

# Hierarchical Graph Learning with Small-World Brain Connectomes for Cognitive Prediction

Yu Jiang<sup>1</sup>, Zhibin He<sup>2</sup>, Zhihao Peng<sup>1</sup>, Yixuan Yuan<sup>1</sup>

Electronic Engineering Department, The Chinese University of Hong Kong, Hong Kong SAR, China

[yxyuan@ee.cuhk.edu.hk](mailto:yxyuan@ee.cuhk.edu.hk)

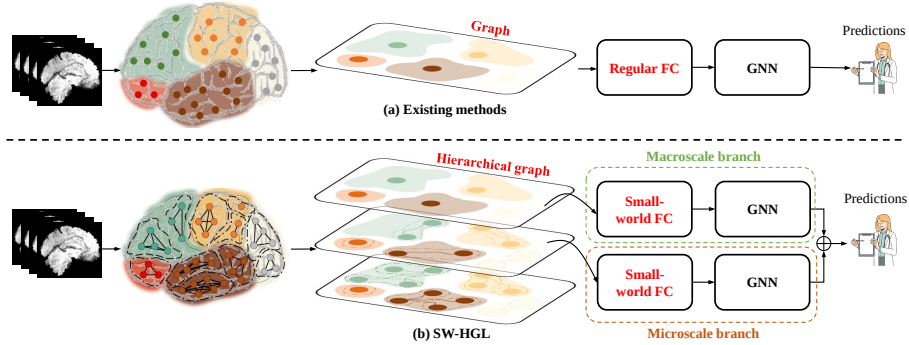
School of Automation, Northwestern Polytechnical University, Xi'an, China

**Abstract.** Functional MRI is capable of assessing an individual’s cognitive ability by blood oxygen level dependence. Due to the complexity of brain function, exploring the relationship between cognitive ability and brain functional connectivity is extremely challenging. Recently, graph neural networks have been employed to extract functional connectivity features for predicting cognitive scores. Nevertheless, these methods have two main limitations: 1) Ignore the hierarchical nature of brain: discarding fine-grained information within each brain region, and overlooking supplementary information on the functional hierarchy of the brain at multiple scales; 2) Ignore the small-world nature of brain: current methods for generating functional connectivity produce regular networks with relatively low information transmission efficiency. To address these issues, we propose a *Hierarchical Graph Learning with Small-World Brain Connectomes* (SW-HGL) framework for cognitive prediction. This framework consists of three modules: the pyramid information extraction module (PIE), the small-world brain connectomes construction module (SW-BCC), and the hierarchical graph learning module (HGL). Specifically, PIE identifies representative vertices at both micro-scale (community level) and macro-scale (region level) through community clustering and graph pooling. SW-BCC simulates the small-world nature of brain by rewiring regular networks and establishes functional connections at both region and community levels. MSFEF is a dual-branch network used to extract and fuse micro-scale and macro-scale features for cognitive score prediction. Compared to state-of-the-art methods, our SW-HGL consistently achieves outstanding performance on HCP dataset. The code is available at <https://github.com/CUHK-AIM-Group/SW-HGL>.

**Keywords:** Cognitive score prediction · Hierarchical graph · Small-world brain connectomes.

## 1 Introduction

Cognitive scores serve as indicators of intellectual ability and are strongly associated with health and mortality rates [4, 20]. Understanding the impact of brain connectomes on cognitive scores is crucial for unraveling the complexities



**Fig. 1.** Illustration of the cognitive scores prediction. (a) Previous works use regular graph structures to predict cognitive scores at a single scale. (b) The proposed SW-HGL introduces a hierarchical graph learning with small-world brain connectomes for cognitive prediction

of human brain. Consequently, a considerable amount of research has focused on exploring the relationship between brain connectomes and cognitive abilities [3, 8, 11, 15, 17, 23]. Recently, researchers have introduced graph neural network (GNN) to leverage the topological characteristics of brain connectomes, enabling the extraction of connection features between cortical regions [5, 9, 13, 22].

Despite promising initial results, accurate prediction of individual-level cognitive faces two main challenges (Fig. 1(a)): 1) **Neglect of the hierarchical characteristics:** Current methods only extract functional connectivity between the average levels of different regions, discarding fine-grained information within each region. In fact, functional connectivity exists not only on a macroscopic scale but also on a microscopic scale [7]. Hierarchical functional connectivity can provide complementary topological information about the brain and help improve the learning performance of models. 2) **Neglect of the small-world nature:** Current methods only consider local connections to generate regular brain connectomes [5, 13, 22]. Compared with small-world networks, the information transfer efficiency of regular networks is relatively low. The reason for choosing to build a small-world network is because it can process and transmit information quickly, which is similar to the brain’s high-level cognitive functions. If we consider generating brain connectomes with small-world properties, the graph learning model can transmit and process information more rapidly.

To address these issues, we propose a novel framework, the Hierarchical Graph Learning with Small-World Brain Connectomes (SW-HGL), as depicted in Fig. 1(b). First, to capture pyramid information, we cluster all vertices in one region into several communities to obtain micro-scale community representative vertices based on the Louvain algorithm [1]. While macro-scale region representative vertices are obtained through feed-forward GNN and graph pooling. Then macro-scale and micro-scale features are integrated by a dual-branch GNN network architecture. Secondly, due to high degree of clustering and short

path length of small-world brain network allows human to process information efficiently and integrate information quickly, we propose a small-world brain connectomes construction method to simulate the small-world nature of brain. This involves using the  $k$ -nearest neighbor method to construct a regular network with high degree of clustering. Then the regular network is rewired to introduce disordered edges to shorten path length. The rewiring method is used because it is more in line with the original definition of the small-world network - an intermediate state between regular networks and random networks.

## 2 Method

In Fig. 2, we introduce the overall architecture of SW-HGL pipeline. First, given one preprocessed fMRI subject  $X$ , we use pyramid information extraction module to extract the region information  $X_{re} \in \mathbb{R}^{N_1 \times t}$  at the macro-scale and community information  $X_{com} \in \mathbb{R}^{N_2 \times t}$  at micro-scale, where  $N_1$  and  $N_2$  are the numbers of the brain region and community,  $t$  is the length of time series. Then, we put  $X_{re}$  and  $X_{com}$  into small-world brain connectomes construction module respectively, obtaining the functional connection matrix  $\mathbf{G}_{re} \in \mathbb{R}^{N_1 \times N_1}$  and the  $\mathbf{G}_{com} \in \mathbb{R}^{N_2 \times N_2}$ . Finally, we process the hierarchical information extracted earlier to obtain macro-scale features  $H_{re}$  and micro-scale features  $H_{com}$ . Finally,  $H_{re}$  and  $H_{com}$  are connected and input into a fully connected layer to obtain fusion features  $F$ , which is put into a fully connected layer for cognition prediction.

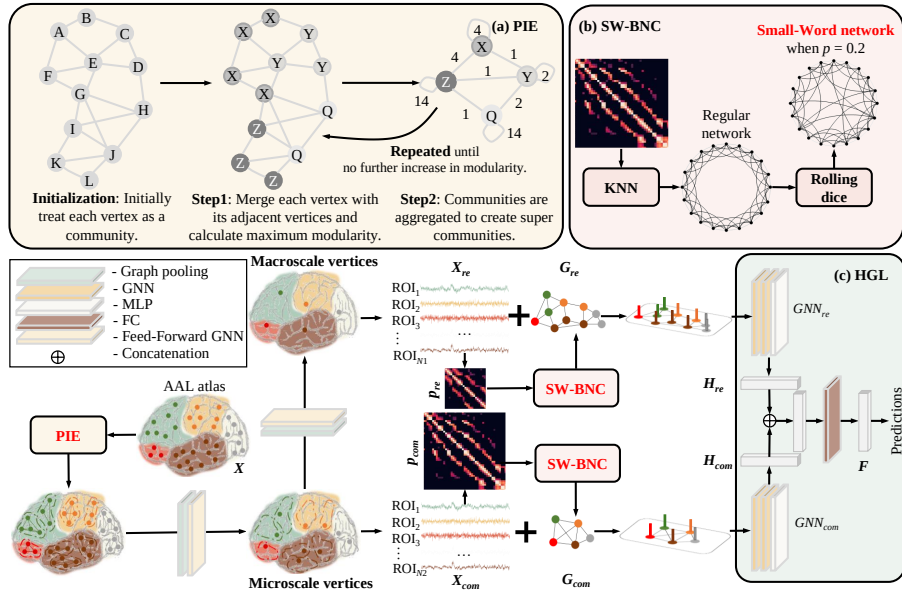
### 2.1 Pyramid information extraction

To exploit hierarchical spatial patterns of brain connectomes [7], we use the Louvain algorithm to discover pyramid structures. First, we calculate the Pearson correlation coefficient (PCC) matrix  $\mathbf{P}$  of all vertices in one brain region, and only retain the top 10% nearest neighbors (KNN) of each vertex [13]. Then, each vertex is assigned to its community (itself). Subsequently, each vertex is considered in turn to see whether moving it to a neighbor community would increase the modularity. The local optimization and community aggregation steps are then repeated until no further increase in modularity can be achieved.

Given the large number of cortical vertices, calculating the modularity increment by moving isolated vertex  $i$  to community  $com$  would result in a large computation. The reason for choosing the Louvain algorithm is that the modularity increment  $\Delta Q$  obtained is simple to calculate and the calculation amount is small. It can be calculated by Eq. (2):

$$Q = \frac{1}{2E} \sum_{i,j \in \mathcal{V}} (\mathbf{P}_{ij} - \frac{k_i k_j}{2E}) \mathbb{1}_{[com_i=com_j]}, \quad (1)$$

$$\Delta Q = [\frac{\sum_{in} + s_{i,in}}{2E} - (\frac{\sum_{tot} + s_i}{2E})^2] - [\frac{\sum_{in}}{2E} - (\frac{\sum_{tot}}{2E})^2 - (\frac{s_i}{2E})^2], \quad (2)$$



**Fig. 2.** Illustration of the cognitive score prediction based on SW-HGL. (a) PIE: Pyramid information extraction. (b) SW-BNC: Small-world brain connectomes construction. (c) HGL: Hierarchical graph learning.

where  $E$  is the sum of edge weights of the graph,  $k_i$  and  $com_i$  is the degree and the community of vertex  $i$ ,  $\mathbb{1}_{[com_i=com_j]}$  is indicator function,  $\sum_{in}$  is the sum of edge weights within community  $com$ ,  $\sum_{tot}$  is the sum of the weights of the edges connected to vertexes in community  $com$ ,  $s_i$  is the sum of the weights of the edges connected to vertex  $i$ ,  $s_{i,in}$  is the sum of the weights of the edges connecting vertex  $i$  to vertexes in community  $com$ .

After the community division of each region is obtained, we first extract the micro-scale information. For each community, we calculate PCC between each vertex within the community and input  $\mathbf{P}$  and  $X$  into a feed-forward GNN layer [18] for message propagation. The vertex with the highest degree within the community is identified as the micro-scale vertex by graph pooling. The functional time series  $X_{com}$  of this micro-scale vertex already incorporates the temporal and structural information of all vertices within the community. We then acquire the macro-scale information. For all micro-scale vertices within each region, we calculate the PCC between each micro-scale vertex and input  $\mathbf{P}_{com}$  and  $X_{com}$  into a feed-forward GNN layer for vertex message propagation. The vertex with the highest degree within the brain region is identified as the macro-scale vertex. The time series  $X_{re}$  of macro-scale vertex already incorporates the temporal and structural information of all vertices within the region.

## 2.2 Small-World brain connectomes construction

As we know, there are a large number of local connections in the human brain connectomes, but there are also a small number of long-distance connections in brain connectomes [14]. This small-world nature of brain have an important impact on human cognitive ability. The high degree of clustering of the brain connectomes allows the human brain to perform efficient information processing and learning, while the short path length allows the human brain to quickly integrate information and make decisions. This efficient information processing capability is the basis of human advanced cognitive functions.

To simulate the small-world nature of brain, we propose a small-world functional connectivity generation algorithm. Firstly, calculate the PCC between each vertex on both community and region level. Then use KNN method and retain the top 10 % of connection information for each vertex, to initially construct the functional connectivity  $\mathbf{G}$ . The formula is as follows:

$$\mathbf{G}(v_i, v_j) = \begin{cases} 1, & \text{if } j \in \text{Neigh}_k(i), \\ 0, & \text{if } j \notin \text{Neigh}_k(i), \end{cases} \quad (3)$$

where  $\text{Neigh}_k(v_i)$  is the set of  $k$  vertices closest to the vertex  $i$ . Secondly, each edge is randomly reconnected, for any vertex  $i \in V$ , a connection is randomly selected to replace its original connection. The selected edge, which connects vertex  $i$  and vertex  $j$  ( $j \in V$ ), is reconnected to another node  $u$  ( $u \in V, j \neq u$ ) with a probability of  $p$ , transforming the original edge  $(i, j)$  into the edge  $(i, u)$ . When  $p = 0$ , the functional connectivity is not reconnected, and the resulting brain connectomes remain a regular network. When  $p \neq 0$ , all edges will be reconnected with a probability of  $p$ , and as the value of  $p$  increases, the small-world of the network becomes more pronounced. When  $p = 1$ , the brain connectomes transform into a random network.

## 2.3 Hierarchical graph learning

The hierarchical graph learning consists of two different GNN-based modules,  $GNN_{com}$  and  $GNN_{re}$ , each used to process the micro-scale and macro-scale information. Both consist of two layers of GNN and one layer of MLP. Following the idea of neighborhood aggregation and updating, GNN is used to process the spatial patterns to obtain the representations of vertices through the message function  $M(\text{Message})$  and update function  $U(\text{Update})$  after  $k$  rounds of the message propagation mechanism [21]. The message propagation process is as follows:

$$m_i^{k+1} = \sum_{j \in N(i)} M(h_i^k, h_j^k, w_{ij}), \quad (4)$$

$$h_i^{k+1} = U(h_i^k, m_i^{k+1}), \quad (5)$$

where  $h_i^k$  and  $h_i^{k+1}$  represent the features of vertex  $i$  of the  $k$ -th and  $(k+1)$ -th layers,  $h_j^k$  is the feature of vertex  $j$  in the  $k$ -th layer,  $w_{ij}$  is the weight on the edge

of vertex  $i$  and  $j$ ,  $m_i^{k+1}$  is the aggregated feature between vertex  $i$  and its neighbor  $j$  in the  $k$ -th layer. After conducting hierarchical graph learning, we obtained the features of each subject at two different scales. To fully utilize the potential complementary information provided by different spatial scales, we designed a feature fusion module to integrate hierarchical features and used the fused graph features for prediction tasks. The feature fusion module is represented as:

$$F = H_f \parallel H_c \in \mathbb{R}^{2D}, \quad (6)$$

where  $\parallel$  represents the concatenation operation. The hierarchical graph representation of each subject is a  $2D$  representation vector, which is then fed into a fully connected layer for prediction.

#### 2.4 Model Optimization:

To train the model, we use the root mean square error (RMSE) to optimize the model parameters. We added orthogonal constraints in the vertex feature matrix to avoid the problem of gradient vanishing or gradient explosion during the optimization process. The total loss function can be represented as Eq. (7):

$$Loss = L_{RMSE} + \lambda L_F, \quad (7)$$

$$L_F = \left\| \frac{1}{m} F^T F - I \right\|_2, \quad (8)$$

where  $L_{RMSE}$  is the RMSE loss function,  $L_F$  is the orthogonal constraint at two scales,  $m = \max(F^T F)$  and  $\lambda$  is the hyperparameter used to balance the contributions of the two terms in Eq. (7).

### 3 Experiments and Results

#### 3.1 Experimental setup

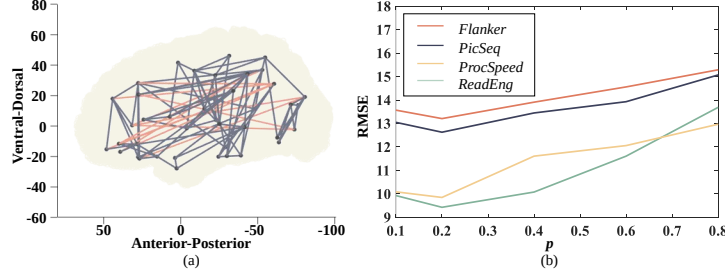
**Dataset:** We validated the SW-HGL on HCP dataset. This dataset includes 900 fMRI scans. After eliminating some invalid subjects, we ultimately used 841 subjects for experiments. Then, We map all data to a common space, the cortical surface partition of Anatomical Automatic Labeling (AAL) atlas [19]. Further, we follow an established procedure to process the left and right cerebral cortex only. To quantify the learning behavior cognitive level of subjects, we used four cognitive scores, namely Visual Episodic Memory (*PicSeq*), Inhibition (*Flanker*), Processing Speed (*ProcSpeed*), and Vocabulary (*ReadEng*).

**Evaluation:** To measure the model performance more objectively, we compared the cognitive scores and prediction data after denormalization, and then employed three evaluation metrics: RMSE, mean absolute percentage error (MAPE), and PCC. The smaller the values of RMSE and MAPE, and the larger the PCC value, the higher the model performance. In addition, we compared SW-HGL with several state-of-the-arts: 1) GNN-based fMRI data processing methods, including BrainGNN [13], RegGNN [6], Meta-RegGNN [9] and BrainGB [5], 2) spatial-temporal GNN-based methods ST-SSL [10] and ST-GAT [2].

**Table 1.** Comparison with state-of-the-art methods and the ablation studies on HCP dataset. The best performance is highlighted in **bold** and the suboptimal performance is highlighted in underline.

| Metrics \ Score   | Metrics | <i>Flanker</i>                    | <i>PicSeq</i>                      | <i>ProcSpeed</i>                   | <i>ReadEng</i>                     |
|-------------------|---------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Methods           |         |                                   |                                    |                                    |                                    |
| BrainGNN          | RMSE    | 10.18 $\pm$ 0.98                  | <u>13.46 <math>\pm</math> 0.73</u> | <u>15.06 <math>\pm</math> 1.29</u> | <u>10.43 <math>\pm</math> 0.61</u> |
|                   | MAPE    | <u>7.33 <math>\pm</math> 0.69</u> | <u>9.97 <math>\pm</math> 0.81</u>  | <u>10.31 <math>\pm</math> 0.85</u> | <u>7.54 <math>\pm</math> 0.65</u>  |
|                   | PCC     | 0.13 $\pm$ 0.15                   | 0.15 $\pm$ 0.04                    | 0.15 $\pm$ 0.02                    | 0.18 $\pm$ 0.10                    |
| RegGNN            | RMSE    | 10.22 $\pm$ 0.58                  | 18.55 $\pm$ 3.33                   | 15.61 $\pm$ 0.92                   | 10.83 $\pm$ 0.77                   |
|                   | MAPE    | <u>7.31 <math>\pm</math> 0.36</u> | 12.35 $\pm$ 1.84                   | 11.20 $\pm$ 1.25                   | 7.52 $\pm$ 0.30                    |
|                   | PCC     | <u>0.13 <math>\pm</math> 0.02</u> | 0.18 $\pm$ 0.09                    | 0.15 $\pm$ 0.10                    | 0.26 $\pm$ 0.03                    |
| Meta-RegGNN       | RMSE    | 14.82 $\pm$ 0.80                  | 18.55 $\pm$ 3.33                   | 17.93 $\pm$ 1.35                   | 17.20 $\pm$ 2.89                   |
|                   | MAPE    | 10.81 $\pm$ 0.76                  | 12.35 $\pm$ 1.85                   | 12.84 $\pm$ 1.13                   | 11.70 $\pm$ 2.37                   |
|                   | PCC     | 0.08 $\pm$ 0.7                    | 0.08 $\pm$ 0.09                    | 0.18 $\pm$ 0.10                    | 0.28 $\pm$ 0.05                    |
| BrainGB           | RMSE    | 17.05 $\pm$ 1.76                  | 15.97 $\pm$ 0.89                   | 19.01 $\pm$ 4.00                   | 12.18 $\pm$ 1.00                   |
|                   | MAPE    | 11.59 $\pm$ 1.13                  | 10.91 $\pm$ 0.95                   | 13.58 $\pm$ 2.44                   | 8.44 $\pm$ 1.72                    |
|                   | PCC     | 0.15 $\pm$ 0.03                   | <u>0.18 <math>\pm</math> 0.07</u>  | 0.14 $\pm$ 0.10                    | 0.21 $\pm$ 0.07                    |
| ST-SSL            | RMSE    | 45.57 $\pm$ 32.89                 | 43.96 $\pm$ 31.81                  | 65.02 $\pm$ 25.26                  | 10.75 $\pm$ 0.02                   |
|                   | MAPE    | 30.68 $\pm$ 22.32                 | 6.12 $\pm$ 6.13                    | 27.77 $\pm$ 10.76                  | 7.42 $\pm$ 0.01                    |
|                   | PCC     | <u>0.17 <math>\pm</math> 0.02</u> | 0.15 $\pm$ 0.01                    | 0.10 $\pm$ 0.11                    | 0.14 $\pm$ 0.10                    |
| ST-GAT            | RMSE    | 11.54 $\pm$ 2.17                  | 15.45 $\pm$ 2.77                   | 28.26 $\pm$ 8.12                   | 10.73 $\pm$ 0.01                   |
|                   | MAPE    | 7.45 $\pm$ 0.87                   | 10.98 $\pm$ 0.84                   | 17.41 $\pm$ 4.94                   | <u>7.41 <math>\pm</math> 0.01</u>  |
|                   | PCC     | 0.15 $\pm$ 0.09                   | 0.07 $\pm$ 0.11                    | 0.15 $\pm$ 0.09                    | 0.13 $\pm$ 0.10                    |
| <i>w/o</i> PIE    | RMSE    | 10.02 $\pm$ 0.13                  | 12.90 $\pm$ 0.06                   | 13.71 $\pm$ 0.01                   | 11.76 $\pm$ 0.30                   |
|                   | MAPE    | 7.13 $\pm$ 0.10                   | 9.80 $\pm$ 0.03                    | 8.89 $\pm$ 0.03                    | 8.13 $\pm$ 0.29                    |
|                   | PCC     | 0.22 $\pm$ 0.01                   | 0.20 $\pm$ 0.02                    | 0.18 $\pm$ 0.04                    | 0.24 $\pm$ 0.01                    |
| <i>w/o</i> SW-BCC | RMSE    | 9.72 $\pm$ 0.12                   | 12.82 $\pm$ 0.05                   | 13.74 $\pm$ 0.03                   | 10.30 $\pm$ 0.34                   |
|                   | MAPE    | 6.90 $\pm$ 0.10                   | 9.75 $\pm$ 0.04                    | 8.93 $\pm$ 0.04                    | 6.99 $\pm$ 0.26                    |
|                   | PCC     | 0.24 $\pm$ 0.02                   | 0.25 $\pm$ 0.02                    | 0.18 $\pm$ 0.02                    | 0.35 $\pm$ 0.01                    |
| <i>w/o</i> HGL    | RMSE    | 9.71 $\pm$ 0.11                   | 12.96 $\pm$ 0.06                   | 13.82 $\pm$ 0.01                   | 12.28 $\pm$ 0.57                   |
|                   | MAPE    | 6.90 $\pm$ 0.07                   | 9.80 $\pm$ 0.04                    | 9.04 $\pm$ 0.02                    | 8.15 $\pm$ 0.39                    |
|                   | PCC     | 0.25 $\pm$ 0.01                   | 0.25 $\pm$ 0.03                    | 0.18 $\pm$ 0.02                    | 0.28 $\pm$ 0.03                    |
| Our               | RMSE    | <b>9.43 <math>\pm</math> 0.12</b> | <b>12.63 <math>\pm</math> 0.05</b> | <b>13.21 <math>\pm</math> 0.01</b> | <b>9.85 <math>\pm</math> 0.34</b>  |
|                   | MAPE    | <b>6.60 <math>\pm</math> 0.08</b> | <b>9.64 <math>\pm</math> 0.05</b>  | <b>8.63 <math>\pm</math> 0.02</b>  | <b>6.61 <math>\pm</math> 0.26</b>  |
|                   | PCC     | <b>0.27 <math>\pm</math> 0.01</b> | <b>0.28 <math>\pm</math> 0.01</b>  | <b>0.19 <math>\pm</math> 0.02</b>  | <b>0.38 <math>\pm</math> 0.01</b>  |

**Implementation Details:** To train our model, we implemented it through PyTorch [16] and Adam [12] selected as the optimizer, periodically adjusting the weight decay learning rate between 1e-2 and 1e-5. The batch size was set to 64, and the maximum training epoch was set to 500. We set the hyperparameters  $p=0.2$  to generate the small-world brain connectomes, the macro-scale hidden layer node feature channel  $d=32$ , and the micro-scale hidden layer node feature channel  $d=64$ . We used 650 subjects as the training set, 51 subjects as the valid set and 130 subjects as the test set. During the testing phase, we report the average and standard deviation of the results from 10 experiments.



**Fig. 3.** Visualize results. (a) Small-World brain connectomes. (b) Sensitivity analysis.

### 3.2 Experimental Results

**Comparison with State-of-the-arts:** We listed the comparison results in Table 1. ST-SSL and ST-GAT have almost the worst performance. Compared with the fMRI processing method based on GNN, SW-HGL performs the best in the four cognitive score prediction tasks. Especially for the PCC metric, our improvements in *Flanker*, *ProcSpeed* and *ReadEng* were around 10%, which shows that SW-HGL has a stronger fitting ability. And the variance of our 10-time results is small, which is related to the fact that the brain connectome we generated is a small-world network. Because small-world networks are highly robust, even if some nodes or connections are corroded or distorted, the overall function of the network will not be greatly affected. Then, we select a subject to construct a small-world brain connectomes, and visualize it in Fig. 3(a). Vertices are placed sequentially based on their anatomical location in the brain. The dark blue and orange connections represent short-distance (Distance < 75 mm) and long-distance (Distance > 75 mm) connections. The results show that the proposed method can indeed increase the number of long-distance connections.

**Ablation Studies:** Table 1 presents the ablation study validating the effectiveness of each module in the proposed SW-HGL. Results demonstrate that pyramid information extraction module enhances the effectiveness of our model, as evidenced by the removal of it RMSE  $10.02 \pm 0.13$ . Introducing small-word nature improves RMSE by reducing error tolerance, as indicated by removing it accuracy  $9.72 \pm 0.12$ . Utilizing hierarchical graph learning improves model performance by considering supplementary information on the functional hierarchy, as indicated by removing it RMSE  $9.71 \pm 0.11$ . Then we conducted sensitivity experiments by setting the hyperparameter  $p$  to different values to study the impact of the degree of small-world nature of the constructed brain connectomes on model performance. We record the RMSE results in Fig. 3(b). Our research results show that the highest prediction accuracy can be achieved when  $p=0.2$ . However, setting  $p$  too small can lead to insufficient small-worldness of the functional connectivity, causing a slight decline in performance, conversely, setting  $p$  too large results in the overly chaos, leading to a decrease in model performance.

## 4 Conclusion

This paper proposes a novel framework SW-HGL for cognitive score prediction. Specifically, the pyramid information extraction module captures macro-scale brain region representative vertex information and micro-scale community representative vertex information. The small-world brain connectomes construction module constructs functional connectivity between brain region representative vertices and between community representative vertices. The hierarchical graph learning module uses the dual-branch GNN to extract and fuse macro-scale and micro-scale features to predict cognitive scores. Experimental results demonstrate the superiority of the proposed model.

**Acknowledgment.** This work was supported by Hong Kong Research Grants Council (RGC) General Research Fund 14204321.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008)
2. Chen, H., Jiang, Y., Zhang, X., Zhou, Y., Wang, L., Wei, J.: Spatio-temporal graph attention network for sintering temperature long-range forecasting in rotary kilns. *IEEE Transactions on Industrial Informatics* **19**, 1923–1932 (2023)
3. Cole, M.W., Yarkoni, T., Repovš, G., Anticevic, A., Braver, T.S.: Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *Journal of Neuroscience* **32**, 8988–8999 (2012)
4. Colom, R., Escorial, S., Shih, P.C., Privado, J.: Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences* **42**, 1503–1514 (2007)
5. 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. *IEEE Transactions on Medical Imaging* **42**, 493–506 (2023)
6. Hanik, M., Demirtaş, M.A., Gharsallaoui, M.A., Rekik, I.: Predicting cognitive scores with graph neural networks through sample selection learning. *Brain Imaging and Behavior* **16**, 1123–1138 (2021)
7. He, Z., Li, W., Zhang, T., Yuan, Y.: H2gm: A hierarchical hypergraph matching framework for brain landmark alignment. In: *Medical Image Computing and Computer Assisted Intervention*. pp. 548–558. Springer Nature Switzerland (2023)
8. Huang, H., Hu, X., Zhao, Y., Makkie, M., Dong, Q., Zhao, S., Guo, L., Liu, T.: Modeling task fmri data via deep convolutional autoencoder. *IEEE Transactions on Medical Imaging* **37**, 1551–1561 (2018)
9. Jegham, I., Rekik, I.: Meta-reggnn: Predicting verbal and full-scale intelligence scores using graph neural networks and meta-learning. In: *Predictive Intelligence in Medicine*. pp. 203–211. Springer Nature Switzerland (2022)

10. Ji, J., Wang, J., Huang, C., Wu, J., Xu, B., Zhang, Z.W.J., Zheng, Y.: Spatio-temporal self-supervised learning for traffic flow prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 4356–4364 (2023)
11. Jiang, R., Calhoun, V.D., Fan, L., Zuo, N., Jung, R., Qi, S., Lin, D., Li, J., Zhuo, C., Song, M., Fu, Z., Jiang, T., Sui, J.: Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores. *Cerebral Cortex* **30**, 888–900 (2019)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations (2015)
13. 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)
14. Liao, X., Vasilakos, A.V., He, Y.: Small-world human brain networks: Perspectives and challenges. *Neuroscience and Biobehavioral Reviews* **77**, 286–300 (2017)
15. Pamplona, G.S.P., Neto, G.S.S., Rosset, S.R.E., Rogers, B.P., Salmon, C.E.G.: Analyzing the association between functional connectivity of the brain and intellectual performance. *Frontiers in Human Neuroscience* **9**, 61 (2015)
16. 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)
17. Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T.: Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols* **12**, 506–518 (2017)
18. Song, X., Lian, J., Huang, H., Luo, Z., Zhou, W., Lin, X., Wu, M., Li, C., Xie, X., Jin, H.: xgcnn: An extreme graph convolutional network for large-scale social link prediction. In: Proceedings of the ACM Web Conference 2023. p. 349–359. Association for Computing Machinery (2023)
19. 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**, 273–289 (2002)
20. Wang, Q., Wu, M., Fang, Y., Wang, W., Qiao, L., Liu, M.: Modularity-constrained dynamic representation learning for interpretable brain disorder analysis with functional mri. In: Medical Image Computing and Computer Assisted Intervention. pp. 46–56 (2023)
21. Xiao, T., Chen, Z., Wang, D., Wang, S.: Learning how to propagate messages in graph neural networks. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. p. 1894–1903. Association for Computing Machinery (2021)
22. Yan, J., Chen, Y., Xiao, Z., Zhang, S., Jiang, M., Wang, T., Zhang, T., Lv, J., Becker, B., Zhang, R., Zhu, D., Han, J., Yao, D., Kendrick, K.M., Liu, T., Jiang, X.: Modeling spatio-temporal patterns of holistic functional brain networks via multi-head guided attention graph neural networks (multi-head gagnns). *Medical Image Analysis* **80**, 102518 (2022)
23. Zhao, Y., Li, X., Huang, H., Zhang, W., Zhao, S., Makkie, M., Zhang, M., Li, Q., Liu, T.: Four-dimensional modeling of fmri data via spatio-temporal convolutional neural networks (st-cnns). *IEEE Transactions on Cognitive and Developmental Systems* **12**, 451–460 (2020)