QUANTIFYING STRUCTURE-FUNCTION COUPLING IN THE HUMAN BRAIN USING VARIATIONAL GRAPH CONTRASTIVE LEARNING

Yihan Zhang
School of Science, China Pharmaceutical University, Nanjing 210009, China
zyh@stu.cpu.edu.cn

October 26, 2025

ABSTRACT

This study proposes a novel method based on variational graph contrastive learning to quantify structure–function coupling at the regional level of the brain. We obtained whole-cortex scale structural connectivity matrices from publicly available studies and constructed matched functional connectivity matrices using resting-state functional MRI (rs-fMRI) from the WU-Minn Human Connectome Project (HCP). The core of the model is a dual-branch variational graph convolutional network, which aligns the latent representations of the same brain region across structural and functional modalities via contrastive learning, augmented with distance constraints and regularization. We define the Gaussian kernel output of the latent representations of a brain region in structural and functional connectivity as its structure–function coupling (SFC). The results indicate that the SFC is strongest in the Visual network and weakest in the Orbito-affective network, with its distribution pattern aligning with known cognitive hierarchy principles. Further ablation experiments and single run experiments validate the effectiveness of the model components and the robustness of the SFC metric. This study provides a new computational framework for extracting stable multimodal coupling features from complex brain network data.

***K*eywords** structure-function Coupling (SFC), graph neural networks (GNN), contrastive learning

1 Introduction

The structural connectivity of the brain provides the fundamental anatomical scaffold for its functional activity. Understanding this structure–function coupling is one of the central challenges in uncovering the organizational principles of the brain. The precise quantification of structure–function coupling not only aids in understanding the neural basis of cognitive functions but also offers a critical perspective to reveal the pathological mechanisms of neuropsychiatric disorders [1].

In recent years, the use of deep learning models, particularly graph neural networks, to study brain networks has become an emerging trend. However, we observe that most studies that employ deep learning models to compute the structure-function coupling follow a similar paradigm: using structural connectivity as input to train a model aimed at predicting or reconstructing functional connectivity [2, 3]. In this paradigm, the relationship (e.g., correlation coefficient) between the model-predicted functional connectivity and the empirical functional connectivity is often used as a metric for the strength of the structure-function coupling. While this paradigm has achieved success, it might underutilize the rich information in FC that is independent of SC. Therefore, we aim to explore a more symmetric modeling framework to fully leverage the potentially abundant information in FC not entirely explained by SC.

To this end, we designed a dual-branch variational graph neural network model based on contrastive learning. This model learns low-dimensional latent representations for each brain region from SC and FC separately, and uses the contrastive loss function to directly constrain the latent representations of the same brain region across the two modalities to be as close as possible. This design forces the model to explore the shared, deep correspondences between SC and

FC, rather than deriving one modality from the other. Furthermore, the model preserves the geometric consistency between the latent space and the original connectivity patterns through a distance loss, and enhances robustness via the variational design.

Ultimately, from this aligned cross-modal latent space, we define a region-level SFC as the Gaussian kernel output of a brain region's latent representations in SC and FC. Utilizing large-sample data from the HCP, we validated that the distribution of this metric across different brain networks aligns with known neurobiological principles, and further explored its sex differences and stability across different fMRI runs. Ablation experiments confirmed the crucial roles of the contrastive loss and variational design in our model. The model proposed in this work provides a powerful new tool for finely quantifying structure–function coupling at the regional level of brain, and its symmetric design philosophy opens up new avenues for the integrative analysis of multimodal brain imaging data.

2 Methods

2.1 Dataset

2.1.1 SC

We utilized publicly available data from the study *A Whole-Cortex Probabilistic Diffusion Tractography Connectome* to obtain preprocessed structural connectivity (SC) data for 1065 subjects [4]. In that study, the authors used T1-weighted structural MRI and diffusion MRI (dMRI) data from the WU-Minn Human Connectome Project (HCP) [5]. They performed probabilistic fiber tracking using FSL [6] and generated SC matrices based on the HCP-MMP 1.0 atlas (360 brain regions) [7]. Furthermore, the authors applied fractional scaling to the SC, defined by the formula:

$$F(DTI_{i,j}) = \frac{DTI_{i,j}}{\sum_{x=1}^{N} DTI_{i,x} + \sum_{y=1}^{N} DTI_{y,j}}, where\ x \neq i\ \&\ y \neq j. \tag{1}$$

This means the fractionally scaled connection between two distinct regions defined as the number of streamlines starting in A and ending in B, divided by the total number of streamlines connected to either A or B while excluding within-parcel connections. Subsequently, the authors also symmetrized the SC matrix by taking the arithmetic mean of the A–B and B–A fractionally scaled connection weights, i.e.:

$$F_{i,j} = \frac{F_{i,j} + F_{j,i}}{2}. \tag{2}$$

Additionally, the authors noted that probabilistic tractography values are approximately log-normal distribution, thus requiring a log-transformation prior to subsequent analysis. However, the SC data we obtained from the public data were not log-transformed. Therefore, in this study, we firstly log-transformed (log10) the acquired SC data, followed by Min-Max scaling to the range [0, 1] using the formula:

$$F_{i,j} = \frac{F_{i,j} - \min(\mathbf{F})}{\max(\mathbf{F}) - \min(\mathbf{F})}. \tag{3}$$

2.1.2 FC

We acquired resting-state functional MRI (rs-fMRI) data from the WU-Minn Human Connectome Project (HCP) [5], which had been preprocessed with ICA-FIX denoising. The dataset includes 1003 subjects, with each subject having four runs: `REST1_LR`, `REST1_RL`, `REST2_LR`, and `REST2_RL`, representing different sessions and phase encoding directions. To maintain consistency with the SC data, we firstly used Workbench [8] to convert the rs-fMRI data from the `fs_LR` space to the fsaverage space and subsequently separated the data by hemisphere. Then, using the NiBabel library, we extracted the mean BOLD time series for each brain region according to the same HCP-MMP 1.0 atlas. Finally, we computed the Pearson correlation coefficients between the time series of all brain regions to construct the functional connectivity matrices. As the resulting FC values were already approximately normal distribution, no further transformations were applied.

2.1.3 Sparsification and Sample pairs

Since only a minority of strong connections dominate in both SC and FC, and to reduce computational load, we independently sparsified all the connectivity matrices generated above. Specifically, for SC, we retained the top 20% of strongest connections and set the diagonal elements to zero. For FC, we retained connections with absolute values greater than 0.2 and preserved the diagonal elements.

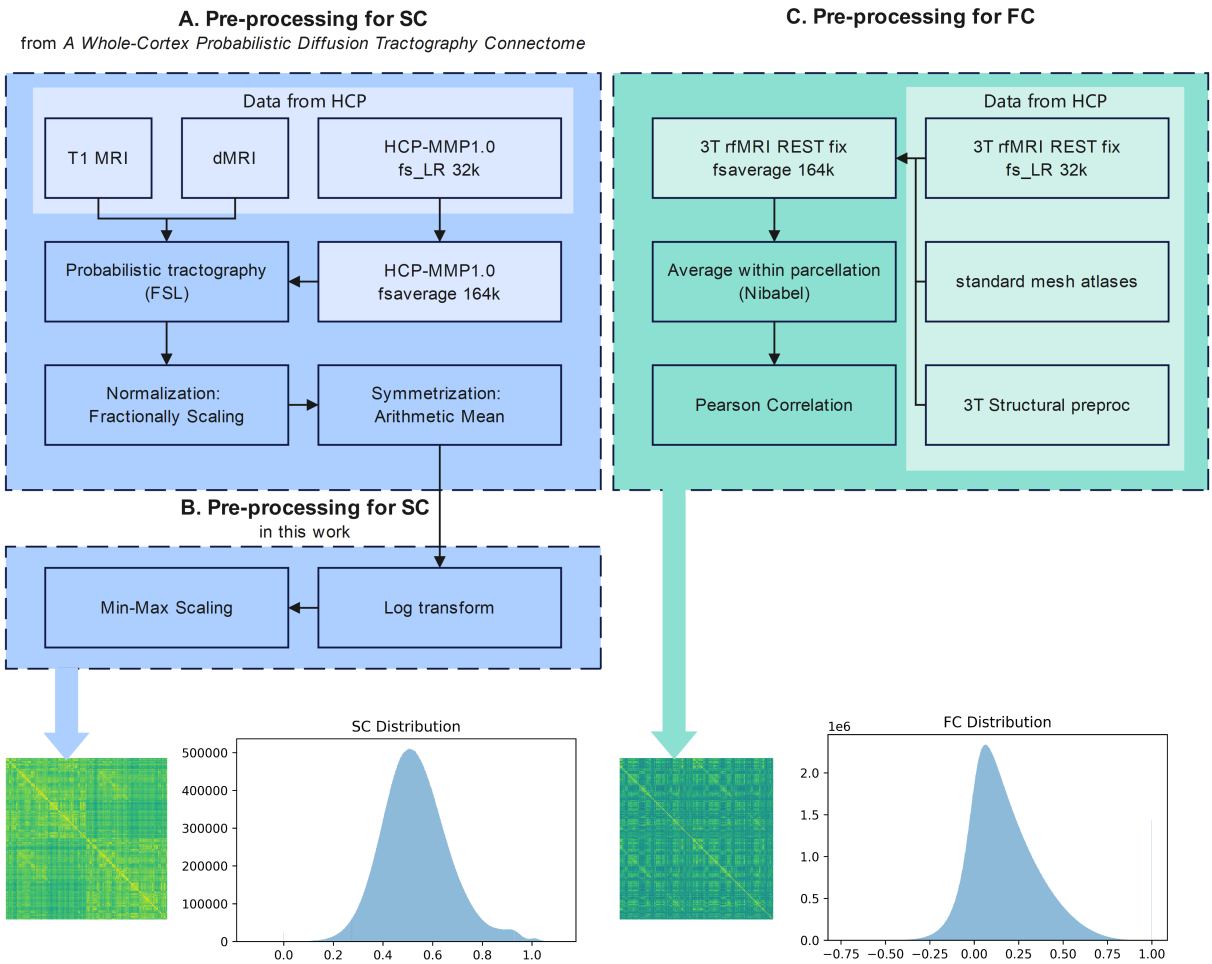

Figure 1: The process of data preprocessing. The data of the distribution graphs come from all SC or FC values, It can be seen that they all approach normal distribution after transformation.

After the aforementioned processing, we obtained SC data for 1065 subjects and FC data for 1003 subjects. Among these, 998 subjects had both SC and FC data. Furthermore, each subject's FC data comprised four different runs. To fully utilize all available runs, we replicated each subject's SC data four times, pairing each copy with a different FC run to form sample pairs. Consequently, we ultimately obtained 3992 sample pairs. This operation can be considered a form of data augmentation, and its rationality is validated in the results section through experiments using single runs.

Finally, we represented all SC and FC matrices as graphs. In these graphs, each node represents a brain region, the node features are one-hot vectors, and the connection strengths were used as edge weights.

2.2 Model

Overall, the model consists of two variational graph neural networks (GNNs), which respectively accept SC and FC, and are associated with contrastive loss, and ultimately output latent representations of brain regions.

2.2.1 Variational GCN

Both the SC and FC use an identical GNN architecture, utilizing variational graph convolutional network (GCN) for feature extraction. GCN is a type of Graph Neural Network that performs feature extraction by applying aggregation

and update operations on the graph without altering the graph structure [9]. It is computed as follows:

$$\begin{aligned} \mathbf{X}' &= \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}\mathbf{X}\mathbf{\Theta} \\ \hat{\mathbf{A}} &= \mathbf{A} + \mathbf{I} \end{aligned} \tag{4}$$

Here, $\mathbf{X}$ represents the node features, $\mathbf{\Theta}$ denotes the trainable weights, $\hat{\mathbf{D}}$ is the degree matrix of the nodes, $\mathbf{A}$ is the adjacency matrix, and $\mathbf{I}$ is the identity matrix, indicating the addition of self-loops. Self-loops allow nodes to consider their own features during aggregation.

As shown in the formula, the standard GCN typically uses node degree for normalization. However, in this study, since both SC and FC are weighted graphs, we use the absolute-valued weighted degree for normalization. The absolute-valued weighted degree of a node is defined as the sum of the absolute values of the weights of all its connecting edges.

During preprocessing, we set the diagonal elements of the SC matrix to zero because within-parcel connectivity were not examined [4], so we do not add self-loops for SC in the GCN. In contrast, the diagonal elements of the FC matrix represent perfect autocorrelation within each brain region, so self-loops are retained for FC.

In addition to the standard GCN, our model incorporates a variational component. This is computed as follows:

$$\mathbf{z} = \mu + \sigma \odot \epsilon \tag{5}$$

Here, μ and σ are outputs of the GCN, representing the mean and standard deviation of the latent distribution. $\mathbf{z}$ is the final latent representation, and ϵ is random noise sampled from a standard normal distribution. This technique, commonly used in Variational Autoencoder (VAE) [10], enables the model to learn a distribution rather than a fixed representation, thereby enhancing the model's robustness. The above equation is also known as the reparameterization.

We use the Parametric Rectified Linear Unit (PReLU) as the activation function, defined by the formula:

$$\text{PReLU}(x) = \max(0, x) + a * \min(0, x) \tag{6}$$

PReLU is a variant of the ReLU activation function, where the coefficient a controls the slope in the negative region.

2.2.2 Loss

The model incorporates three distinct loss: Contrastive Loss, Distance Loss, and KL Loss.

The Contrastive Loss [11] is shared between the SC and FC model. It quantifies the distance between the latent representations of corresponding SC and FC nodes. By minimizing this loss, the latent representations of all brain regions from SC and FC are pulled closer together. It is calculated as follows:

$$ContrastiveLoss = \|\mathbf{z}_{sc} - \mathbf{z}_{fc}\|_2^2 \tag{7}$$

The Distance Loss is an internal loss specific to either the SC model or the FC model. It measures the discrepancy between the Euclidean distance of the latent representations of two connected nodes (within the same SC or FC graph) and the absolute value of their connection weight (from the pre-sparsified matrix). The formula is:

$$DistanceLoss = \text{abs}(\|\mathbf{z_i} - \mathbf{z_j}\|_2 - (1 - \text{abs}(attr_{i,j}))) \tag{8}$$

Here, $\mathbf{z_i}$ and $\mathbf{z_j}$ are the latent representations of two connected nodes within the same graph (SC or FC), and $attr_{i,j}$ is the weight of the edge connecting them. This loss encourages nodes connected by stronger edges (if negative edges, considering the absolute value) to have latent representations that are closer in Euclidean space.

The KL Loss is the Kullback-Leibler divergence between the learned latent distribution of the nodes and a standard normal distribution. The general definition of KL divergence for continuous variables is:

$$D_{KL}(P\|Q) = \int P(x)\log(\frac{P(x)}{Q(x)})\,dx \tag{9}$$

It represents the information loss when using the approximate distribution Q to model the true distribution P. In our study, since both distributions are Gaussian, the KL divergence has a closed-form solution. Therefore, the KL Loss is computed as:

$$KLLoss = -\frac{1}{2}\sum(1 + \log(\sigma^2) - \mu^2 - \sigma^2) \tag{10}$$

Here, μ and σ are the mean and standard deviation of the latent distribution, and they are also used in the reparameterization formula mentioned earlier. For our model, the KL Loss acts as a regularization term introduced by the variational component.

Consequently, the total loss function is a weighted sum of these individual losses:

$$TotalLoss = w_1 \cdot ContrastiveLoss + w_2 \cdot DistanceLoss_{sc} + w_3 \cdot KLLoss_{sc} + w_4 \cdot DistanceLoss_{fc} + w_5 \cdot KLLoss_{fc} \tag{11}$$

where $w_i, i \in \{1, 2, ..., 5\}$ are hyperparameters that control the contribution of each loss component to the total loss.

2.2.3 Training

Our model consists of three GCN layers with dimensions of 360, 128, and 32, respectively. We incorporated Dropout layers (with a rate of p=0.2) and Batch Normalization layers. The weights for the loss components were set as follows: 1 for Contrastive Loss, 10 for SC Distance Loss, 1 for SC KL Loss, 15 for FC Distance Loss, and 1 for FC KL Loss. The higher weights assigned to the Distance Loss components reflect their greater importance compared to the Contrastive Loss. Furthermore, the FC Distance Loss was assigned a higher weight than the loss of SC, acknowledging that learning latent representations from FC is more challenging. Both the SC and FC models were optimized using the Adam optimizer with identical hyperparameters: a learning rate of 0.01 and weight decay of 0.001. The models were trained for only 2 epochs, as they converged rapidly, reaching a state where all loss components stabilized and achieved balance after just one epoch.

2.3 SFC

We define the Structure-Function Coupling (SFC) of a brain region i as the Gaussian kernel output of its latent representations from the SC and FC modalities:

$$\begin{aligned} SFC_i &= \exp(-\|\mathbf{z}_{i_{sc}} - \mathbf{z}_{i_{fc}}\|_2^2 / 2\sigma^2) \\ \sigma &= \text{median}(\|\mathbf{z}_{i_{sc}} - \mathbf{z}_{i_{fc}}\|_2^2),\ i \in \{1, 2, \ldots, 360\} \end{aligned} \tag{12}$$

This definition is consistent with the model's optimization objectives, as both the contrastive loss and distance loss are computed based on Euclidean distances. The sigma parameter for the Gaussian kernel is set as the median of the Euclidean distances between the SC and FC latent representations across all brain regions, ensuring robustness against potential outliers.

3 Results

3.1 The distribution of SFC within different brain network

We first performed 10-fold cross-validation on the model. The average SFC values from the test set of each fold were then weighted by inter-fold correlation in cross-validation. Finally, with reference to existing study [4], the 360 brain regions were categorized into 20 distinct brain networks, and the average SFC for each network was computed. Results demonstrated that SFC was strongest in the Visual network and weakest in the Orbito-affective network (Figure 3). Overall, the SFC gradually lost along the cognitive-representational hierarchies, consistent with known patterns of structure-function coupling.

Furthermore, to examine the distinctions in SFC across different brain networks, we calculated the mean SFC for each brain network across all samples and conducted separate one-way ANOVAs for each of the four fMRI runs. All runs revealed a strong main effect of brain network (`REST1_LR`: F(19, 19940) = 212.28, η^2 = 0.168; `REST1_RL`: F(19, 19940) = 233.98, η^2 = 0.182; `REST2_LR`: F(19, 19940) = 221.63, η^2 = 0.174; `REST2_RL`: F(19, 19940) = 202.12, η^2 = 0.161; all $p < 0.001$).

Considering the dependency of measurements across different runs within the same subject, we further used a mixed-effects model for analysis. After controlling for subject-level random effects, the model confirmed a significant fixed effect of brain network (all 19 comparisons $p < 0.001$, except for the right Cingulo-opercular network) and revealed significant differences across runs. Specifically, the Visual network had the highest SFC (Right Visual: β = +0.103), while the Orbito-affective network had the lowest values (Left Orbito-affective: β = -0.212). Run comparisons indicated that the `REST2_LR` run had the lowest global SFC (β = -0.043 vs. `REST1_LR`) Table 1.

These results consistently indicate that there are significant differences in SFC among different brain networks, and these differences remain stable under different data acquisition conditions.

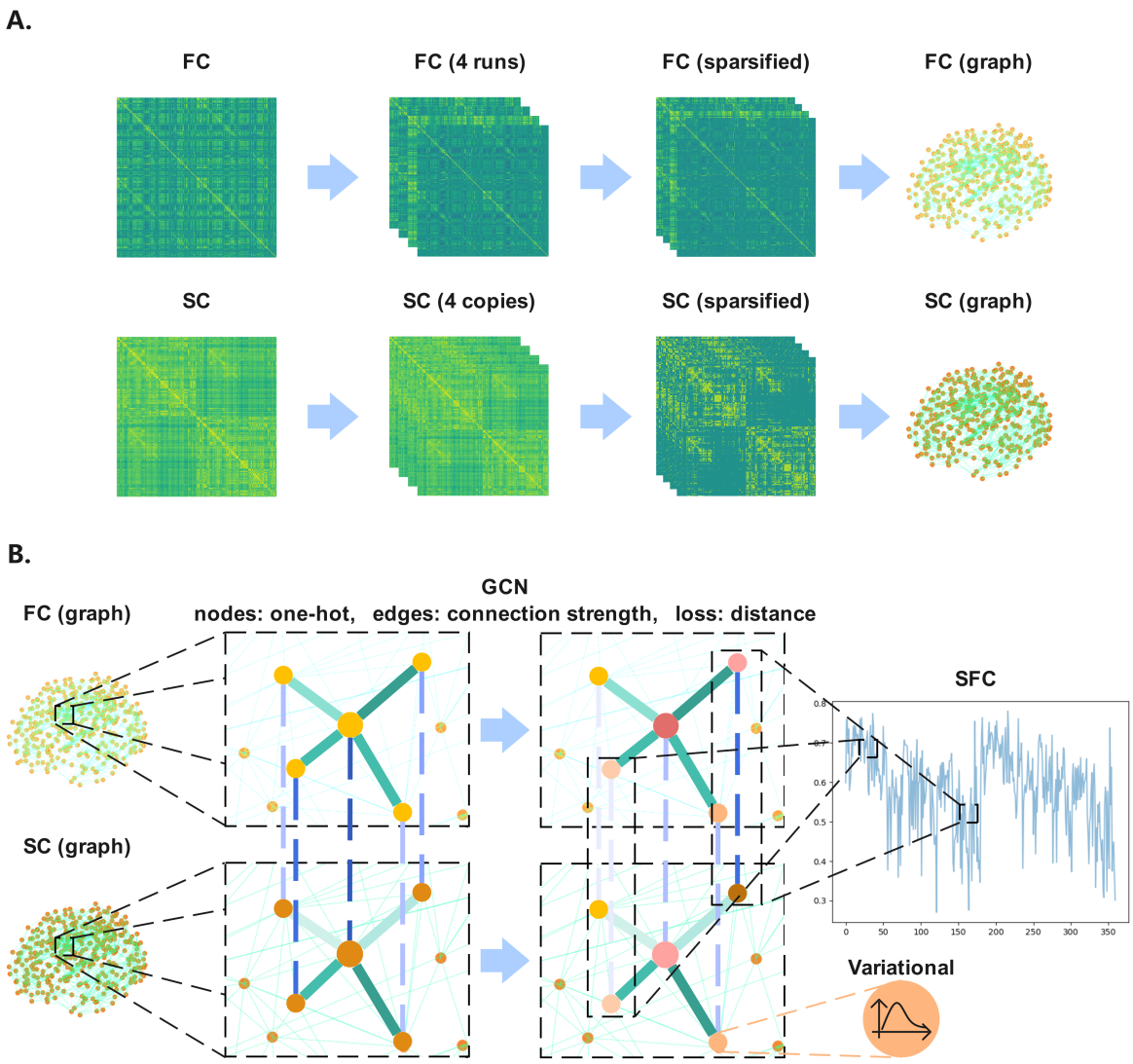

Figure 2: ***A***, The SC was replicated four times to pair with the FCs of different runs, and then they were all sparsified and represented as graphs. ***B***, During the training of the model, the Euclidean distance of the latent representations of the edge endpoints in the same graph gradually approaches their edge weights. The Euclidean distance of the latent representations of the nodes corresponding to the same brain region in SC and FC gradually decreases, eventually, the closer the distance of the brain region, the stronger the structure-function coupling.

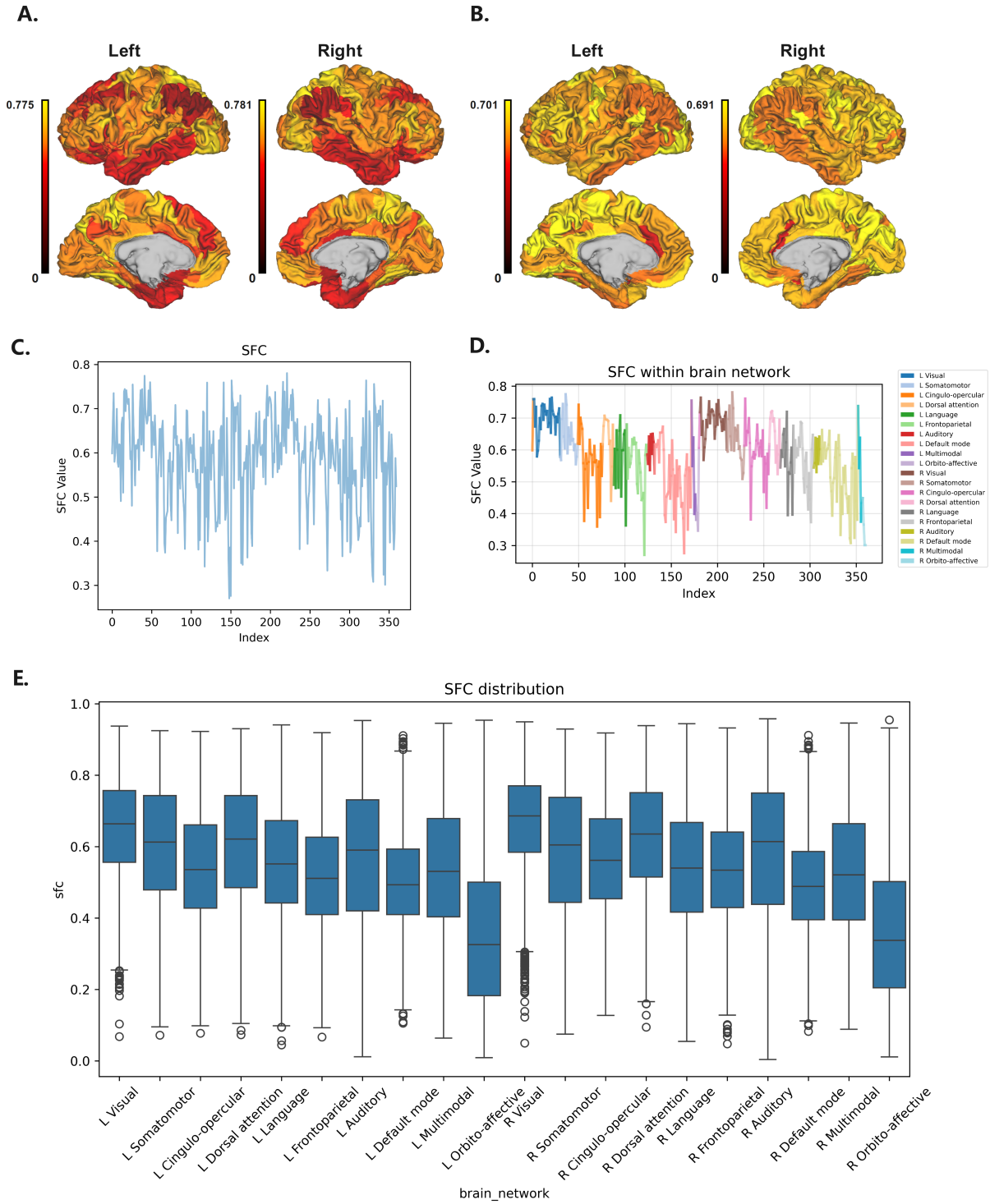

Figure 3: ***A***, The average SFC of the test set weighted by cross-validated inter-fold correlation is presented on the cerebral cortex. ***B***, Average the SFC of ***A*** according to the brain networks. ***C***, The SFC values used in ***A*** were plotted as a line graph according to the original brain region index of the HCP-MMP1.0 atlas. ***D***, Rearrange the line graph of ***C*** according to brain networks. ***E***, The box plot plotted by the average SFC of the brain networks of all individuals.

Network	REST1_LR SFC	REST1_RL SFC	REST2_LR SFC	REST2_RL SFC
L Visual	0.6672	0.6406	0.6344	0.6536
L Somatomotor	0.6293	0.5815	0.5828	0.6105
L Cingulo-opercular	0.5737	0.5271	0.5317	0.5459
L Dorsal attention	0.6361	0.5920	0.5896	0.6093
L Language	0.5803	0.5398	0.5435	0.5527
L Frontoparietal	0.5474	0.5144	0.5099	0.5179
L Auditory	0.6089	0.5197	0.5605	0.5723
L Default mode	0.5311	0.4876	0.4940	0.5049
L Multimodal	0.5689	0.5203	0.5227	0.5543
L Orbito-affective	0.3790	0.3389	0.3254	0.3692
R Visual	0.6854	0.6631	0.6526	0.6705
R Somatomotor	0.6135	0.5700	0.5605	0.6049
R Cingulo-opercular	0.5900	0.5570	0.5465	0.5777
R Dorsal attention	0.6486	0.6192	0.6040	0.6312
R Language	0.5637	0.5355	0.5214	0.5439
R Frontoparietal	0.5541	0.5379	0.5196	0.5401
R Auditory	0.6075	0.5624	0.5529	0.6073
R Default mode	0.5095	0.4918	0.4755	0.5074
R Multimodal	0.5559	0.5228	0.5046	0.5429
R Orbito-affective	0.3971	0.3378	0.3466	0.3735
Mean	0.5723	0.5330	0.5289	0.5545

Table 1: All the data in this table are output by the mixed-RUN model. After averaging the SFC of all samples in the brain networks, they are grouped by RUN.

3.2 The differences in SFC between different genders

Based on the cross-validation results, we further analyzed sex differences in SFC. In the dataset we utilized, males generally exhibited higher SFC compared to females Table 2, which aligns with known patterns of structure-function coupling in young adults [12, 13, 14, 15, 16].

To examine the specific effect of sex on this SFC metric, we constructed a mixed-effects model incorporating brain network, sex, fMRI run, and their interaction terms as fixed effects, with subject included as a random effect. The model revealed a significant main effect of sex ($\beta = 0.052$, $p < 0.001$), and indicating that overall SFC was higher in males than in females. Analysis of the brain network × sex interaction further uncovered the network specificity of these sex differences. Specifically, the male advantage was significantly stronger than the overall level in the left Cingulo-opercular network, right Language network, and right Multimodal network (all $p < 0.05$). Conversely, in the right Visual network, the male advantage was significantly weaker than the overall level ($p = 0.001$). The left Language network and left Multimodal network showed trends toward a stronger male advantage that approached marginal significance ($p = 0.074$ and $p = 0.052$, respectively).

These results indicate that while a general male advantage in SFC exists, this advantage is more pronounced in brain networks involved in cognitive control (Cingulo-opercular network), language processing, and multimodal information integration, and is relatively attenuated in the visual processing network.

3.3 Ablation study

To verify the effectiveness of each component, we conducted the following ablation study.

3.3.1 Without variational parts

After removing the variational components from the model, the test set loss showed little change. However, the distribution pattern of SFC across different brain networks was altered Figure 4, and the differences in SFC between networks were reduced (std = 0.1055 vs. std = 0.0799). This indicates that the variational design aids in learning more distinct SFC patterns.

Network	Male SFC	Female SFC	Rank
L Visual	0.6720	0.6287	19
L Somatomotor	0.6297	0.5759	11
L Cingulo-opercular	0.5814	0.5123	1
L Dorsal attention	0.6330	0.5838	16
L Language	0.5869	0.5253	5
L Frontoparietal	0.5517	0.4967	9
L Auditory	0.5928	0.5413	12
L Default mode	0.5286	0.4832	17
L Multimodal	0.5749	0.5124	4
L Orbito-affective	0.3794	0.3301	15
R Visual	0.6858	0.6523	20
R Somatomotor	0.6174	0.5608	7
R Cingulo-opercular	0.5992	0.5403	6
R Dorsal attention	0.6489	0.6055	18
R Language	0.5762	0.5104	2
R Frontoparietal	0.5654	0.5139	13
R Auditory	0.6120	0.5567	8
R Default mode	0.5231	0.4723	14
R Multimodal	0.5657	0.5016	3
R Orbito-affective	0.3925	0.3385	10

Table 2: All individuals obtained from cross-validation were averaged by brain network, and then grouped by gender and averaged to obtain this table. The second and third columns represent the average SFC of each brain network for males and females, respectively. It can be seen that in this dataset, the averaged SFC of all networks in males is higher than that in females. After sorting the differences, we obtained the ranking in the fourth column, which shows that the L Cingulo-opercular network has the largest difference and the R Visual network has the smallest difference.

3.3.2 Without contrastive loss

After setting the weight of the contrastive loss to zero, the contrastive loss itself decreased as the distance losses decreased. However, we observed that the SFC output by the model in this scenario was inconsistent with the known distribution pattern (e.g., lower in the Visual network) and showed less variation across different brain regions (std = 0.0812) Figure 4. This occurs because the SC and FC models are optimized separately without aligned latent spaces when the contrastive loss is absent. Therefore, the contrastive loss is necessary for the current model.

3.3.3 Cosine similarity

When cosine similarity was used to replace Euclidean distance, the model was still able to converge normally. The SFC pattern learned using cosine similarity was generally similar to the results based on Euclidean distance, but the former is generally higher than the latter Table 3. Theoretically, Euclidean distance imposes a stronger constraint on the latent space. Therefore, we consider the results obtained using Euclidean distance to be potentially more reliable.

3.4 Single run

As previously mentioned, the operation of combining different runs resembles a form of data augmentation rather than pseudoreplication. This is due to the presence of the contrastive loss in the model, which ensures that even the same SC input contributes differently to the SC model's loss when paired with different FC runs. To further substantiate this point, we conducted experiments using individual runs and performed a two-way mixed-effects intraclass correlation coefficient (ICC(3,1)) analysis. The results demonstrated that while the performance of individual runs during cross-validation was less stable compared to that of combined runs Figure 5, the resulting SFC distribution patterns were largely consistent with the latter (ICC = 0.844, 95% CI [0.82, 0.87], $p < 0.001$). This confirms the stability of the SFC distribution pattern learned by the model and validates the effectiveness of combining multiple runs as a data augmentation strategy.

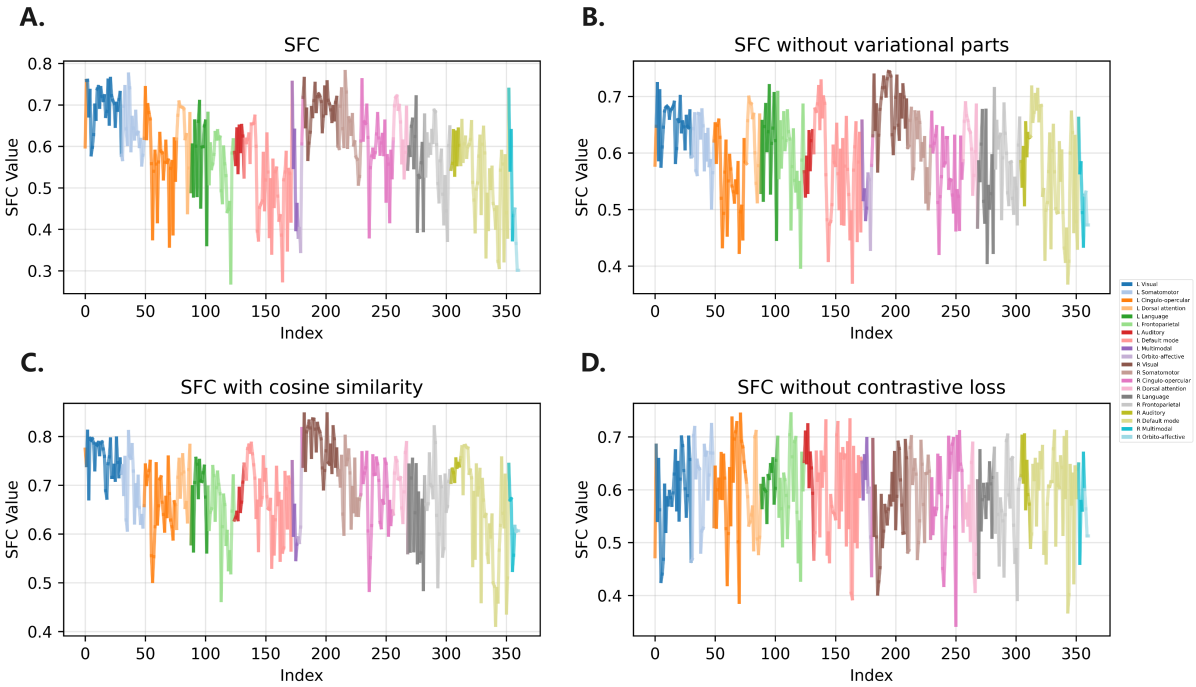

Figure 4: All four models tested in the ablation experiment underwent 10-fold cross-validation, and the average SFC of the test set were all weighted according to the inter-fold correlation of cross-validation. Finally, they were all rearranged by brain networks to obtain these four line graphs. ***A***, The model without removing or replacing any components; ***B***, The model without variational parts (including the variational latent space and KL loss); ***C***, The Euclidean distances of both the contrastive loss and the distance loss are replaced by cosine similarity, and the Gaussian kernel for output was removed, directly using cosine similarity as SFC; ***D***, The model without contrastive loss.

3.5 Distance matrix

Furthermore, we analyzed the distance matrices of the latent representations from SC and FC (Each element of the distance matrix represents the Euclidean distance between the latent representations of the row and column brain regions). Due to the presence of the distance loss, these distance matrices are expected to approximate their corresponding connectivity matrices. We aggregated all the SC and FC distance matrices from the test sets of the 10-fold cross-validation, weighting them by correlation to obtain average distance matrices. It can be observed that these average distance matrices closely resemble the original SC and FC connectivity matrices, demonstrating that the model successfully learned meaningful information. By examining the difference matrices, we found that, compared to the SC model, the FC model learned less effectively, particularly within the Visual, Somatomotor, and Cingulo-opercular networks Figure 6.

4 Discussion

4.1 How to understand this model

Our model aims to learn a latent space capable of distinguishing the SFC of different brain regions. During model training, this space is primarily constrained by the distance loss, which pulls brain regions with stronger connections closer together in the latent representation. Consequently, the latent spaces generated by the SC and FC models are respectively imbued with information from the true structural and functional connectivity patterns. Subsequently, under the constraint of the contrastive loss, the SC and FC latent spaces are aligned as closely as possible. For a given brain region, the better the alignment between the SC and FC of its connected regions, the more likely the region itself is to be well-aligned. Therefore, the model can potentially differentiate brain regions based on their varying degrees of SFC strength.

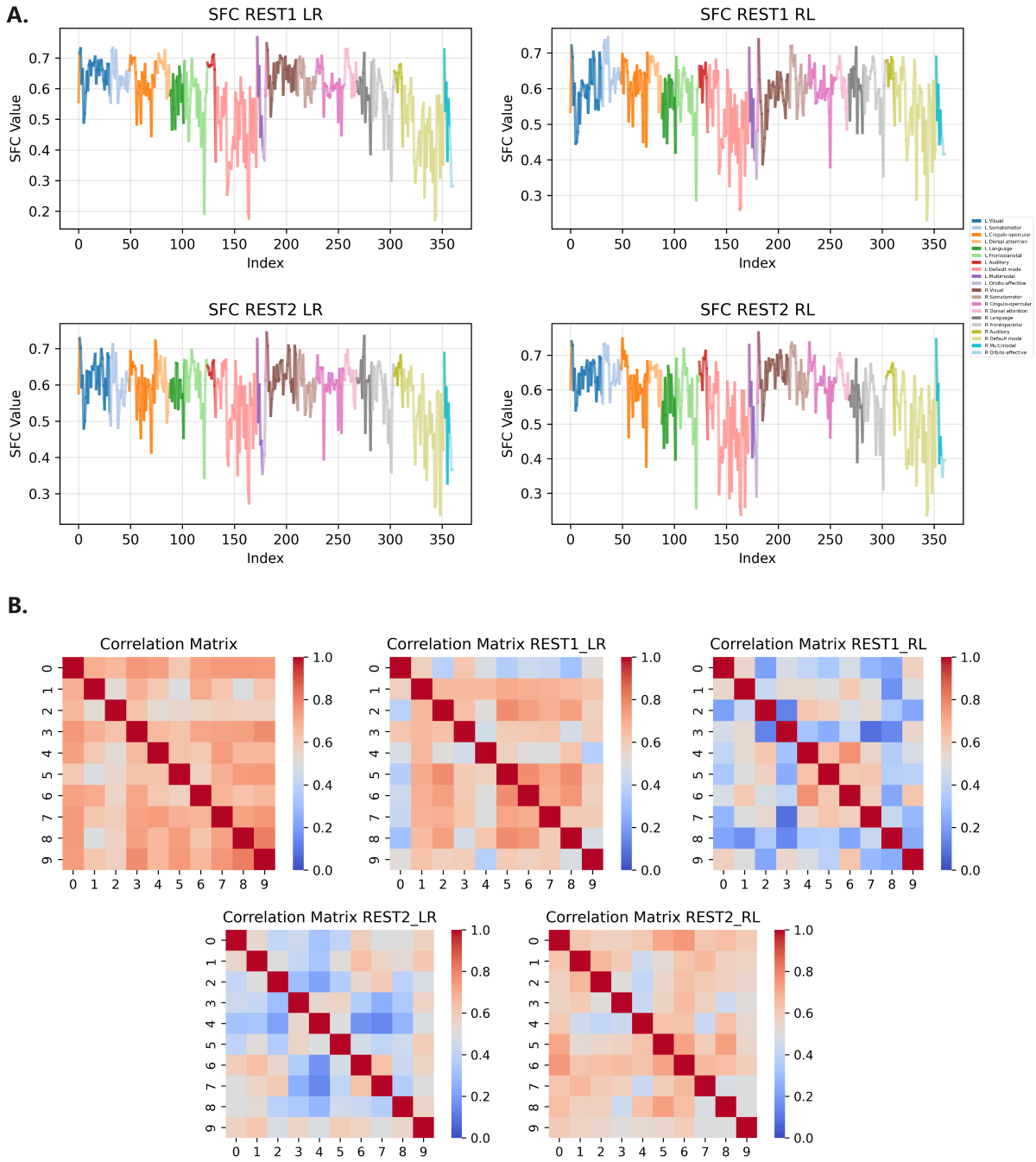

Figure 5: ***A***, The average SFC obtained by weighting the cross-validated inter-fold correlation of the models trained on four single-run datasets. It can be seen that the distribution of SFC is generally similar. ***B***, The cross-validated inter-fold correlation matrices of the mixed-run model and the four single-run models respectively. It shows that the mixed-RUN model is more stable among different folds of cross-validation.

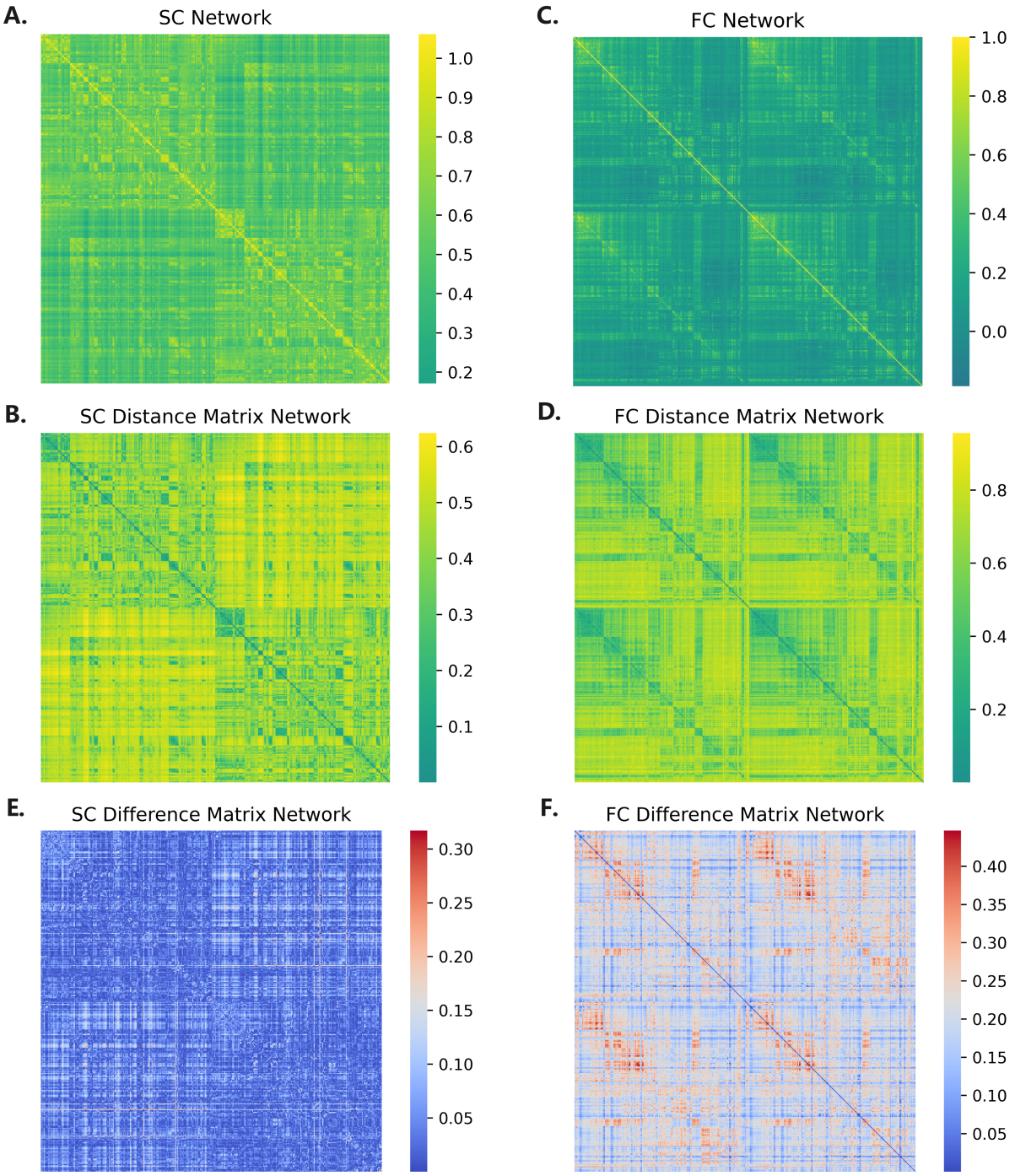

Figure 6: For ***A*** and ***C***, the average connectivity matrix of all unsparse samples; for ***B*** and ***D***, the weighted average distance matrix of all samples; for ***E*** and ***F***, the difference matrix between the average connectivity matrix and the average distance matrix, the higher the value, the greater the gap between the distance matrix and the connectivity matrix, and also indicates that the model has not learned well. All matrices are rearranged according to the brain networks. It can be found that the values of the distance matrix and the connectivity matrix are almost opposite, because the model assumes that the stronger the connection between the brain regions, the closer they are in the latent space.

Network	Distance SFC	Cosine similarity SFC	Diff.
L Visual	0.7007	0.7555	+0.0548
L Somatomotor	0.6405	0.6950	+0.0545
L Cingulo-opercular	0.5661	0.6634	+0.0973
L Dorsal attention	0.6407	0.7023	+0.0615
L Language	0.5721	0.6851	+0.1129
L Frontoparietal	0.5490	0.6397	+0.0907
L Auditory	0.6040	0.6560	+0.0520
L Default mode	0.5127	0.6773	+0.1646
L Multimodal	0.5604	0.6262	+0.0658
L Orbito-affective	0.3764	0.5962	+0.2198
R Visual	0.6913	0.7848	+0.0934
R Somatomotor	0.6322	0.6984	+0.0661
R Cingulo-opercular	0.5932	0.6845	+0.0913
R Dorsal attention	0.6483	0.7109	+0.0625
R Language	0.5672	0.6266	+0.0593
R Frontoparietal	0.5657	0.6758	+0.1101
R Auditory	0.5882	0.7310	+0.1427
R Default mode	0.5134	0.6354	+0.1219
R Multimodal	0.5452	0.6334	+0.0881
R Orbito-affective	0.3719	0.6091	+0.2372

Table 3: The data in the second column is obtained by weighting the average SFC of the model without removed or replaced components on each fold test set in cross-validation according to the inter-fold correlation. The data in the third column is calculated in the same way for the model with Euclidean distance replaced by cosine similarity. The fourth column is the difference between the second column and the third column. It can be seen that the SFC obtained by using cosine similarity is generally higher than that obtained by using Euclidean distance.

4.2 The division of brain networks

Since the SC data was sourced from the study *A Whole-Cortex Probabilistic Diffusion Tractography Connectome* [4], which already provided a brain network parcellation, we adopted its partitioning scheme in this study. However, it is noteworthy that this scheme assigns the V1 region (the first region in the HCP-MMP 1.0 atlas) to the Cingulo-opercular network instead of the Visual network. While we suspect this might be a clerical error by the original authors, we retained this assignment because our calculated average SFC for the brain networks also indicated that the SFC strength of the V1 region was closer to the average level of the Cingulo-opercular network.

5 Acknowledgments

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

[1] Panagiotis Fotiadis, Linden Parkes, Kathryn A. Davis, Theodore D. Satterthwaite, Russell T. Shinohara, and Dani S. Bassett. Structure–function coupling in macroscale human brain networks. *Nature Reviews Neuroscience*, 25:688–704, 2024.

[2] Peiyu Chen, Hang Yang, Xin Zheng, Hai Jia, Jiachang Hao, Xiaoyu Xu, Chao Li, Xiaosong He, Runsen Chen, Tatsuo S. Okubo, and Zaixu Cui. Group-common and individual-specific effects of structure–function coupling in human brain networks with graph neural networks. *Imaging Neuroscience*, 2:imag–2–00378, 2024.

[3] T. Sarwar, Y. Tian, B.T.T. Yeo, K. Ramamohanarao, and A. Zalesky. Structure-function coupling in the human connectome: A machine learning approach. *NeuroImage*, 226:117609, 2021.

[4] B.Q. Rosen and E. Halgren. A whole-cortex probabilistic diffusion tractography connectome. *eNeuro*, 8(1):ENEURO.0416–20.2020, 2021.

[5] David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, and for the WU-Minn HCP Consortium. The wu-minn human connectome project: An overview. *NeuroImage*, 80:62–79, 2013.

[6] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, and S.M. Smith. Fsl. *NeuroImage*, 62(2):782–790, 2012.

[7] M. Glasser, T. Coalson, E. Robinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536:171–178, 2016.

[8] D.S. Marcus, J. Harwell, T. Olsen, M. Hodge, M.F. Glasser, F. Prior, M. Jenkinson, T. Laumann, S.W. Curtiss, and D.C. Van Essen. Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics*, 5:4, 2011.

[9] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[10] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes, 2022.

[11] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019.

[12] B. Vázquez-Rodríguez, L.E. Suárez, R.D. Markello, G. Shafiei, C. Paquola, P. Hagmann, M.P. van den Heuvel, B.C. Bernhardt, R.N. Spreng, and B. Misic. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences*, 116(42):21219–21227, 2019.

[13] M.G. Preti and D. Van De Ville. Decoupling of brain function from structure reveals regional behavioral specialization in humans. *Nature Communications*, 10:4747, 2019.

[14] N. Luo, J. Sui, A. Abrol, J. Chen, J.A. Turner, E. Damaraju, Z. Fu, L. Fan, D. Lin, C. Zhuo, Y. Xu, D.C. Glahn, A.L. Rodrigue, M.T. Banich, G.D. Pearlson, and V.D. Calhoun. Structural brain architectures match intrinsic functional networks and vary across domains: A study from 15000+ individuals. *Cerebral Cortex*, 30(10):5460–5470, 2020.

[15] P. Fotiadis, M. Cieslak, X. He, et al. Myelination and excitation-inhibition balance synergistically shape structure-function coupling across the human cortex. *Nature Communications*, 14:6115, 2023.

[16] Z. Gu, K.W. Jamison, M.R. Sabuncu, and A. Kuceyeski. Heritability and interindividual variability of regional structure-function coupling. *Nature Communications*, 12:4894, 2021.