Interpretable Graph Neural Networks for Connectome-Based Brain Disorder Analysis

Hejie Cui[1], Wei Dai[1], Yanqiao Zhu[2], Xiaoxiao Li[3], Lifang He[4], and Carl Yang[1](✉)

[1] Emory University [2] University of California, Los Angeles
[3] The University of British Columbia [4] Lehigh University
j.carlyang@emory.edu

Abstract. Human brains lie at the core of complex neurobiological systems, where the neurons, circuits, and subsystems interact in enigmatic ways. Understanding the structural and functional mechanisms of the brain has long been an intriguing pursuit for neuroscience research and clinical disorder therapy. Mapping the connections of the human brain as a network is one of the most pervasive paradigms in neuroscience. Graph Neural Networks (GNNs) have recently emerged as a potential method for modeling complex network data. Deep models, on the other hand, have low interpretability, which prevents their usage in decision-critical contexts like healthcare. To bridge this gap, we propose an interpretable framework to analyze disorder-specific Regions of Interest (ROIs) and prominent connections. The proposed framework consists of two modules: a brain-network-oriented backbone model for disease prediction and a globally shared explanation generator that highlights disorder-specific biomarkers including salient ROIs and important connections. We conduct experiments on three real-world datasets of brain disorders. The results verify that our framework can obtain outstanding performance and also identify meaningful biomarkers. All code for this work is available at https://github.com/HennyJie/IBGNN.git.

Keywords: Interpretation · Graph neural network · Brain networks

1 Introduction

Brain networks (a.k.a the connectome) are complex graphs with anatomic regions represented as nodes and connectivities between the regions as links. Interpretable models on brain networks for disorder analysis are vital for understanding the biological functions of neural systems, which can facilitate early diagnosis of neurological disorders and neuroscience research [27]. Previous work on brain networks has studied models from shallow to deep, such as graph kernels [14], tensor factorizations [22], and convolutional neural networks [16, 17, 20].

Recently, Graph Neural Networks (GNNs) attract broad interest due to their established power for analyzing graph-structured data [19, 34]. Compared with shallow models, GNNs are suitable for brain network analysis with universal

expressiveness to capture the sophisticated connectome structures [4, 26, 38, 43]. However, GNNs as a family of deep models are prone to overfitting and lack transparency in predictions, preventing their usage in decision-critical areas like disorder analysis. Although several methods have been proposed for GNN explanation [24, 36, 39], most of them focus on node-level prediction tasks and will produce a unique explanation for each subject when applied to graph-level tasks. However, for graph-level connectome-based disorder analysis, it is recognized that subjects having the same disorder share similar brain network patterns [15], which means disorder-specific explanations across instances are preferable. Moreover, brain networks have unique properties such that directly applying vanilla GNN models will obtain suboptimal performance.

In this work, we propose an interpretable GNN framework to investigate disease-specific patterns that are common across the group and robust to individual image quality. Meanwhile, the group-level interpretation can be combined with subject-specific brain networks for different levels of interpretation. As shown in Fig. 1, it is composed of two modules: a backbone model IBGNN which adapts a message passing GNN designed for connectome-based disease prediction and an explanation generator that learns a globally shared mask to highlight disorder-specific biomarkers including salient Regions of Interest (ROIs) and important connections. Furthermore, we combine the two modules by enhancing the original brain networks with the learned explanation mask and further tune the backbone model. The resulting model, which we term IBGNN+ for brevity, produces predictions and interpretations simultaneously.

Through experiments on three real-world brain disorder datasets (i.e. HIV, BP, and PPMI), we show our backbone model performs well across brain networks constructed from different neuroimaging modalities. Also, it is demonstrated that the explanation generator can reveal disorder-specific biomarkers coinciding with neuroscience findings. Last, we show that the combination of explanation generator and backbone model can further boost disorder prediction performance.

2 The Proposed Model

Problem definition. The input to the proposed framework is a set of N weighted brain networks. For each network $G = (V, E, \boldsymbol{W})$, $V = \{v_i\}_{i=1}^{M}$ is the node set of size M defined by the Regions Of Interest (ROIs) on a specific brain parcellation [10, 32], with each v_i initialized with the node feature $\boldsymbol{x}_i$, $E = V \times V$ is the edge set of brain connectome, and $\boldsymbol{W} \in \mathbb{R}^{M \times M}$ is the weighted adjacency matrix describing the connection strengths between ROIs. The model outputs a brain disorder prediction $\hat{y}_n$ for each subject n and learns a disorder-specific interpretation matrix $\boldsymbol{M} \in \mathbb{R}^{M \times M}$ that is shared across all subjects to highlight the disorder-specific biomarkers.

The backbone model IBGNN. Edge weights in brain networks are often determined by the signal correlation between brain areas, which may have both positive and negative values, and thus cannot be handled correctly by conventional GNNs. To

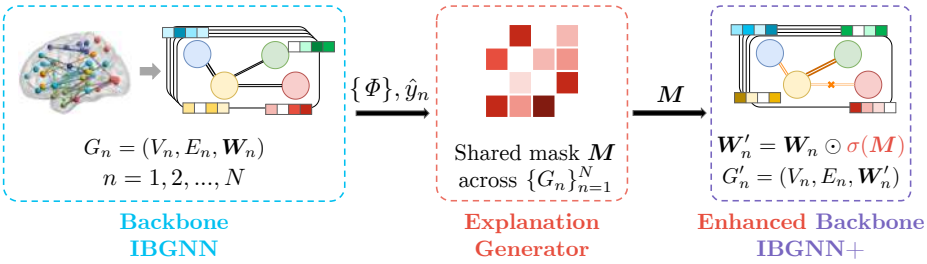


Fig. 1: An overview of our proposed framework. The backbone model is firstly trained on the original data. Then, the explanation generator learns a globally shared mask across subjects. Finally, we enhance the backbone by applying the learned explanation mask and fine-tune the whole model.

avoid this issue and better utilize edge weights in the GNN model, we design an edge-weight-aware message passing mechanism specifically for brain networks. Specifically, we first construct a message vector $\boldsymbol{m}_{ij} \in \mathbb{R}^D$ by concatenating embeddings of a node v_i and its neighbor v_j, and the edge weight w_{ij}:

$$\boldsymbol{m}_{ij}^{(l)} = \text{MLP}_1\left(\left[\boldsymbol{h}_i^{(l)}; \boldsymbol{h}_j^{(l)}; w_{ij}\right]\right), \tag{1}$$

where l is the index of the GNN layer. Then, for each node v_i, we aggregate messages from all its neighbors $\mathcal{N}_i$ with the following propagation rule:

$$\boldsymbol{h}_i^{(l)} = \xi\left(\sum\nolimits_{v_j \in \mathcal{N}_i \cup \{v_i\}} \boldsymbol{m}_{ij}^{(l-1)}\right), \tag{2}$$

where ξ is a non-linear activation function such as ReLU, and $\boldsymbol{h}_i^{(0)}$ is initialized with node feature $\boldsymbol{x}_i$ reflecting the connectivity information in brain networks [5]. After stacking L layers, a readout function summarizing all node embeddings is employed to obtain a graph-level embedding $\boldsymbol{g}$. Formally, we instantiate this function with another Multi-Layer Perceptron (MLP) and residual connections:

$$\boldsymbol{z} = \sum\nolimits_{i \in V} \boldsymbol{h}_i^{(L)}, \qquad \boldsymbol{g} = \text{MLP}_2(\boldsymbol{z}) + \boldsymbol{z}. \tag{3}$$

This backbone model IBGNN can be trained with the conventional supervised cross-entropy objective towards ground-truth disorder prediction, defined as

$$\mathcal{L}_{\text{CLF}} = -\frac{1}{N}\sum\nolimits_{n=1}^{N}\left(y_n \log\left(\hat{y}_n\right) + \left(1 - y_n\right)\log\left(1 - \hat{y}_n\right)\right). \tag{4}$$

The globally shared explanation generator. A general paradigm to generate explanations for GNNs is to find an explanation graph G' that has the maximum agreement with the label distribution on the original graph $G = (V, E, \boldsymbol{W})$, where G' can be a subgraph [39] or other variations of G [24, 40]. However, these explanation methods for GNNs mostly work on node-level prediction tasks

and will produce a unique explanation graph for each subject when applied to graph-level tasks. On the other hand, directly using attention weights in some attention-based GNN models [34, 41] as explanations is known to be problematic [1, 13]. Note that brain networks have some unique properties. For example, the node number and order are fixed under a given atlas. Also, brain networks assume that subjects with the same brain disorder have similar brain connection patterns. Therefore, a globally shared explanation graph G' capture common patterns for specific disorders at the group level is preferable.

In this work, we propose to learn a globally shared edge mask $\boldsymbol{M} \in \mathbb{R}^{M \times M}$ that is applied to all brain network subjects in a dataset. Specifically, we maximize the agreement between the predictions $\hat{y}$ on the original graph G and $\hat{y}'$ on an explanation graph $G' = (V, E, \boldsymbol{W}')$ induced by a masking matrix $\boldsymbol{M}$, where $\boldsymbol{W}' = \boldsymbol{W} \odot \sigma(\boldsymbol{M})$, $\odot$ denotes element-wise multiplication, and σ denotes the sigmoid function. Formally this objective is implemented as a cross-entropy loss:

$$\mathcal{L}_{\text{MASK}} = -\frac{1}{N} \sum\nolimits_{n=1}^{N} \sum\nolimits_{c=1}^{C} \mathbb{1}[\hat{y}_n = c] \log P_{\Phi}\left(\hat{y}'_n = \hat{y}_n \mid G'_n\right), \tag{5}$$

where $\sum_{n=1}^{N} P_{\Phi}\left(\hat{y}'_n = \hat{y}_n \mid G'_n\right)$ represents the conditional probability that the backbone model Φ's prediction $\hat{y}'_n$ on the masked graph G'_n is consistent with the prediction $\hat{y}_n$ on the original graph G_n, C is the number of possible prediction labels. Besides, following the practice in GNNExplainer [39], we further apply two regularization terms $\mathcal{L}_{\text{SPS}}$ and $\mathcal{L}_{\text{ENT}}$ to encourage the compactness of the explanation and the discreteness of the mask values, respectively:

$$\mathcal{L}_{\text{SPS}} = \sum\nolimits_{i,j} \boldsymbol{M}_{i,j}, \qquad \mathcal{L}_{\text{ENT}} = -(\boldsymbol{M} \log(\boldsymbol{M}) + (1 - \boldsymbol{M}) \log(1 - \boldsymbol{M})). \tag{6}$$

The final training objective is given as:

$$\mathcal{L} = \mathcal{L}_{\text{CLF}} + \alpha \mathcal{L}_{\text{MASK}} + \beta \mathcal{L}_{\text{SPS}} + \gamma \mathcal{L}_{\text{ENT}}, \tag{7}$$

where α, β and γ scale the numerical value of each loss item to the same order of magnitude to balance their influence. Our explanation generator will generate a globally shared edge mask that can be used for all testing graphs to investigate neurological biomarkers and highlight disorder-specific salient connections.

Enhancing the backbone with the learned explanations. The learned explanation mask can further improve the disorder prediction considering that raw brain networked data inevitably contain random noise. Specifically, we enhance the original backbone by applying essential disorder-specific signals. We note that this strategy is compatible with any backbone model, not limited to our proposed IBGNN. We combined the aforementioned two modules so that predictions and interpretations are produced in a closed-loop for brain disorder analysis. We term the enhanced model by IBGNN+ hereafter.

The whole training pipeline is summarized in Fig. 1. The original brain networks are firstly input to train the backbone model. Then, a globally shared explanation mask is learned based on the backbone model Φ and prediction $\hat{y}_n$. Finally, we enhance the backbone model by highlighting salient ROIs and important connections on the raw data and tune the backbone model again.

3 Experiments

Dataset acquisition and preprocessing. We evaluate our framework using three real-world neuroimaging datasets of different modalities. Specifically, groups in each dataset have balanced age and gender portions and are collected with the same image acquisition procedure.

- *Human Immunodeficiency Virus Infection (HIV):* This dataset is collected from Early HIV Infection Study at Northwestern University. It includes fMRI imaging of 70 subjects, 35 of which are early HIV patients, and the others are seronegative controls. We perform image preprocessing using the DPARSF[1] toolbox. The images are realigned to the first volume, followed by slice timing correction, normalization, spatial smoothness using an 8-mm Gaussian kernel, band-pass filtering (0.01-0.08 Hz), and linear trend removing of the time series. We focus on the 116 anatomical regions of interest (ROI), and extract a sequence of responses from them. Finally, brain networks with 90 cerebral regions are constructed, where each node represents a brain region and links are created based on correlations between different brain regions.
- *Bipolar Disorder (BP):* This DTI imaging dataset is collected from 52 bipolar I subjects and 45 healthy controls. We use the FSL toolbox[2] for preprocessing which includes distortion correction, noise filtering, and repetitive sampling from the distributions of principal diffusion directions for each voxel. Each subject is parcellated into 82 regions based on FreeSurfer-generated cortical/subcortical gray matter regions.
- *Parkinson's Progression Markers Initiative (PPMI):* This large-scale, publicly available dataset[3] is from a collaborative study[4] to improve PD therapeutics. We consider brain imaging in the DTI modality of 754 subjects, 596 of whom are Parkinson's disorder patients, and the rest 158 are healthy controls. The raw data are aligned using the FSL eddy-correct tool to correct head motion and eddy current distortions. The brain extraction tool (BET) from FSL is used to remove non-brain tissue. The skull-stripped images are linearly aligned and registered using Advanced Normalization Tools (ANTs[5]). 84 ROIs are parcellated from T1-weighted structural MRI using FreeSurfer[6] and the brain network connectivity is reconstructed using the deterministic 2nd-order Runge-Kutta (RK2) whole-brain tractography algorithm [42].

Experimental settings. The proposed model is implemented using PyTorch 1.10.2 [29] and PyTorch Geometric 2.0.3 [9]. A Quadro RTX 8000 GPU with 48GB of memory is used for our model training. Hyper-parameters are selected automatically with the open source AutoML toolkit NNI[7]. We refer readers of

[1] http://rfmri.org/DPARSF/
[2] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
[3] https://www.ppmi-info.org/
[4] https://www.michaeljfox.org/
[5] http://stnava.github.io/ANTs/
[6] https://surfer.nmr.mgh.harvard.edu/
[7] https://github.com/microsoft/nni

Table 1: Experimental results (%) on three datasets, where * denotes a significant improvement according to paired t-test with $p = 0.05$ compared with baselines. The best performances are in bold and the second runners are underlined.

Method	HIV			BP			PPMI		
	Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC
M2E	$57.14_{\pm 19.17}$	$53.71_{\pm 19.80}$	$57.50_{\pm 18.71}$	$52.56_{\pm 13.86}$	$51.65_{\pm 13.38}$	$52.42_{\pm 13.83}$	$78.69_{\pm 1.78}$	$45.81_{\pm 4.17}$	$50.39_{\pm 2.59}$
MIC	$54.29_{\pm 18.95}$	$53.63_{\pm 19.44}$	$55.42_{\pm 19.10}$	$62.67_{\pm 20.92}$	$63.00_{\pm 21.61}$	$61.79_{\pm 21.74}$	$79.11_{\pm 2.16}$	$49.65_{\pm 5.10}$	$52.39_{\pm 2.94}$
MPCA	$67.14_{\pm 20.25}$	$64.28_{\pm 23.47}$	$69.17_{\pm 20.17}$	$52.56_{\pm 13.12}$	$50.43_{\pm 14.99}$	$52.42_{\pm 13.69}$	$79.15_{\pm 0.57}$	$44.18_{\pm 0.18}$	$50.00_{\pm 0.00}$
MK-SVM	$65.71_{\pm 7.00}$	$62.08_{\pm 7.49}$	$65.83_{\pm 7.41}$	$57.00_{\pm 8.89}$	$41.08_{\pm 13.44}$	$53.75_{\pm 8.00}$	$79.15_{\pm 0.57}$	$44.18_{\pm 0.18}$	$50.00_{\pm 0.00}$
GCN	$70.00_{\pm 12.51}$	$68.35_{\pm 13.28}$	$73.58_{\pm 9.49}$	$55.56_{\pm 13.86}$	$50.71_{\pm 11.75}$	$61.55_{\pm 28.77}$	$78.55_{\pm 1.58}$	$47.87_{\pm 4.40}$	$59.43_{\pm 8.64}$
GAT	$71.43_{\pm 11.66}$	$69.79_{\pm 10.83}$	$77.17_{\pm 9.42}$	$63.34_{\pm 9.15}$	$60.42_{\pm 7.56}$	$67.07_{\pm 5.98}$	$79.02_{\pm 1.25}$	$45.85_{\pm 3.16}$	$64.40_{\pm 6.87}$
PNA	$57.14_{\pm 12.78}$	$45.09_{\pm 19.62}$	$57.14_{\pm 12.78}$	$63.71_{\pm 11.34}$	$55.54_{\pm 14.06}$	$60.30_{\pm 11.89}$	$79.36_{\pm 1.84}$	$51.76_{\pm 10.32}$	$54.71_{\pm 6.77}$
BrainNetCNN	$69.24_{\pm 19.04}$	$67.08_{\pm 11.11}$	$72.09_{\pm 19.01}$	$65.83_{\pm 20.64}$	$64.74_{\pm 17.42}$	$64.32_{\pm 13.72}$	$55.20_{\pm 12.63}$	$55.45_{\pm 9.15}$	$52.54_{\pm 10.21}$
BrainGNN	$74.29_{\pm 12.10}$	$73.49_{\pm 10.75}$	$75.00_{\pm 10.56}$	$68.00_{\pm 12.45}$	$62.33_{\pm 13.01}$	$74.20_{\pm 12.93}$	$69.17_{\pm 0.00}$	$44.19_{\pm 0.00}$	$45.26_{\pm 3.65}$
IBGNN	$82.14_{\pm 10.81}$*	$82.02_{\pm 10.86}$*	$86.86_{\pm 11.65}$*	$73.19_{\pm 12.20}$	$72.87_{\pm 12.09}$*	$83.64_{\pm 9.61}$*	**$79.82_{\pm 1.47}$**	$51.58_{\pm 4.66}$	$70.65_{\pm 6.55}$*
IBGNN+	**$84.29_{\pm 12.94}$***	**$83.86_{\pm 13.42}$***	**$88.57_{\pm 10.89}$***	**$76.33_{\pm 13.00}$***	**$76.13_{\pm 13.01}$***	**$84.61_{\pm 9.08}$***	$79.55_{\pm 1.67}$	**$56.58_{\pm 7.43}$**	**$72.76_{\pm 6.73}$***

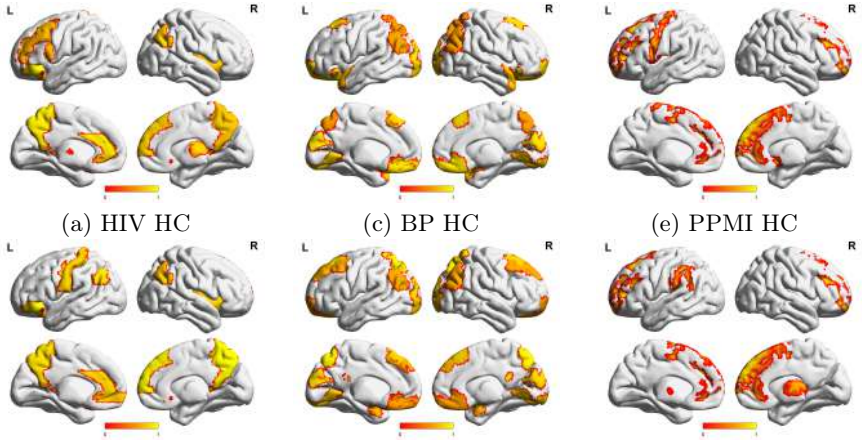
interest to supplementary materials for implementation details. All reported results are averaged of ten-fold cross validation.

Baselines. We compare our proposed models, i.e., the backbone model IBGNN and the explanation enhanced IBGNN+, with competitors of both shallow and deep models. Shallow methods include M2E [22], MIC [31], MPCA [23], and MK-SVM [7], where the output graph-level embeddings are evaluated using logistic regression classifiers. We also include three representative deep graph models: GAT [35], GCN [19], PNA [3] and two state-of-the-art deep models specifically design for brain networks: BrainNetCNN [17] and BrainGNN [20].

Prediction performance. The overall results are presented in Table 1. Both our proposed models yield impressive improvements over SOTA shallow and deep baselines. Compared with shallow models such as MK-SVM, our backbone model IBGNN outperforms them by large margins, with up to 11% absolute improvements on BP. Besides, the effectiveness of our brain network-oriented design is supported by its superiority compared with other SOTA deep models. Moreover, the performance of the explanation enhanced model IBGNN+ can further increase the backbone by about 9.7% relative improvements, which demonstrates that IBGNN+ effectively highlights the disorder-specific signals while also achieving the benefit of restraining random noises in individual graphs.

4 Interpretation Analysis

Neural system mapping. The ROIs on brain networks can be partitioned into neural systems based on their structural and functional roles under a specific parcellation atlas, which facilitates the understanding of generated explanations from a neuroscience perspective. In this paper, we map the ROI nodes as defined on each dataset into eight commonly used neural systems, including Visual Network (VN), Auditory Network (AN), Bilateral Limbic Network (BLN), Default Mode



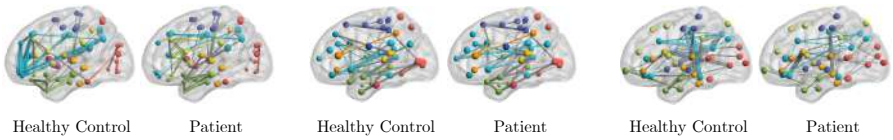
(a) HIV HC (c) BP HC (e) PPMI HC

(b) HIV Patient (d) BP Patient (f) PPMI Patient

Fig. 2: Visualization of salient ROIs on the explanation enhanced brain connection networks for Health Control (HC) and Patient. The color of regions represents ROI's average importance in the given group. The bright-yellow color indicates a high score, while dark-red indicates a low score.

Network (DMN), Somato-Motor Network (SMN), Subcortical Network (SN), Memory Network (MN), and Cognitive Control Network (CCN).

Salient ROIs. We provide both group-level and individual-level interpretations to understand which ROIs contribute most to the prediction of a specific disorder. On the group level, we rank the most salient ROIs on the learned explanation mask by calculating the sum of the edge weights connected to each node. Then on the individual level, we use the BrainNet Viewer [37] to plot the salient ROIs on the average brain connectivity graph enhanced by the learned explanation mask. For the HIV disease, anterior cingulate, paracingulate gyri, and inferior frontal gyrus are selected as salient ROIs. This complies with scientific findings that the regional homogeneity value of the anterior cingulate and paracingulate gyri are decreased [25] and lower gray matter volumes are found in inferior frontal gyrus in HIV patients [21]. The individual-level visualizations in Fig. 2(a)(b) show the difference between Health Control (HC) and HIV patients in those salient ROIs. For the BP disease, secondary visual cortex and medial to superior temporal gyrus are selected as salient ROIs. This observation is in line with existing studies that visual processing abnormalities have been characterized in bipolar disorder patients [28, 30], which is also confirmed in Fig. 2(c)(d). For the PPMI disease, rostral middle frontal gyrus and superior frontal gyrus are selected as salient ROIs and Fig. 2(e)(f) display the difference. This is in accordance with MRI analysis revealing a significant decrease in PD patients in the rostral medial frontal gyrus and superior, middle, and inferior frontal gyri [18]. All these observed salient ROIs can be potential biomarkers to identify brain disorders from each cohort.



(a) HIV (b) BP (c) PPMI

Fig. 3: Visualization of important connections on the explanation enhanced brain connection network. Edges connecting nodes within the same neural system (VN, AN, BLN, DMN, SMN, SN, MN, CCN) are colored accordingly, while edges across different systems are colored gray. Edge width indicates its weight in the explanation graph.

Important connections. The globally shared explanation mask $\boldsymbol{M}$ provides interpretations of important connections. We obtain an explanation subgraph G'_s by taking the top 100 weighted edges from the masked G' with all other edges removed. The connection comparisons are shown in Fig. 3, which helps identify connections related to specific disorders. For the HIV dataset, the explanation subgraph of patients excludes rich interactions within the DMN (colored blue) system. Also, interactions within the VN (colored red) system of patients are significantly less than those of HCs. These patterns are consistent with the findings in earlier studies [11, 12] that connectivity alterations within- and between-network DMN and VN may relate to known visual processing difficulties for HIV patients. For the BP dataset, compared with tight interactions within the BLN (colored green) system of the healthy control, the connections within BLN system of the patient subject are much sparser, which may signal pathological changes in this neural system. This observation is in line with previous studies [6], which finds that the parietal lobe, one of the major lobes in the brain roughly located at the upper back area in the skull and is in charge of processing sensory information received from the outside world, is mainly related to Bipolar disorder attack. Since parietal lobe ROIs are contained in BLN under our parcellation, the connections missing within the BLN system in our visualization are consistent with existing clinical understanding. For the PPMI dataset, the connectivity in the patient group decreases in the SMN (colored purple) system, which integrates primary sensorimotor, premotor, and supplementary motor areas to facilitate voluntary movements. This observation confirms existing neuroimaging studies that have repeatedly shown disorder-related alteration in sensorimotor areas of Parkinson's patients [2]. Furthermore, individuals with PD have lower connectivity within the DMN (colored blue) system compared with healthy controls, which is consistent with the cognition recession study on Parkinson's patients [8, 33].

5 Conclusion

In this work, we propose a novel interpretable GNN framework for connectome-based brain disorder analysis, which consists of a brain network-oriented GNN

predictor and a globally shared explanation generator. Experiments on real-world neuroimaging datasets show the superior prediction performance of both our backbone and the explanation enhanced models and validate the disorder-specific interpretations from the generated explanation mask. The limitation of the proposed framework might arise from the small size of neuroimaging datasets, which restraints the effectiveness and generalization ability of deep learning models. A direct future direction based on this work is to utilize pre-training and transfer learning techniques to learn across datasets. This allows for the sharing of information and explanations across different cohorts, which could lead to a better understanding of cross-disorder commonalities.

Acknowledgement

This research was partly supported by the internal funds and GPU servers provided by the Computer Science Department of Emory University and the University Research Committee of Emory University. Xiaoxiao Li was supported by NSERC Discovery Grant (DGECR-2022-00430). Lifang He was supported by ONR N00014-18-1-2009 and Lehigh's accelerator grant S00010293.

References

1. Bai, B., et al.: Why attentions may not be interpretable? In: SIGKDD (2021)
2. Caspers, J., et al.: Within-and across-network alterations of the sensorimotor network in parkinson's disease. Neuroradiology (2021)
3. Corso, G., et al.: Principal neighbourhood aggregation for graph nets. In: NeurIPS (2020)
4. Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A.A.C., Lukemire, J., Zhan, L., He, L., Guo, Y., Yang, C.: Braingb: A benchmark for brain network analysis with graph neural networks. arXiv preprint arXiv:2204.07054 (2022)
5. Cui, H., Lu, Z., Li, P., Yang, C.: On positional and structural node features for graph neural networks on non-attributed graphs. arXiv preprint arXiv:2107.01495 (2021)
6. Das, T.K., et al.: Parietal lobe and disorganisation syndrome in schizophrenia and psychotic bipolar disorder: A bimodal connectivity study. Psychiatry Res. Neuroimaging (2020)
7. Dyrba, M., et al.: Multimodal analysis of functional and structural disconnection in alzheimer's disease using multiple kernel svm. Hum. Brain Mapp. (2015)
8. van Eimeren, T., et al.: Dysfunction of the default mode network in parkinson disease: a functional magnetic resonance imaging study. Arch. Neurol. (2009)
9. Fey, M., et al.: Fast graph representation learning with pytorch geometric. In: RLGM@ICLR (2019)
10. Figley, T.D., et al.: Probabilistic white matter atlases of human auditory, basal ganglia, language, precuneus, sensorimotor, visual and visuospatial networks. Front. Hum. Neurosci. (2017)

11. Flannery, J.S., et al.: Hiv infection is linked with reduced error-related default mode network suppression and poorer medication management abilities. medRxiv.org (2021)
12. Herting, M.M., et al.: Default mode connectivity in youth with perinatally acquired hiv. Medicine (2015)
13. Jain, S., et al.: Attention is not explanation. In: NAACL-HLT (2019)
14. Jie, B., et al.: Sub-network based kernels for brain network classification. In: ACM BCB (2016)
15. Kan, X., Cui, H., Lukemire, J., Guo, Y., Yang, C.: Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In: MIDL (2022)
16. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. arXiv preprint (2022)
17. Kawahara, J., et al.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage (2017)
18. Kendi, A.K., et al.: Altered diffusion in the frontal lobe in parkinson disease. AJNR Am. J. Neuroradiol. (2008)
19. Kipf, T.N., et al.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)
20. Li, X., et al.: Braingnn: Interpretable brain graph neural network for fmri analysis. Med. Image Anal. (2021)
21. Li, Y., et al.: Structural gray matter change early in male patients with hiv. Int. J. Clin. Exp. Med. (2014)
22. Liu, Y., et al.: Multi-view multi-graph embedding for brain network clustering analysis. In: AAAI (2018)
23. Lu, H., et al.: Mpca: Multilinear principal component analysis of tensor objects. IEEE Trans. Neural Netw. (2008)
24. Luo, D., et al.: Parameterized explainer for graph neural network. In: NeurIPS (2020)
25. Ma, Q., et al.: Hiv-associated structural and functional brain alterations in homosexual males. Front. Neurol. (2021)
26. Maron, H., et al.: Invariant and equivariant graph networks. In: ICLR (2018)
27. Martensson, G., et al.: Stability of graph theoretical measures in structural brain networks in alzheimer's disease. Sci. Rep. (2018)
28. O'Bryan, R.A., et al.: Disturbances of visual motion perception in bipolar disorder. Bipolar Disord. (2014)
29. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: NeurIPS (2019)
30. Reavis, E.A., et al.: Structural and functional connectivity of visual cortex in schizophrenia and bipolar disorder: a graph-theoretic analysis. Schizophr. Bull. Open (2020)
31. Shao, W., et al.: Clustering on multi-source incomplete data via tensor modeling and factorization. In: PAKDD (2015)
32. Shirer, W.R., et al.: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex (2012)

33. Tessitore, A., et al.: Default-mode network connectivity in cognitively unimpaired patients with parkinson disease. Neurology (2012)
34. Veličković, P., et al.: Graph attention networks. In: ICLR (2018)
35. Veličković, P., et al.: Deep graph infomax. In: ICLR (2019)
36. Vu, M.N., et al.: Pgm-explainer: Probabilistic graphical model explanations for graph neural networks. In: NeurIPS (2020)
37. Xia, M., et al.: Brainnet viewer: a network visualization tool for human brain connectomics. PloS One (2013)
38. Yang, Y., Zhu, Y., Cui, H., Kan, X., He, L., Guo, Y., Yang, C.: Data-efficient brain connectome analysis via multi-task meta-learning. In: KDD (2022)
39. Ying, Z., et al.: Gnnexplainer: Generating explanations for graph neural networks. In: NeurIPS (2019)
40. Yuan, H., et al.: Explainability in graph neural networks: A taxonomic survey. arXiv.org (2020)
41. Yun, S., et al.: Graph transformer networks. In: NeurIPS (2019)
42. Zhan, L., et al.: Comparison of nine tractography algorithms for detecting abnormal structural brain networks in alzheimer's disease. Front. Aging Neurosci. (2015)
43. Zhu, Y., Cui, H., He, L., Sun, L., Yang, C.: Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In: EMBC (2022)