Biologically Plausible Brain Graph Transformer

Ciyuan Peng[1], Yuelong Huang[2], Qichao Dong[3], Shuo Yu[2], Feng Xia[4], Chengqi Zhang[5], Yaochu Jin[6]

[1]Federation University Australia, [2]Dalian University of Technology, [3]Zhejiang Gongshang University, [4]RMIT University, [5]Hong Kong Polytechnic University, [6]Westlake University

BioBGT GitHub

Introduction

Motivation

- Brain graphs exhibit small-world architecture accompanied by the presence of (1) **hubs** and (2) **functional modules**.
- Current methods lack tailored design for small-world brain graphs, leading to **lacking sufficient biological plausibility** for the learned brain graph representations.

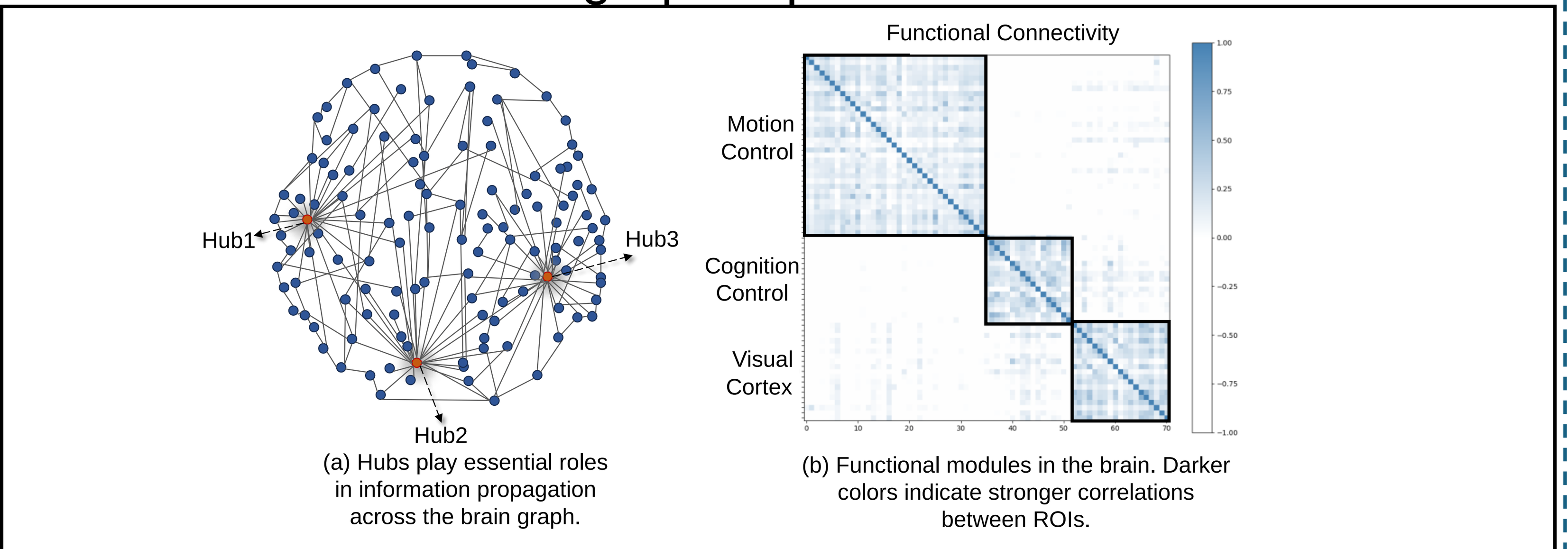

Fig: Small-world architecture of brain graphs

Proposal and Contribution

BioBGT encodes the small-world architecture of brain graphs to enhance the biological plausibility of the learned representations:

- **Node importance encoding:** Capturing node importance in the information propagation across brain graphs.
- **Functional module encoding:** Preserving the functional segregation and integration characteristics of brain graphs in the learned representations.

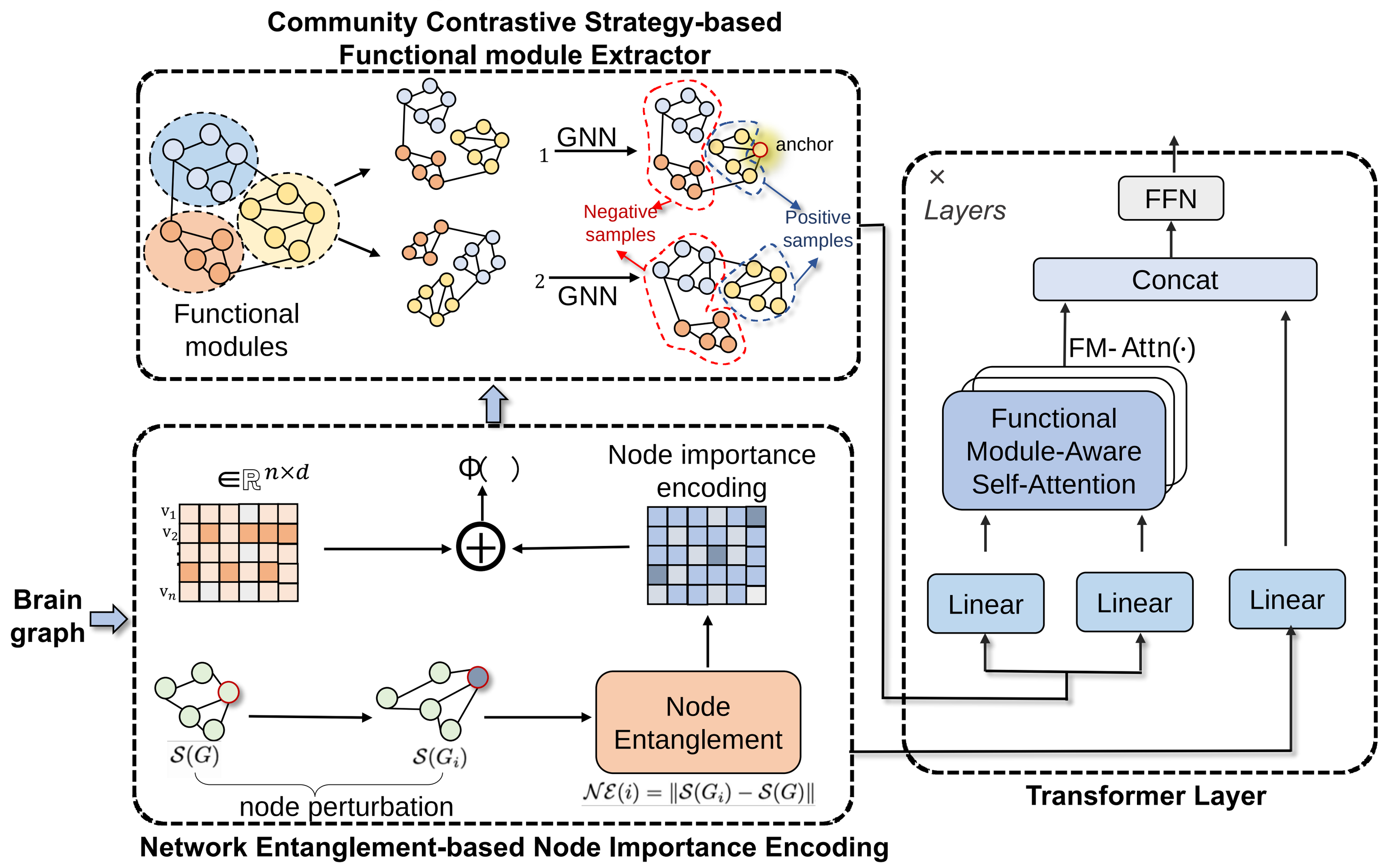

Fig: Overall framework of BioBGT.

Method

1 Network Entanglement (NE)-based Node Importance Encoding

- **Node importance degree calculation (NE value):** $\mathcal{NE}(i) = \|\mathcal{S}(G_i) - \mathcal{S}(G)\|$
  - $\mathcal{S}(G)$ and $\mathcal{S}(G_i)$ are the density matrix-based spectral entropy of G and G_i, G_i is the i-control graph obtained after the perturbation of node i.
- **Node representation with node importance encoded:**

$$\mathbf{x}'_i = \Phi(\mathbf{x}_i) = \mathbf{x}_i + \mathbf{x}_{\mathcal{NE}(i)} \qquad (1)$$

  - $\mathbf{x}_i$ is the original representation of node i.
  - $\mathbf{x}_{\mathcal{NE}(i)}$ is the learnable embedding vector specified by $\mathcal{NE}(i)$.
  - $\Phi(\cdot)$ denotes our node importance encoding.

2 Functional Module Encoding

- **Community contrastive strategy-based functional module extractor:** We design a functional module extractor $\psi(\cdot)$ to obtain the updated functional module-aware representation:

$$\mathbf{h}_i := \psi(i, \mathcal{M}_i) \qquad (2)$$

  - $\mathbf{h}_i$ is the functional module-aware representation of node i.
  - $\mathcal{M}_i$ is the functional module node i belongs to. Functional modules are detected by using Louvain algorithm.
  - We adopt the InfoNCE as the contrastive loss function to update representations.
- **Functional Module-aware Self-attention:** We define trainable exponential kernels on functional module-aware node representations:

$$\text{FM-}Attn(i) = \sum_{j \in V} \frac{exp\left(\langle \mathbf{W}_Q\mathbf{h}_i, \mathbf{W}_K\mathbf{h}_j\rangle/\sqrt{d_K}\right)}{\sum_{u\in V} exp\left(\langle \mathbf{W}_Q\mathbf{h}_i, \mathbf{W}_K\mathbf{h}_u\rangle/\sqrt{d_K}\right)} f(\mathbf{h}_j). \qquad (3)$$

Therefore, for node i, the node representation fed to FFN is: $\tilde{\mathbf{x}}_i = \Phi(\mathbf{x}_i) + \text{FM-}Attn(i)$.

Results and Analysis

Model Comparison

- We compare BioBGT with other state-of-the-art baselines on graph classification tasks across three datasets.

Dataset	Group	Number	No. ROI
ADNI	NC	190	90
	MCI	170	90
	AD	47	90
ADHD-200	NC	230	190
	ADHD	229	190
ABIDE	NC	493	200
	Autism	516	200

Table: Statistic information of three datasets.

Method		ADHD-200		ABIDE		ADNI	
		ACC	AUC	ACC	AUC	ACC	AUC
ML Methods	SVM	53.56±2.73	54.66±3.40	49.01±1.70	49.05±1.94	32.29±2.63	49.88±3.10
	Random Forest	58.96±2.77	59.49±2.38	51.14±3.08	51.41±3.23	49.03±1.27	58.18±2.31
Graph Transformer Models	SAN	51.09±2.00	51.22±2.21	49.80±1.97	50.20±2.34	34.44±4.61	49.23±2.67
	Graph Trans.	50.76±2.07	51.49±1.15	50.20±0.50	48.20±0.16	40.28±4.17	52.31±2.04
	Graphormer	61.60±0.90	58.64 ±1.50	58.40±0.68	57.61±0.72	35.64±2.17	48.19±12.69
	SAT-PE	60.00±2.73	59.68±2.60	60.60±3.11	59.14±4.56	39.96±1.51	48.17±6.57
	SAT+PE	64.44±3.45	64.21±3.40	58.76±4.88	69.29±5.48	41.51±4.01	42.13±5.74
	BRAINNETTF	70.80±2.70	**79.36±3.43**	68.24±2.24	**78.38±3.43**	47.39±3.11	55.72±7.13
	Polynormer	64.78±2.34	63.61±2.43	57.03±0.96	56.42±1.56	41.85±2.12	54.34±4.37
	Gradformer	68.94±3.18	67.83±4.66	61.56±4.13	61.75±4.29	46.54±2.72	53.88±2.37
	GTSP	61.70±3.81	61.41±2.90	61.37±3.59	60.43±3.47	47.27±3.81	53.59±3.26
Graph Neural Networks	GAT	55.38±3.18	54.97±3.28	53.51±2.54	53.41±2.48	34.99±7.43	51.73±6.66
	BrainGNN	55.76±1.20	58.00±0.49	51.34±1.17	54.27±0.66	43.33±4.08	50.21±2.97
	BrainGB	68.20±7.81	74.64±10.10	65.12±3.90	70.32±3.66	44.34±3.90	62.24±4.68
	MCST-GCN	59.06±2.69	59.05±3.89	54.22±2.40	55.18±2.35	48.44±3.12	62.25±2.93
	GroupBNA	69.87±3.02	71.16±4.53	63.14±2.65	71.30±3.81	46.72±1.33	50.85±8.10
Our Model	**BioBGT**	**71.06±0.08**	71.64±1.14	**74.00±2.01**	73.33±2.37	**52.08±2.08**	**62.33±5.98**

Table: Model comparison Results.

Results and Analysis (cont.)

Ablation Studies

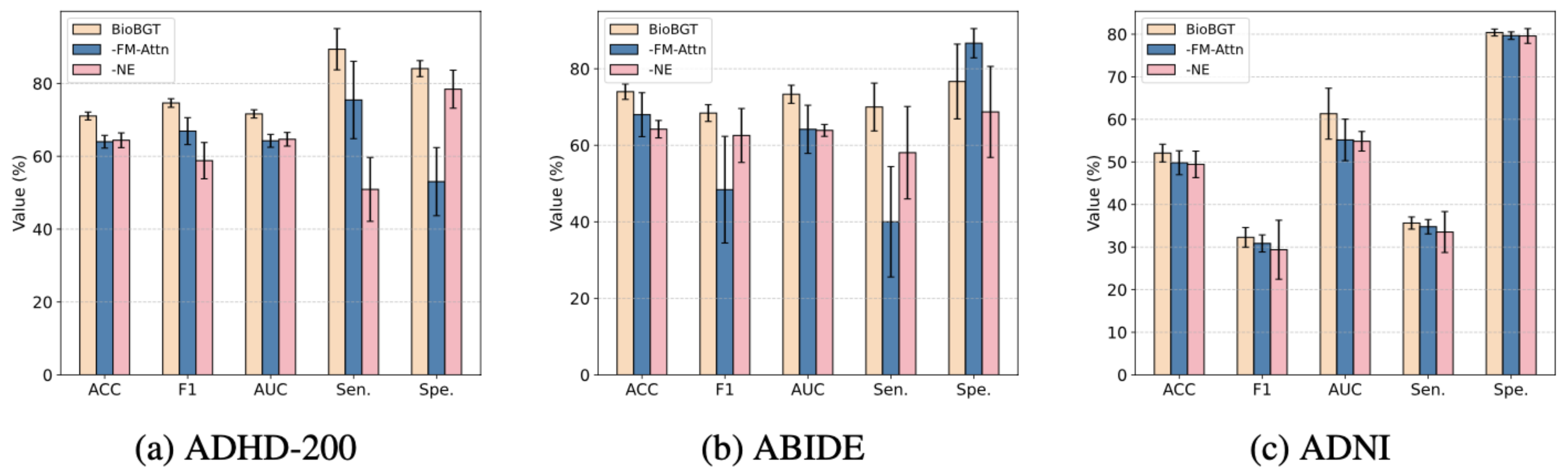

Fig: Performance of BioBGT and its altered models.

Comparative analysis of node importance measurement

	ABIDE			ADNI			ADHD-200		
	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC
+PE	54.00±2.97	60.60±2.25	60.91±2.05	30.09±3.36	49.43±2.42	52.14±2.41	69.21±7.14	67.56±3.24	67.39±2.80
+DC	59.73±4.23	61.20±1.88	61.28±1.80	27.27±1.57	50.57±1.64	55.45±3.59	74.22±1.35	69.78±2.33	70.18±2.28
+PE+DC	56.64±2.40	63.00±1.63	63.32±1.55	27.77±1.08	50.94±1.25	55.08±3.94	73.60±4.28	70.67±4.00	70.79±4.10
+BC	52.77±1.30	70.00±6.12	65.62±7.29	26.70±4.26	48.11±3.20	51.38±7.88	72.20±3.61	71.12±0.88	70.09±1,05
+CC	62.43±1.53	73.75±6.50	70.84±7.65	25.84±5.50	48.11±3.61	50.68±11.24	73.78±5.11	**72.69±0.88**	69.38±0.94
+EC	53.13±1.62	71.25±8.20	66.67±9.88	27.75±9.12	47.64±3.08	54.39±8.41	73.09±4.98	71.11±1.11	68.99±0.99
BioBGT	**68.41±2.19**	**74.00±2.01**	**73.33±2.37**	**32.29±2.31**	**52.08±2.08**	**61.33±5.98**	**74.63±1.18**	71.06±0.08	**71.64±1.14**

Table: Results of BioBGT and its variants

Biological Plausibility in Node Importance Encoding

- We compare the NE values of each node with their node efficiency (NEff) values to evaluate the validity of NE.
- The changing trend of the NE curve aligns closely with that of the NEff curve.

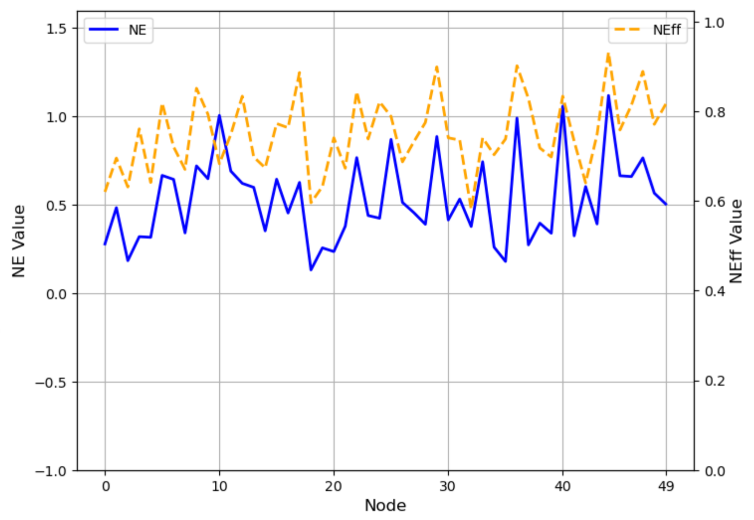

Fig: The NE and NEff values of 50 randomly selected nodes from a sample in the ABIDE dataset.

Biological Plausibility in Functional Module-Aware Self-Attention

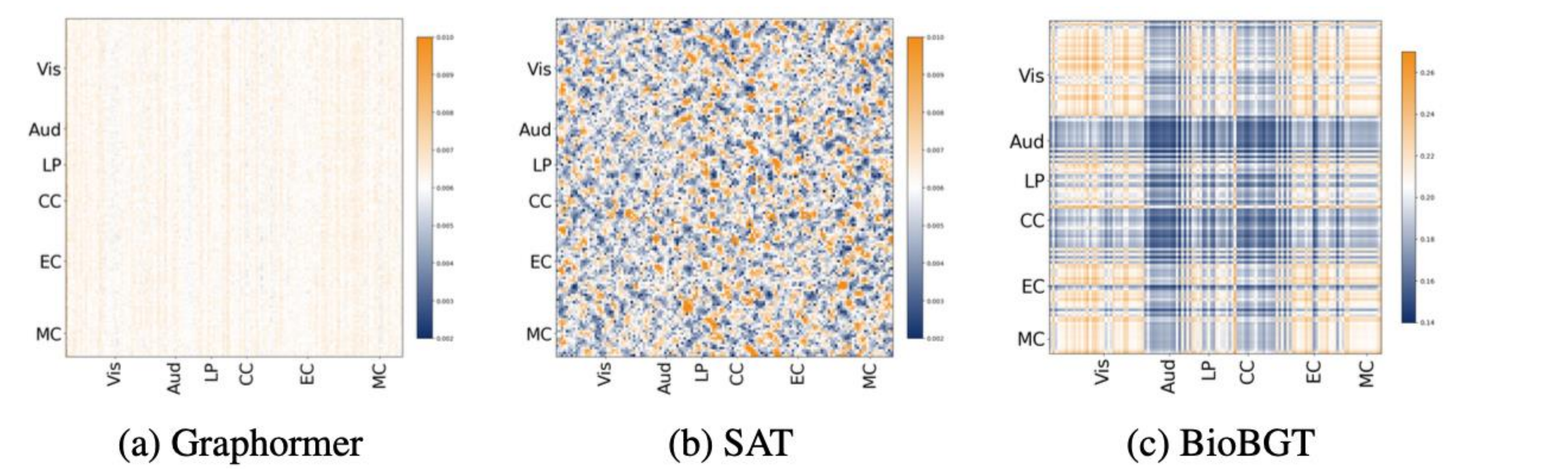

Fig: The average attention score of BioBGT (on ADHD-200) align better with the division of functional modules.