1 **Connectome-based individual prediction of cognitive behaviors via**
2 **the graph propagation network reveals directed brain network**
3 **topology**

4 **Dongya Wu¹, Xin Li^{2*}, Jun Feng^{1,3*}**

5 ¹School of Information Science and Technology, Northwest University, Xi'an, 710127,
6 China

7 ²School of Mathematics, Northwest University, Xi'an, 710127, China

8 ³State-Province Joint Engineering and Research Center of Advanced Networking and
9 Intelligent Information Services, School of Information Science and Technology,
10 Northwest University, Xi'an, 710127, China

11

12 *** Correspondence:**

13 Xin Li, lixin@nwu.edu.cn

14 Jun Feng, fengjun@nwu.edu.cn

15

16

17

Abstract

The brain connectome supports the information flow underlying human cognitions and should reflect the individual variability in human cognitive behaviors. Various studies have utilized the brain connectome to predict individual differences in human behaviors. However, traditional studies viewed the brain connectome feature as a vector of one dimension, a method which neglects topological structures of the brain connectome. To utilize topological properties of the brain connectome, we proposed that graph neural network which combines graph theory and neural network can be adopted. Different from previous node-driven graph neural networks that parameterize on the node feature transformation, we designed an edge-driven graph neural network named graph propagation network that parameterizes on the information propagation within the brain connectome. We compared various models in predicting the individual total cognition based on the resting-state functional connectome. The edge-driven graph propagation network showed the highest prediction accuracy and outperformed the node-driven graph neural network and traditional partial least square regression. The graph propagation network also revealed a directed network topology encoding the information flow, indicating that the high-level association cortices are responsible for the information integration underlying the total cognition. These results suggest that the edge-driven graph propagation network can better explore the topological structure of the brain connectome and can serve as a new method to associate the brain connectome and human behaviors.

Keywords: graph propagation network, individual prediction, brain network topology, information flow, functional connectome, cognitive behavior

1 **Introduction**

2 Brain connectivity network provides pathways for transferring information
 3 between brain regions and serves as the basis of human cognition at a network level.
 4 Hence, individual differences in human cognitive behaviors should be reflected in the
 5 brain connectivity network. Investigating the brain connectivity network basis of
 6 these individual differences facilitates the understanding of brain network information
 7 processing mechanism that supports human cognitions and may benefit the
 8 developments of individualized medicine.

9 Numerous researchers have investigated the relationship between brain
 10 connectivity networks and individual differences in human cognitive behaviors. Smith
 11 et al. (2015) revealed a positive-negative axis linking human behavioral measures to a
 12 specific pattern of brain functional connectivity. Researchers have further adopted
 13 resting-state functional connectivity to predict human intelligence (Finn et al., 2015),
 14 sustained attention (Rosenberg et al., 2016) and creative ability (Beaty et al., 2018), or
 15 adopted functional connectivity derived from different task states to predict cognitive
 16 behaviors (Greene et al., 2018), and a connectome-based predictive modeling
 17 framework has been built to predict individual behaviors from brain connectivity
 18 (Shen et al., 2017). Zimmermann et al. (2018) also compared the unique and
 19 overlapping mapping of functional and structural connectivity on cognitive behaviors.

20 However, most current researches that investigate the relationship between brain
 21 networks and human behaviors treat the brain network features as a vector of one
 22 dimension, a method which neglects topological properties of the graph-structured
 23 brain network. To utilize these topological properties, one should treat the brain
 24 network as a graph instead of a vector, with connections between brain regions
 25 represented as edge features and regional properties represented as node features.
 26 Numerous studies have used graph theory to analyze the graph-represented brain
 27 network (Bullmore and Sporns, 2009), for instance, calculating the node degrees and

central brain regions (van den Heuvel and Sporns, 2011), analyzing the small-world properties (Sporns and Zwi, 2004) and modularity properties of brain networks (Bullmore and Sporns, 2012), investigating the high order proximity properties of brain regions (Becker et al., 2018), or extracting various other graph metrics such as clustering, transitivity, integration (Kazeminejad and Sotero, 2018) and so on. However, the graph theory method needs to extract the graph features subjectively and cannot analyze the graph-structured data in an end-to-end fashion.

In view of the excellent performance achieved by the neural network and deep learning in extracting data features automatically (LeCun et al., 2015), it is promising to apply the neural network method to the graph-structured brain network to automatically learn the relationship between brain networks and human behaviors in an end-to-end fashion. The combination of neural network and graph-structured data is known as graph neural network (GNN). GNN has achieved rapid developments in recent years (Defferrard et al., 2016; Kipf and Welling, 2017; Veličković et al., 2018; Gong and Cheng, 2019; Xu et al., 2019) and obtained state-of-the-art results in many areas, such as the citation network, social network, and protein network, but the applications of GNN to the brain network are still very limited to date (Ktena et al., 2018; Parisot et al., 2018; Yang et al., 2019; Kim and Ye, 2020). In addition, these applications are based on the node-driven GNNs with parameters concentrating on the transformation of node features, a strategy that is not specialized for utilizing topological properties of the brain connectome. Edge-driven GNN as used by Wu et al. (2021) is needed to explore topological structures of the brain connectome.

To apply the GNN to the brain network analysis in a better way, we proposed an edge-driven GNN model named graph propagation network (GPN) (Fig. 1). The regional properties propagate within the brain network via connections between neighboring regions to form the representation of human behaviors. The relationship between brain networks and human behaviors is encoded in the propagation coefficients that are automatically learned by the error backpropagation. In this work,

1 we utilized resting-state functional magnetic resonance imaging (rfMRI) data from the
2 human connectome project (HCP) to construct the brain functional connectome.
3 Different structures of the GPN were explored to investigate the connectome-based
4 individual prediction of the total cognition. We further compared the GPN with the
5 node-driven GNNs and partial least square regression (PLSR) that is widely adopted
6 by traditional studies to demonstrate the priority of the GPN. Finally, we adopted the
7 saliency map to investigate topological properties of the brain connectome underlying
8 the total cognition.

9 **Materials and Methods**

10 **HCP imaging and behavioral data**

11 We adopted the rfMRI and behavioral data in the minimally pre-processed data
12 provided by the HCP S1200 release (Glasser et al., 2013). We selected 997 subjects
13 that have the rfMRI and behavioral acquisitions.

14 The rfMRI data were projected onto the 2 mm standard CIFTI grayordinates
15 space, with the multimodal surface matching algorithm based on areal features used
16 for accurate inter-subject registration (Robinson et al., 2014). Parameters acquisition
17 and processing of the rfMRI data are described in Smith et al. (2013). Briefly, the
18 rfMRI scans were acquired at 2 mm isotropic resolution, with a fast TR sampling rate
19 at 0.72s using multiband pulse sequences (Ugurbil et al., 2013). Each subject had four
20 15-minute rfMRI runs, with a total of 1,200 time points per run. The rfMRI data were
21 further processed by the ICA-FIX algorithm to remove the nuisances (Griffanti et al.,
22 2014; Salimi-Khorshidi et al., 2014).

23 The behavioral data are described in Barch et al. (2013) and we adopted
24 behaviors from the cognitive domain. As many cognitive measures are correlated, we
25 used a single cognitive behavior “Cognition Total Composite Score” to measure the
26 total cognition and further normalized this cognitive behavior to zero mean and unit

variance. The “Cognition Total Composite Score” composes of both fluid and crystallized cognitive measures, with the fluid cognitive measures involving paradigms such as Flanker, Dimensional Change Card Sort, Picture Sequence Memory, List Sorting and Pattern Comparison, and the crystallized cognitive measures involving paradigms such as Picture Vocabulary and Reading Tests.

Calculation of the functional connectome

We adopted the Human Connectome Project Multi-Modal Parcellation version 1.0 (HCP-MMP1.0) that contains 360 brain regions to define parcels of the brain (Glasser et al., 2016). Each subject’s four runs of rfMRI time series were demeaned, variance-normalized, and concatenated along the time axis successively. The time series of all vertices within each parcel were averaged. The averaged time series of each brain region were correlated with each other to form the functional connectome. We neither applied global signal regression on the rfMRI time series nor set any thresholds on the functional connectome, so as to avoid additional processing.

Edge-driven graph propagation network

Fig. 1A shows the architecture of a 2-layer GPN. The GPN composes of three parts, i.e., the input layer, the middle layers (first and second layer in Fig. 1A), and the readout layer. In the input layer, the functional connectivity network constructed from the rfMRI time series is fed into the GPN as graph-structured data, with the node values initialized as ones. Each middle layer contains a graph convolution computation followed by an activation function. The graph convolution computation is central to the GPN and Fig. 1B shows examples of this computation. Each node integrates information from its neighbors to update its value, with the integration weighted by the strength of connections and graph convolution coefficients. The graph convolution computation can be realized by the matrix multiplication operation (Fig. 1C). In the readout layer, a linear model utilizes the node values of all layers to

output the prediction of behavior.

Mathematically, a middle layer of the GPN can be represented in matrix form as follows:

$$X_{n \times 1}^k = \sigma(W_{n \times n}^k \odot A_{n \times n} X_{n \times 1}^{k-1}), \quad (1)$$

where σ is the activation function, k represents the k -th middle layer of the network, n represents the number of nodes in the GPN ($n = 360$ in this study), A represents the functional connectome, W indicates the extent to which each connection participates in the information propagation, and X represents the node value in the GPN. W exerts on A via the operation $\odot$ which represents element-wise multiplication. The initialized node value in the 0-th layer is a vector of ones: $X_{n \times 1}^0 = [1, 1, \dots, 1]_{n \times 1}$. The node value in the k -th layer encodes the functional connectivity feature of k orders. To enrich the feature representation of the GPN, we concatenated the output of each layer to obtain the final representation as:

$$X_G = \text{concatenate}(\{X_{n \times 1}^k | k = 1, 2, \dots, K\}). \quad (2)$$

The readout layer utilizes a linear regression model to transform the final representation as:

$$\hat{y} = \beta_{1 \times nK} X_G + b, \quad (3)$$

where β represents the transformation vector, b represents the bias term and $\hat{y}$ is the prediction result.

Node-driven graph neural network

We briefly described the node-driven GNN to make a comparison with the edge-driven GPN. A matrix representation of the node-driven GNN is as follows:

$$X_{n \times 1}^k = \text{MLP}(\hat{A}_{n \times n} X_{n \times d}^{k-1}), \quad (4)$$

where MLP is the abbreviation of multi-layer perceptron that is used for node feature transformation, d is the dimension of node feature, and X represents the node feature that is the functional connectivity feature of each node. If we used a single-layer perceptron for feature transformation, model (4) can be represented as:

$$X_{n \times 1}^k = \sigma(\hat{A}_{n \times n} X_{n \times d}^{k-1} \theta_{d \times l}^k), \quad (5)$$

where θ is the feature transformation parameter that transforms the node feature with d dimensions into a new feature with l dimensions, and σ is an activation function such as the rectified linear unit. If $\hat{A}$ is the adjacency matrix A_{adj} that is symmetrically normalized by the node degree, then model (5) corresponds to the graph convolutional network (Kipf and Welling, 2017). If $\hat{A} = A_{adj} + (1 + \epsilon)I$, where I is the identity matrix, then model (4) and model (5) correspond to the graph isomorphism network (GIN) (Xu et al., 2019). We also concatenated the node features of all layers and utilized linear regression to obtain the graph prediction output. Comparing model (2) with model (5), one sees that parameters of the edge-driven GPN concentrate on the propagation coefficient W and parameters of the node-driven GNN concentrate on the feature transformation θ . These two parameterization strategies are not conflicted and one can also combine both parameterizations in one GNN. The combination of both edge-driven and node-driven strategies leads to a GNN as follows:

$$X_{n \times 1}^k = \sigma(W_{n \times n}^k \odot A_{n \times n} X_{n \times d}^{k-1} \theta_{d \times l}^k). \quad (6)$$

This combination method is also tested in this study.

18 Training and testing procedures for the graph neural network

The dataset was randomly divided into a training set and a testing set with a ratio of nine to one. In the training stage, the mean squared error between the actual behavior and prediction was minimized to learn the parameters of the GPN. To ensure that the learning rate is not affected by the variance of each training batch, the mean squared error was normalized by the actual behavior's variance of each training batch. In the testing stage, we used the Pearson correlation (r) to assess the prediction performance of the GPN. The value of r assesses the similarity between the actual and prediction value, and r^2 approximates the proportion of individual behavioral

1 differences that can be explained by the model. We performed this process 10 times
2 with different random divisions of the dataset and used the mean Pearson correlation
3 coefficient to assess the model.

4 **Saliency map that identifies the important connections**

5 The propagation coefficient W cannot be directly used to indicate the extent to
6 which each connectivity contributes to the individual prediction, as the GPN contains
7 the batch normalization and the parameters of batch normalization also affects the
8 information propagation. In order to calculate the importance of each connectivity
9 feature, we adopted the saliency map method. The saliency map estimates each
10 connectivity feature's importance by calculating the extent to which the prediction
11 result changes when each connectivity feature changes an infinitely small value.
12 Mathematically, this amounts to calculating the gradient of each connectivity feature
13 with respect to the prediction result. In practice, after we finished the training of the
14 GPN, we calculated the gradient using each subject's functional connectivity feature,
15 and the final saliency map was calculated as the mean gradient across all subjects and
16 10 repetitions.

17 **Implementation details of the graph neural network**

18 We implemented the GPN and GIN with PyTorch. The propagation coefficients
19 in the GPN were initialized by the Xavier normal distribution with a gain of 0.1. The
20 model was trained via the stochastic gradient descent optimizer with a Nesterov
21 momentum of 0.9 and the training batch size was 128. The initial learning rate was
22 0.0001 for the GPN and a 0.1 multiplicative factor of learning rate decay was set at
23 250 epochs (300 total training epochs). The initial learning rate was 0.001 for the GIN
24 and a 0.1 multiplicative factor of learning rate decay was set at 50 epochs (100 total
25 training epochs). We adopted the batch normalization technique before the activation
26 function and adopted the dropout technique in each layer's output with a default rate

1 of 0.5. We also added Gaussian random noise to the functional connectome of each
 2 training batch. The variance of the Gaussian random noise was equal to the variance
 3 of each connectome feature across subjects. We trained the models on an NVIDIA
 4 GeForce GTX 1080 Ti graphic processing unit.

5 **Results**

6 **Comparison between graph propagation networks with different structures**

7 We focused on three aspects of the GPN's structures, i.e., shared or unshared
 8 parameter strategy, depth of the network, and choice of the activation function.
 9 Different middle layers can have different graph propagation coefficients (W) or
 10 shared the same graph propagation coefficient. As the size of the propagation
 11 coefficient in each layer is 360×360 , the total parameters of unshared strategy grow
 12 more rapidly as the network gets deeper but the shared strategy controls the size of
 13 total parameters. We tested the performance of GPNs with different parameter
 14 strategies as the depth of the network gets larger under different choices of activation
 15 functions. As shown in Table 1, even though the total parameters of unshared
 16 strategies are several times than those of shared strategies, the GPNs with unshared
 17 strategies did not exhibit serious overfitting problems and showed comparable or even
 18 better prediction performance than those with shared strategies. In addition, GPNs
 19 that adopt linear activation functions showed better prediction performance than those
 20 that adopt rectified linear activation functions. In contrast to previous neural networks
 21 that need nonlinear activation functions to make the whole model nonlinear, the GPN
 22 can still achieve nonlinear mapping when linear activation functions are used. This
 23 nonlinear nature of the GPN resides in the information propagation process through
 24 consecutive layers.

1 Contrasting the graph propagation network with the baseline model

2 We selected the GIN as a representation of node-driven GNNs, as the GIN is
3 powerful than other GNN variants (Xu et al., 2019). The parameter ϵ is set as 0 for
4 simplicity. We used each brain region’s functional connectivity features as the input
5 node features. We derived the adjacency matrix in the GIN by setting a threshold on
6 the functional connectivity matrix, i.e., the connections that have absolute values
7 larger than the threshold are conceived to be positive or negative adjacent. To make
8 the parameter size of GPN and that of GIN comparable, we firstly adopted a single
9 layer MLP for node feature transformation and the dimension of features was set as
10 360 in the GIN. In this model setting, the parameter size is 360×360 in each layer of
11 the GIN. In addition, we tested a 2-layer MLP for node feature transformation and the
12 dimension of features was set as 64. The parameter size is 360×64+64×64 in each
13 layer under this model setting. In these GINs, we adopted rectified linear activation
14 functions rather than linear functions, as nonlinear activation functions are necessary
15 for MLP to make the transformation nonlinear. Table 2 recorded the prediction
16 performance of GINs under different settings. Setting a larger threshold on the
17 functional connectivity network leads to GINs that showed better prediction
18 performance. Under various model settings, the GPNs outperforms GINs when the
19 parameter size is comparable or when GINs adopted a 2-layer neural network for
20 node feature transformation.

21 We also compared the GPN with the PLSR, as the PLSR performs well in the
22 individual prediction based on the brain connectome (Yoo et al., 2018; Chen et al.,
23 2019). As the functional connectivity network is symmetric, only the upper triangular
24 part of the network is extracted and viewed as a vector for the PLSR. Divisions of the
25 dataset for PLSR are the same as those for GNNs. We determined the number of
26 PLSR components via ten-fold cross-validation on the training dataset and the model
27 that achieved minimum cross-validation error was used for assessment on the testing
28 dataset. We implemented the PLSR using the function “plsregress” in Matlab. Table 2

1 showed the prediction performance of the PLSR. All GPNs showed higher prediction
2 accuracy than the PLSR but GINs showed lower prediction accuracy than the PLSR.

3 **Testing the combination of edge-driven and node-driven strategies**

4 The edge-driven and node-driven GNNs differ in their parameterization strategies,
5 with the edge-driven method focusing on the modeling of information propagation
6 and the node-driven method focusing on the transformation of node features. We
7 tested whether the combination of both strategies can improve the individual
8 prediction of cognition based on the functional connectivity network (Table 3).
9 Similar to the performance of GPNs, the combination strategy also showed better
10 performance when adopting a linear activation function. The combination strategy
11 showed higher prediction performance than GINs and showed lower prediction
12 performance than GPNs. This result showed that adding graph propagation coefficient
13 to the GINs can facilitate the utilization of graph's topological structures and improve
14 the individual prediction based on the brain connectome. However, adding the node
15 feature transformation to GPNs reduces the performance of GPNs and the graph
16 propagation parameterization is sufficient to utilize the brain network's structure.

17 **The directed network topology underlying the total cognition**

18 We used the saliency map to visualize the functional connectivity network
19 features of the 2-layer GPN (adopting unshared strategy and linear activation function)
20 in Fig.2. As the functional connectivity network is symmetric, the feature pattern of
21 traditional models such as the PLSR is also symmetric. However, the functional
22 network features of the GPN showed an asymmetric pattern that derives from the
23 asymmetric information propagation. Since the information propagation and
24 integration are realized by the graph convolution in Fig. 1C, rows of the saliency
25 matrix represent regions that integrate information from neighboring regions, and
26 columns of the saliency matrix represent regions that send information out. Thus the

1 saliency matrix revealed a directed network topology that encodes the information
 2 flow. One can identify each functional connection's importance relative to the total
 3 cognition via the saliency matrix in Fig.2. We further calculated the node degree of
 4 each region as the summation of the absolute saliency matrix vertically or horizontally.
 5 The vertical node degree (on the left side of the saliency matrix) represents the
 6 importance of each region in integrating information from other regions, and the
 7 horizontal node degree (on the bottom side of the saliency matrix) represents the
 8 importance of each region in sending information out to other regions. The horizontal
 9 node degree showed a relatively homogeneous distribution pattern across cortices
 10 compared with the vertical node degree. The vertical node degree is plotted on the
 11 cortical surface and showed small values in the primary visual, auditory, and motor
 12 cortices, while showed large values in the high-level association cortices such as the
 13 dorsolateral and medial prefrontal, inferior frontal, and parietal cortices. This node
 14 degree distribution pattern indicates that the high-level association brain regions are
 15 responsible for the information integration supporting the total cognition.

16 **Discussion**

17 In the current study, we proposed an edge-driven GNN named GPN that can
 18 capture the individual association between the brain connectome and the human total
 19 cognition in a better way. Compared with either the traditional PLSR model that
 20 views the brain connectivity network as a vector and ignores the topological structure
 21 of the brain connectome or the node-driven GNN that adopts the feature
 22 transformation parameterization, the proposed GPN can model the information flow
 23 within the brain network and facilitate the use of the brain network's topological
 24 structure. The edge-driven GPN exhibits higher performance than the PLSR and the
 25 node-driven GNN in predicting the individual total cognition based on the functional
 26 connectome and reveals a directed information flow supporting the total cognition.
 27 The GPN shows promising future in the field of individualized prediction based on

the brain connectome.

He et al. (2020) showed that the graph convolutional neural network performs worse than traditional regression models, a study in which the authors adopted a node-driven GNN model (Kipf and Welling, 2017; Parisot et al., 2017; Parisot et al., 2018). Similarly, the results in our study also showed that another node-driven GIN performs worse than the traditional PLSR model. The previous node-driven GNNs and traditional regression models are not well adapted to process the graph-structured brain network data and cannot utilize the topological structure of the brain network. Since GNNs usually possess higher model complexities than traditional regression models, node-driven GNNs are prone to suffer overfitting problems and thus exhibited poor generalization abilities than simple regression models such as the PLSR. However, the edge-driven GPN tested in the current study exhibits superior performance compared with the traditional PLSR. This result indicates that GPNs which are specially designed for modeling information propagation in brain networks are more suitable for processing graph-structured brain network data.

We demonstrated that the edge-driven GPN is superior to the node-driven GNN in predicting individual total cognition from the resting-state functional connectivity network. This result suggests that viewing the functional connectome as the node feature may break the topological structure of the brain network, while the edge-driven GPN that can capture the information propagation within the brain network is better specialized for processing the graph-structured brain connectivity network. However, this does not mean that the edge-driven GPN is also better than the node-driven GNN in other conditions. In this study, we only utilized the functional connectome and did not adopt other brain node features, such as the cortical thickness, myelin, task activation and so on. If more node features are available to the GNN, the node-driven strategy that is better at node feature transformation may express a more powerful ability. Future works can incorporate more brain node features into the brain connectome and investigate whether the combination of edge-driven and node-driven

1 strategies can facilitate the processing of multi-modality brain connectome.

2 Compared with the deep learning models that can contain over one hundred
3 layers, the GPN adopted in this study contains only several layers. The reasons behind
4 this shallow structure may be twofold. Since the brain network possesses the
5 small-world property (Sporns and Zwi, 2004; Bassett and Bullmore, 2006), the
6 minimum path length between any pair of brain regions is only several steps. Thus the
7 functional information of each brain region can propagate to any other brain regions
8 in the GPN within several steps. This short information propagation range within the
9 brain may lead to the GPN with a shallow structure. On the other hand, the sample
10 size of the data utilized in this study is only about one thousand, a number that is
11 much smaller than that adopted in deep learning studies. As the depth of GPNs grows
12 larger, the model complexity also gets larger, a condition in which adopting a small
13 sample size is more prone to overfitting. Future works can adopt datasets with more
14 samples, such as the biobank dataset (Alfaro-Almagro et al., 2018), to test whether
15 adding more samples can improve the generalization performance.

16 The GPN revealed a directed information flow via the saliency map and the
17 directed information flow further indicates that the high-level association cortices are
18 responsible for the information integration that supports the total cognition. Even
19 though traditional methods such as the PLSR can also reveal important connectivity
20 features belonging to association cortices (Zimmermann et al., 2018; Wu and Li,
21 2019), the revealed features do not contain information about the direction of the
22 information flow and researchers cannot identify whether association cortices send
23 out or integrate information. The information flow within the brain network is
24 essentially directed, but many functional or structural brain connectivity networks
25 constructed via neuroimaging are undirected. Several methods such as Granger
26 causality (Dhamala et al., 2008) and dynamic causal modeling (Friston et al., 2003)
27 can estimate the directed connectivity based on neuroimaging time series. GPN
28 estimates the directed information flow through a connectome-based individual

prediction that links the brain network with human behaviors, a manner which differentiates GPN with those previous methods. A behavior context is necessary for deriving the directed information flow from the undirected brain connectivity network, as effective connectivity should be defined within a certain context (Park and Friston, 2013). Therefore, the revealed directed information flow is only a subset of the whole brain network and is relevant to the predicted cognitive behavior. The directed information flow revealed by the GPN also coincides with the view of neuroscience, as the high-level association cortices integrate functional information from primary cortices (Purves et al., 2001), which is important for supporting complex cognitions.

In conclusion, we proposed an edge-driven GPN that can greatly improve the individual prediction of the human total cognition from the brain functional connectome. The edge-driven GPN serves as a new kind of GNN relative to node-driven GNNs and focuses on the topological structure of the brain connectome. This advancement contributes to linking the individual differences in human cognitive behaviors to the brain connectome and suggests new perspectives on the information propagation mechanism within the brain connectome.

Data availability: The HCP S1200 data release is publicly available online at <https://www.humanconnectome.org/>.

Conflict of interest: The authors declare that they have no conflict of interest.

Funding: This work was supported by the National Key Research and Development Program of China (grant No. 2017YFB1002504), the Natural Science Foundation of China (grant No. 62073260).

Acknowledgments: Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil;

1 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH
2 Blueprint for Neuroscience Research.

3
4 **Author contributions:** Wu, Li and Feng designed the research, Wu performed the
5 data analysis, Wu, Li and Feng wrote and revised the paper.

7 Reference

- 8 Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R., Griffanti, L., Douaud,
9 G., et al. (2018). Image processing and Quality Control for the first 10,000 brain imaging
10 datasets from UK Biobank. *Neuroimage* 166, 400-424. doi:
11 10.1016/j.neuroimage.2017.10.034.
- 12 Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., et al.
13 (2013). Function in the human connectome: task-fMRI and individual differences in
14 behavior. *Neuroimage* 80, 169-189. doi: 10.1016/j.neuroimage.2013.05.033.
- 15 Bassett, D.S., and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* 12(6),
16 512-523. doi: 10.1177/1073858406293182.
- 17 Beaty, R.E., Kenett, Y.N., Christensen, A.P., Rosenberg, M.D., Benedek, M., Chen, Q., et al.
18 (2018). Robust prediction of individual creative ability from brain functional
19 connectivity. *Proc. Natl. Acad. Sci. U. S. A.* 115(5), 1087-1092. doi:
20 10.1073/pnas.1713532115.
- 21 Becker, C.O., Pequito, S., Pappas, G.J., Miller, M.B., Grafton, S.T., Bassett, D.S., et al. (2018).
22 Spectral mapping of brain functional connectivity from diffusion imaging. *Sci. Rep.* 8(1),
23 1411. doi: 10.1038/s41598-017-18769-x.
- 24 Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of
25 structural and functional systems. *Nat. Rev. Neurosci.* 10(3), 186-198. doi:
26 10.1038/nrn2575.
- 27 Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev.*
28 *Neurosci.* 13(5), 336-349. doi: 10.1038/nrn3214.
- 29 Chen, C., Cao, X., and Tian, L. (2019). Partial Least Squares Regression Performs Well in
30 MRI-Based Individualized Estimations. *Front. Neurosci.* 13(1282), 1282. doi:
31 10.3389/fnins.2019.01282.
- 32 Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). "Convolutional neural networks
33 on graphs with fast localized spectral filtering", in: *Advances in neural information*
34 *processing systems*.

- 1 Dhamala, M., Rangarajan, G., and Ding, M. (2008). Analyzing information flow in brain
2 networks with nonparametric Granger causality. *Neuroimage* 41(2), 354-362. doi:
3 10.1016/j.neuroimage.2008.02.020.
- 4 Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., et al. (2015).
5 Functional connectome fingerprinting: identifying individuals using patterns of brain
6 connectivity. *Nat. Neurosci.* 18(11), 1664-1671. doi: 10.1038/nn.4135.
- 7 Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage*
8 19(4), 1273-1302.
- 9 Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., et al.
10 (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615), 171-178.
11 doi: 10.1038/nature18933.
- 12 Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., et
13 al. (2013). The minimal preprocessing pipelines for the Human Connectome Project.
14 *Neuroimage* 80, 105-124. doi: 10.1016/j.neuroimage.2013.04.127.
- 15 Gong, L., and Cheng, Q. (2019). "Exploiting Edge Features for Graph Neural Networks", in:
16 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- 17 Greene, A.S., Gao, S., Scheinost, D., and Constable, R.T. (2018). Task-induced brain state
18 manipulation improves prediction of individual traits. *Nat Commun* 9(1), 2807. doi:
19 10.1038/s41467-018-04920-3.
- 20 Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E.,
21 et al. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved
22 resting state network imaging. *Neuroimage* 95, 232-247.
- 23 He, T., Kong, R., Holmes, A.J., Nguyen, M., Sabuncu, M.R., Eickhoff, S.B., et al. (2020).
24 Deep neural networks and kernel regression achieve comparable accuracies for
25 functional connectivity prediction of behavior and demographics. *Neuroimage* 206,
26 116276. doi: <https://doi.org/10.1016/j.neuroimage.2019.116276>.
- 27 Kazeminejad, A., and Sotero, R.C. (2018). Topological Properties of Resting-State fMRI
28 Functional Networks Improve Machine Learning-Based Autism Classification. *Front.*
29 *Neurosci.* 12(1018), 1018. doi: 10.3389/fnins.2018.01018.
- 30 Kim, B.H., and Ye, J.C. (2020). Understanding Graph Isomorphism Network for rs-fMRI
31 Functional Connectivity Analysis. *Front. Neurosci.* 14(630), 630. doi:
32 10.3389/fnins.2020.00630.
- 33 Kipf, T.N., and Welling, M. (2017). "Semi-supervised classification with graph convolutional
34 networks", in: *International Conference on Learning Representations*.
- 35 Ktena, S.I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., et al. (2018). Metric
36 learning with spectral graph convolutions on brain connectivity networks. *Neuroimage*
37 169, 431-442.
- 38 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521(7553), 436-444. doi:

1 10.1038/nature14539.

2 Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., et al. (2018). Disease
3 prediction using graph convolutional networks: Application to Autism Spectrum
4 Disorder and Alzheimer's disease. *Med. Image Anal.* 48, 117-130. doi:
5 10.1016/j.media.2018.06.001.

6 Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Moreno, R.G., Glocker, B., et al. (2017).
7 "Spectral graph convolutions for population-based disease prediction", in: *International*
8 *Conference on Medical Image Computing and Computer-Assisted Intervention*.

9 Park, H.J., and Friston, K. (2013). Structural and functional brain networks: from connections
10 to cognition. *Science* 342(6158), 1238411. doi: 10.1126/science.1238411.

11 Purves D, Augustine GJ, D, F., Lawrence C Katz, Anthony-Samuel LaMantia, James O
12 McNamara, et al. (2001). *Neuroscience*. Sunderland (MA): Sinauer Associates.

13 Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., et al.
14 (2014). MSM: a new flexible framework for Multimodal Surface Matching. *Neuroimage*
15 100, 414-426. doi: 10.1016/j.neuroimage.2014.05.069.

16 Rosenberg, M.D., Finn, E.S., Scheinost, D., Papademetris, X., Shen, X., Constable, R.T., et al.
17 (2016). A neuromarker of sustained attention from whole-brain functional connectivity.
18 *Nat. Neurosci.* 19(1), 165-171. doi: 10.1038/nn.4179.

19 Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., and Smith,
20 S.M. (2014). Automatic denoising of functional MRI data: combining independent
21 component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449-468.

22 Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., et al.
23 (2017). Using connectome-based predictive modeling to predict individual behavior
24 from brain connectivity. *Nat. Protoc.* 12(3), 506-518. doi: 10.1038/nprot.2016.178.

25 Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., et
26 al. (2013). Resting-state fMRI in the Human Connectome Project. *Neuroimage* 80,
27 144-168. doi: 10.1016/j.neuroimage.2013.05.039.

28 Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E., Glasser, M.F., et al.
29 (2015). A positive-negative mode of population covariation links brain connectivity,
30 demographics and behavior. *Nat. Neurosci.* 18(11), 1565-1567. doi: 10.1038/nn.4125.

31 Sporns, O., and Zwi, J.D. (2004). The small world of the cerebral cortex. *Neuroinformatics*
32 2(2), 145-162. doi: 10.1385/ni:2:2:145.

33 Ugurbil, K., Xu, J.Q., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., et al.
34 (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the
35 Human Connectome Project. *Neuroimage* 80, 80-104. doi:
36 10.1016/j.neuroimage.2013.05.012.

37 van den Heuvel, M.P., and Sporns, O. (2011). Rich-club organization of the human
38 connectome. *J. Neurosci.* 31(44), 15775-15786. doi:

1 10.1523/JNEUROSCI.3539-11.2011.

2 Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2018). "Graph

3 attention networks", in: *International Conference on Learning Representations*.

4 Wu, D., and Li, X. (2019). "Prediction of Behavioral Traits via Anatomical Connectivity

5 Fingerprint", in: *Proceedings of the 2019 9th International Conference on Biomedical*

6 *Engineering and Technology*.

7 Wu, D., Li, X., and Feng, J. (2021). Multi-Hops Functional Connectivity Improves Individual

8 Prediction of Fusiform Face Activation via a Graph Neural Network. *Front. Neurosci.* 14,

9 596109.

10 Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019). "How Powerful are Graph Neural

11 Networks?", in: *International Conference on Learning Representations*.

12 Yang, H., Li, X., Wu, Y., Li, S., Lu, S., Duncan, J.S., et al. (2019). "Interpretable

13 Multimodality Embedding of Cerebral Cortex Using Attention Graph Network for

14 Identifying Bipolar Disorder", in: *International Conference on Medical Image*

15 *Computing and Computer-Assisted Intervention*.

16 Yoo, K., Rosenberg, M.D., Hsu, W.-T., Zhang, S., Li, C.-S.R., Scheinost, D., et al. (2018).

17 Connectome-based predictive modeling of attention: Comparing different functional

18 connectivity features and prediction methods across datasets. *Neuroimage* 167, 11-22.

19 doi: 10.1016/j.neuroimage.2017.11.010.

20 Zimmermann, J., Griffiths, J.D., and McIntosh, A.R. (2018). Unique Mapping of Structural

21 and Functional Connectivity on Cognition. *The Journal of Neuroscience* 38(45),

22 9658-9667. doi: 10.1523/JNEUROSCI.0900-18.2018

23

24

Tables and Figure legends

Table 1. Prediction performance of GPNs. The accuracy is represented by the mean Pearson correlation and the standard deviation across ten random divisions of the dataset.

Model	Parameter strategy	Layers	Activation function	Accuracy(<i>r</i>)
GPN	Unshared	1	Linear/Relu	0.545±0.055/0.525±0.052
GPN	Unshared	2	Linear/Relu	0.553±0.054/0.542±0.054
GPN	Unshared	3	Linear/Relu	0.551±0.053/0.537±0.051
GPN	Shared	2	Linear/Relu	0.553±0.054/0.524±0.049
GPN	Shared	3	Linear/Relu	0.546±0.047/0.540±0.031

Table 2. Prediction performance of baseline models.

Model	MLP	Layers	Features	Activation function	Threshold	Accuracy(<i>r</i>)
GIN	1	1	360	Relu	0.5	0.286±0.053
GIN	1	2	360	Relu	0.5	0.279±0.085
GIN	1	1	360	Relu	0.9	0.424±0.051
GIN	1	2	360	Relu	0.9	0.430±0.054
GIN	1	3	360	Relu	0.9	0.406±0.063
GIN	2	1	64	Relu	0.9	0.405±0.075
GIN	2	2	64	Relu	0.9	0.392±0.074
PLSR	-	-	-	-	-	0.497±0.063

Table 3. Prediction performance of models with the combination strategy.

Model	MLP	Layers	Features	Activation function	Parameter strategy	Accuracy(<i>r</i>)
GPN+GIN	1	1	360	Relu	Unshared	0.447±0.059
GPN+GIN	1	2	360	Relu	Unshared	0.438±0.063
GPN+GIN	1	1	360	Linear	Unshared	0.494±0.071
GPN+GIN	1	2	360	Linear	Unshared	0.475±0.057

Figure 1. Schematic illustration of the GPN. (A) The graph-structured individual functional connectivity network serves as the input of the GPN. The node values are initialized as all ones. The graph convolution (GC) is the central computation of the GPN. The initial node values update through consecutive layers under the graph convolution. The node values of each layer are concatenated and the readout layer transforms this final graph representation to the individual prediction via a linear regression. (B) Each node updates its value by integrating the information from its neighbors. The graph convolution indicates the extent to which that each connectivity pathway participates in the information propagation. (C) The graph convolution can be realized by the matrix multiplication. The graph convolution parameters exert on

the functional connectivity network via element-wise multiplication. The resulting matrix multiplies the vector that represents node value to update the node information.

Figure 2. The saliency map of functional network features relative to the total cognition. The saliency map of functional network features is shown in the matrix. The saliency matrix value represents the gradient or importance of each connection with respect to the total cognition. The vertical and horizontal node degree of the saliency matrix are shown on the left side and bottom side of the matrix respectively. We grouped brain regions that belong to the same cortex and the cortical name follows the Supplementary Neuroanatomical Results for A Multi-modal Parcellation of Human Cerebral Cortex (Glasser et al., 2016). The vertical node degree is also plotted on the cortical surface to visualize the distribution pattern.