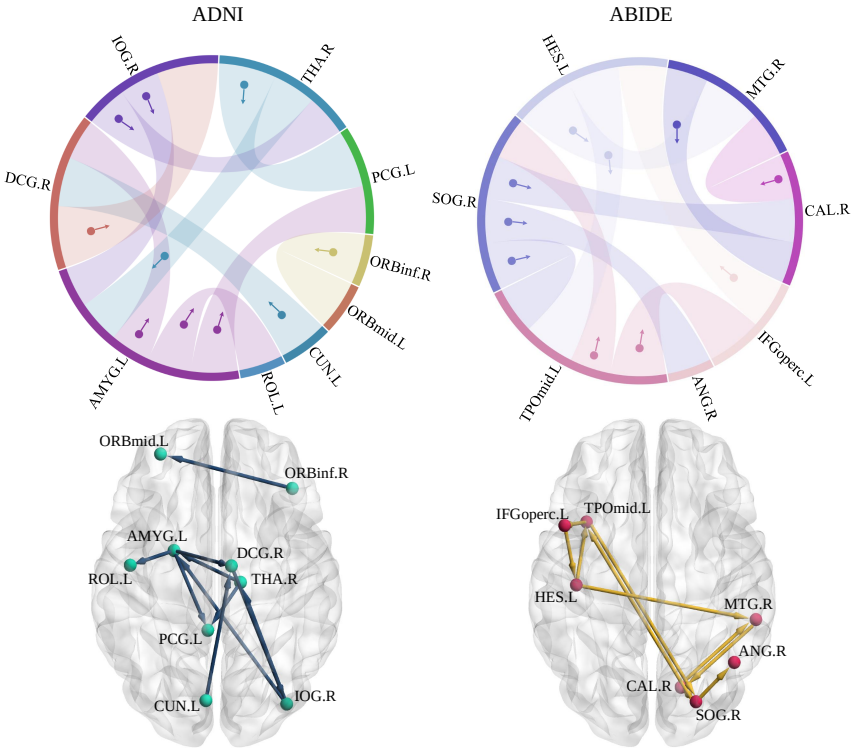

Fig. 2. The top 10 most discriminative connectivities identified by our method for both tasks. The motion direction of the ball in each arc represents the causal relationship between two brain regions (from cause to effect).

3.4 Discriminative Connectivity

Fig. 2 shows the top 10 discriminative connectivities retained by CRGNN in both tasks. Chen et al. found that changes in the middle frontal gyrus and amygdala were strongly associated with the worsening of MCI [4]. Aggarwal et al. demonstrated a robust relationship between changes in the temporal and

occipital lobes with ASD [1]. These findings are consistent with the conclusions drawn in this paper, further validating the reliability of our study.

4 Conclusion

In this paper, we present a novel GNN model for brain disorder identification that integrates graph structure learning and downstream tasks within a unified framework. We have designed a novel block, called CRB, to infer the causal relationships among brain regions. The block consists of multiple gates as learnable parameters that can be jointly optimized with the downstream tasks. We evaluate our method on two public datasets and demonstrate its effectiveness.

Acknowledgments. This work was supported in part by the Natural Science Foundation of Jiangsu Province (No. BK20221487), the National Natural Science Foundation of China (No. 62172228) and the Jiangsu Provincial Key Research and Development Program (BE2021636). This work was also sponsored by Qing Lan Project of Jiangsu Province.

Disclosure of Interests. The authors declare that they have no known financial interests that could potentially influence the findings reported in this paper.

References

1. Aggarwal, P., Gupta, A.: Multivariate graph learning for detecting aberrant connectivity of dynamic brain networks in autism. Medical image analysis **56**, 11–25 (2019)
2. Bacciu, D., Di Sotto, L.: A non-negative factorization approach to node pooling in graph convolutional neural networks. In: AI* IA 2019–Advances in Artificial Intelligence: XVIIIth International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, November 19–22, 2019, Proceedings 18. pp. 294–306. Springer (2019)
3. Chen, D., Liu, M., Shen, Z., Zhao, X., Wang, Q., Zhang, L.: Learnable subdivision graph neural network for functional brain network analysis and interpretable cognitive disorder diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–66. Springer (2023)
4. Chen, X., Zhang, H., Zhang, L., Shen, C., Lee, S.W., Shen, D.: Extraction of dynamic functional connectivity from brain grey matter and white matter for mci classification. Human brain mapping **38**(10), 5019–5034 (2017)
5. Cui, H., Dai, W., Zhu, Y., Li, X., He, L., Yang, C.: Interpretable graph neural networks for connectome-based brain disorder analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 375–385. Springer (2022)
6. Deco, G., Tononi, G., Boly, M., Kringelbach, M.L.: Rethinking segregation and integration: contributions of whole-brain modelling. Nature Reviews Neuroscience **16**(7), 430–439 (2015)
7. Jiang, H., Cao, P., Xu, M., Yang, J., Zaiane, O.: Hi-gcn: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. Computers in Biology and Medicine **127**, 104096 (2020)

8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017)
9. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
10. Mclntosh, A., Gonzalez-Lima, F.: Structural equation modeling and its application to network analysis in functional brain imaging. Human brain mapping **2**(1-2), 2–22 (1994)
11. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
12. Revell, A.Y., Silva, A.B., Arnold, T.C., Stein, J.M., Das, S.R., Shinohara, R.T., Bassett, D.S., Litt, B., Davis, K.A.: A framework for brain atlases: Lessons from seizure dynamics. Neuroimage **254**, 118986 (2022)
13. Tian, Y., Sun, P.: Percolation may explain efficiency, robustness, and economy of the brain. Network Neuroscience **6**(3), 765–790 (2022)
14. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage **15**(1), 273–289 (2002)
15. Vértes, P.E., Bullmore, E.T.: Annual research review: growth connectomics–the organization and reorganization of brain networks during normal and abnormal development. Journal of Child Psychology and Psychiatry **56**(3), 299–320 (2015)
16. Waikhom, L., Patgiri, R.: A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. Artificial Intelligence Review **56**(7), 6295–6364 (2023)
17. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019)
18. Yang, C., Wang, P., Tan, J., Liu, Q., Li, X.: Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks. Computers in Biology and Medicine **139**, 104963 (2021)
19. Zhou, T., Thung, K.H., Liu, M., Shi, F., Zhang, C., Shen, D.: Multi-modal latent space inducing ensemble svm classifier for early dementia diagnosis with neuroimaging data. Medical image analysis **60**, 101630 (2020)