BiSCoT: Behavior-Informed Subgroup-Consistent Connectome Template for Interpretable Brain Network Analysis

Zijian Chen[1], Stefen Beeler-Duden[3], Sophie Lawson[5], Zachary Jacokes[4], John Darrell Van Horn[3,4], Kevin A. Pelphrey[2], and Archana Venkataraman[1]

[1] Department of Electrical and Computer Engineering, Boston University
{zijianc, archanav}@bu.edu
[2] Department of Neurology, University of Virginia
KAP3N@uvahealth.org
[3] Department of Psychology, University of Virginia
{sjb3px, jdv7g}@virginia.edu
[4] School of Data Science, University of Virginia
zj6nw@virginia.edu
[5] Center for Healthy Brain Development, University of Virginia
pkv4hv@virginia.edu

Abstract. We propose a graph information compression framework, called **B**ehavior-**I**nformed **S**ubgroup-consistent **Co**nnectome **T**emplate (BISCoT), that learns interpretable functional subnetworks from resting-state fMRI (rs-fMRI) connectivity, which simultaneously capture the heterogeneity of a diverse patient cohort. BISCoT uses multidimensional behavioral profiles to guide the decomposition of a rs-fMRI connectivity matrices into sparse yet representative subnetworks that are consistent within behavioral sub-groups. In particular, our framework adopts a graph convolution network to capture local connectivity features and applies a subgroup-informed pooling process to extract the final subnetworks. We evaluate BISCoT on an in-house dataset of individuals with autism spectrum disorder and demonstrate that the learned subnetworks *improve the performance of multiple downstream prediction tasks*. In addition, BISCoT extracts consistent connectivity "templates" at the subgroup level, which allows for interpretable biomarker identification [6].

Keywords: Resting-state fMRI · Multimodal Fusion · Graph Neural Networks · Neurobehavioral Signature · Adaptive Connectomics

1 Introduction

Resting-state functional MRI (rs-fMRI) is a non-invasive imaging modality that captures steady-state patterns of co-activation in the brain. Mathematically, these connectivity patterns can be represented as *functional brain networks*,

[6] Code available at https://github.com/zijianch/biscot.

where each node corresponds to a region of interest (ROI) and the edges represent the correlation between them. Alterations in the network organization have been associated with various psychiatric disorders like autism spectrum disorder (ASD) [27] and schizophrenia [17]. Consequently, rs-fMRI begins to be utilized experimentally for diagnostic and prognostic evaluations and in providing a more nuanced picture of behavioral subtypes that exist within a patient cohort.

Robust detection of network alterations in rs-fMRI connectivity remains an open challenge due to high data dimensionality and the subtle but distributed changes associated with many psychiatric disorders. Traditional rs-fMRI studies have focused on statistical approaches for group-wise comparison [28]. More recently, the focus has shifted to predictive analytics, with machine learning being a natural tool to mine complex relationships in the rs-fMRI data for case/control prediction [22, 26]. Going one step further, graph neural networks (GNNs) have been proposed as an end-to-end way to directly model the brain networks. Specifically, the feature updates in GNNs are restricted by the underlying graph structure [11], thus preserving its inherent topology. GNNs have been widely applied to the study of brain disorders. For example, the work of [4] proposed a structurally-regularized edge convolution to predict the behavioral scores of ASD patients, while the work of [32] included both 1-hop and 2-hop neighboring information for ADHD prediction. More recent works include the early diagnosis of Alzheimer's disease [31] and the post-stroke aphasia [1].

While these GNNs demonstrate high classification accuracies, they are difficult to interpret, as most conclusions are derived from the network convolutional weights or the intermediate feature maps, both of which can be noisy and lack robustness. A straightforward solution is to sparsify the full graph to retain only the most salient structures. Common techniques in this vein include traditional pooling [9, 15, 3] and sparse attention/regularization mechanisms [30, 16]. While these methods improve interpretability, they do not naturally account for the heterogeneous phenotypes often present in patient cohorts. As a workaround, recent studies have introduced feature-attentive GNNs [19, 5] to integrate behavioral and demographic information to the graph analysis. While seminal, the learned subnetworks are highly inconsistent across subjects.

In this paper we propose a novel graph information compression framework, **B**ehavior-**I**nformed **S**ubgroup-consistent **Co**nnectome **T**emplate (BISCoT), that uses multidimensional behavioral data to enforce subgroup consistency in its learned sparse representation. Specifically, we use a convolution module to aggregate information from neighboring nodes in the brain network. Based on these learned features, we propose a subgroup-informed pooling mechanism that uses the embeddings of each subject's behavioral profile to guide the selection of edges and nodes in a way that remains consistent within each subgroup. As a result, our method preserves subject-specific information while identifying consistent sub-networks aligned with distinct behavioral phenotypes. Using an in-house dataset, we demonstrate that our framework both preserves the network information in the original rs-fMRI data and learns a generalized and interpretable representation that outperforms baseline models in multiple downstream tasks.

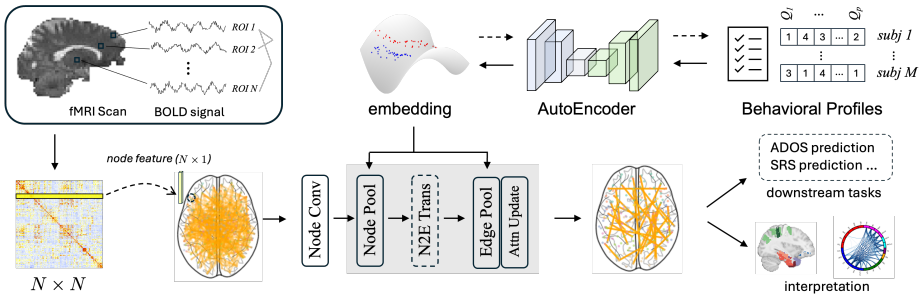

Fig. 1. Schematic of the proposed framework. **Top Left:** Computation of correlation matrix from the fMRI BOLD signal. **Top Right:** Behavioral profile embedding. **Bottom**: (Behavioral) Subgroup-informed template learning.

2 Learning a Subgroup-Informed Connectivity Template

Our proposed BISCoT framework (Fig. 1) takes as input the rs-fMRI connectivity matrix and behavioral profile. Formally, let N be the number of ROIs in the common brain parcellation, and M be the number of subjects in the cohort. $\mathbf{X}_m^{(0)} \in \mathbb{R}^{N\times N}$ is the rs-fMRI connectivity matrix for subject m, and $\mathbf{p}_m \in \mathbb{R}^{p\times 1}$ is the corresponding behavioral profile. The profiles define subgroups of subjects based on specific characteristics, for example, sensory processing in ASD or symptom severity in schizophrenia. The output of BISCoT is a sparse representation $\mathbf{X}_m \in \mathbb{R}^{N\times N}$ for each $\mathbf{X}_m^{(0)}$ that is both consistent within the same behavioral subgroup and can be used for multiple downstream predictive tasks.

2.1 The BISCoT Framework

Profile Embedding: We first map the behavioral profiles $\mathbf{p}_m$ into a low-dimensional embedding $\mathbf{z}_m \in \mathbb{R}^{d_e\times 1}$. Since $\mathbf{p}_m$ can be derived from clinical, observational, and survey data with both continuous and categorical values, we use the embedding $\mathbf{z}_m$ to guide the rs-fMRI analysis. As seen in Fig. 1, we use an autoencoder to construct the profile embeddings. The encoder maps $\mathbf{p}_m$ to $\mathbf{z}_m$, while a decoder reconstructs $\mathbf{p}'_m$ from $\mathbf{z}_m$. We train the autoencoder using a combination of two loss terms. The first term minimizes the reconstruction gap:

$$\mathcal{L}_{\text{dec}} = \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{p}_m - \mathbf{p}'_m\|^2 . \tag{1}$$

The second term encourages the latent embeddings of subjects from different behavioral subgroups to be different via a supervised contrastive loss [13]:

$$\mathcal{L}_{\text{sc}} = \frac{1}{M}\sum_{m=1}^{M} \frac{-1}{|\mathcal{P}_i|}\sum_{p\in\mathcal{P}_m} \log\left(\frac{\exp\left(\mathbf{z}_m\cdot\mathbf{z}_p/\tau\right)}{\sum_{a\neq m}\exp\left(\mathbf{z}_m\cdot\mathbf{z}_a/\tau\right)}\right), \tag{2}$$

where τ is the temperature parameter. Minimizing $\mathcal{L}_{\text{sc}}$ equivalently minimizes the intra-subgroup distance while maximizing the inter-subgroup distance.

Graph Convolution: We use the graph invariant network (GIN) model [29] to map the edge-wise rs-fMRI connectivity matrix onto node embeddings:

$$\mathbf{X}_{i\cdot}^{(\ell)} = \text{MLP}^{(\ell)}\left(\left(1+\epsilon^{(\ell)}\right)\cdot \mathbf{X}_{i\cdot}^{(\ell-1)} + \sum_{j\in\mathcal{N}^*(i)} \mathbf{X}_{j\cdot}^{(\ell-1)}\right) \tag{3}$$

Here, $\mathbf{X}_{i\cdot}^{(\ell)}$ is the feature vector of node i at layer ℓ, the set $\mathcal{N}^*(i)$ denotes the neighboring nodes of i defined by rs-fMRI connectivity > 0.6, and ϵ is a learned parameter. For notational simplicity, we have dropped the subject index m.

Subgroup-Informed Pooling: BISCoT uses two types of pooling to learn behaviorally-informed connectivity templates. The first is node pooling, which identifies and preserves an informative set of brain ROIs. The second is edge pooling, which sparsifies the subgraph created from just the selected nodes.

The first step of node pooling is to learn an "importance score" $\mathbf{f}_i$ for each node (i.e., brain ROI) i that is informed by the subject profile embeddings $\mathbf{z}$:

$$\mathbf{f}_i(\mathbf{X}_{i\cdot}^{(L)} \mid \mathbf{z}) = \phi(\mathbf{w}(\mathbf{z})^\top \mathbf{X}_{i\cdot}^{(L)} + b), \tag{4}$$

where $\mathbf{X}_{i\cdot}^{(L)} \in \mathbb{R}^{d_L\times 1}$ is the output of the last GIN layer in Eq. (3), $\phi(\cdot)$ is a sigmoid activation function to ensure $\mathbf{f}_i \in [0,1]$, and b is a learnable bias.

We constrain the learnable weights $\mathbf{w}(\cdot) \in \mathbb{R}^{d_L\times 1}$ in Eq. (4) to be a parameterized function of the profile embedding $\mathbf{z}$ via the two-layer MLP as follows:

$$\mathbf{w}(\mathbf{z}) = \Theta_2 \cdot \text{Softmax}(\Theta_1\mathbf{z} + \mathbf{b}_1) + \mathbf{b}_2, \tag{5}$$

where $\Theta_1 \in \mathbb{R}^{C\times d_e}$, $\Theta_2 \in \mathbb{R}^{d_L\times C}$ are learnable parameters shared across all nodes and subjects, and $\mathbf{b}_1, \mathbf{b}_2$ are the bias terms. Notice that $\text{Softmax}(\Theta_1\mathbf{z}+\mathbf{b}_1)$ acts as the mixing probability of belonging to each behavioral subgroup. As a result, subjects with similar profiles receive similar weights and therefore similar scores. The scores $\mathbf{f}_i$ are sorted in descending order with the top-k nodes selected as the most important. For stability, we apply a dynamic dropout to the removed nodes by randomly retaining them with a probability derived from how frequently each one is kept for other subjects in the same behavioral subgroup.

These selected node features $\mathbf{X}_{i\cdot}$ are then passed into a node-to-edge transformation block to generate a feature vector $\widetilde{\mathbf{X}}_{ij} \in \mathbb{R}^{d_e\times 1}$ for each edge (i,j) based on its two endpoints. Formally, this is done by concatenating $[\mathbf{X}_{i\cdot}, \mathbf{X}_{j\cdot}]^\top$ and passing this new vector through an MLP with learnable weights.

The edge pooling operation follows the same sequence of steps as node pooling and is summarized briefly. The profile-guided importance score for edge (i,j) is computed as $\mathbf{F}_{ij}(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{z}) = \phi(\widetilde{\mathbf{w}}(\mathbf{z})^\top \widetilde{\mathbf{X}}_{ij} + \widetilde{b})$, with the learnable weight vector $\widetilde{\mathbf{w}}(\mathbf{z}) = \widetilde{\Theta}_2 \cdot \text{Softmax}(\widetilde{\Theta}_1\mathbf{z} + \widetilde{\mathbf{b}}_1) + \widetilde{\mathbf{b}}_2$. In addition to the top-k edge selection

and dynamic dropouts, we redistribute the information from the removed edges (i, j) to the retained ones (u, v) using an attention mechanism, defined as

$$\widetilde{\mathbf{X}}'_{uv} = \mathbf{F}_{uv}\widetilde{\mathbf{X}}^{(\ell)}_{uv} + \alpha\widetilde{\mathbf{X}}^{(\ell)}_{ij}, \tag{6}$$

where the attention score α is computed as follows:

$$\alpha = \text{Softmax}\left(\frac{(W_Q\mathbf{g}_{ij})\,(W_K\mathbf{g}_{uv})^\top}{\sqrt{d}}\right)W_V. \tag{7}$$

Here, W_Q, W_K, and W_V are learnable parameters, and the inputs $\mathbf{g}_{ij}$ and $\mathbf{g}_{uv}$ capture geometric information about the edges (i, j) and (u, v), respectively,

Loss Functions: Similar to the profile embeddings, we minimize a combination of two loss terms. The first term is a reconstruction loss that guarantees a sufficient amount of information is retained in the template $\widetilde{\mathbf{X}}_{ij}$. It is computed as the geodesic distance on the manifold of symmetric positive-definite matrices:

$$\mathcal{L}_{\text{recon}} = \frac{1}{M}\sum_{m=1}^{M} \left\| \log\text{m}(\widehat{\mathbf{X}}^{-1/2}\mathbf{X}^{(0)}\widehat{\mathbf{X}}^{-1/2})\right\|_F^2, \tag{8}$$

where $\widehat{\mathbf{X}}$ is produced by a small reconstruction head, implemented as two convolution layers, on top of the learned representation $\widetilde{\mathbf{X}}_{ij}$. The second term strengthens the subgroup separation introduced by the profile embedding, defined as the cross-entropy between the predicted subgroup $\hat{y}_{m,c}$ and the ground truth:

$$\mathcal{L}_{\text{sep}} = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c=1}^{C} y_{m,c}\log(\hat{y}_{m,c}). \tag{9}$$

Taken together, the overall loss for training the connectivity portion of BISCoT is the sum of Eqs. (8)-(9): $\mathcal{L} = \mathcal{L}_{\text{sep}} + \lambda_1\mathcal{L}_{\text{recon}}$, where λ_1 is a hyperparameter.

2.2 Implementation Details

Optimization: We separately train the profile embeddings and the rs-fMRI connectivity template. The profile embedding module is trained by minimizing $\mathcal{L}_{\text{dec}} + \mathcal{L}_{\text{sc}}$ as defined in Eq. (1) and Eq. (2). It is then frozen during the template learning. The dimension of the profile embedding is set to $d_e = 16$, the temperature to $\tau = 0.2$, and the hidden dimensions for the AutoEncoder to $128, 64$. For the rs-fMRI module, the dimension of the node convolution is 16. The pooling ratio and maximum dropout rate are set to a commonly used value of 0.5 and are not tuned based on performance. The reconstruction layer for Eq. (8) is a convolution with 64 hidden channels, and the classification layer for Eq. (9) includes a global average and max pooling layer followed by an MLP with a hidden dimension of 32. The reconstruction head consists of a convolution layer with 64 hidden channels. We set the hyperparameter λ_1 to 0.03 to balance the scales of the two losses.

The model is trained using Adam with a learning rate of 10^{-3} decayed by a factor of 0.95 every 10 epochs, and a batch size of 8. Additionally, a weight decay of 10^{-3} and early stopping is applied to prevent overfitting.

Evaluation Strategy: We benchmark the learned template $\tilde{\mathbf{X}}$ from BISCoT against ablation models that use different convolution operations (GIN [29], GraphSAGE [12]) and pooling methods (TopK [9], SAGPool [15], EdgePool [3]). We quantify performance via the geodesic and Frobenius distances between the input and learned connectivity matrices, the subgroup separation accuracy, and the difference in the largest eigenvalues of the original and learned matrices.

We also evaluate the utility of the learned template $\tilde{\mathbf{X}}$ on multiple downstream predictive tasks (see Section 3 for details). The downstream tasks operate on the same splits and take the learned subgraph $\widetilde{\mathbf{X}}$ as input. It includes a global average pooling layer followed by an MLP with hidden dimensions of 64 and 32. This multistage procedure isolates the subgraph information from any task-specific tuning to demonstrate generalizability. Beyond the above ablation models, we compare BISCoT with three recent end-to-end models for brain network analysis (BrainGNN [4], MGCN [32], BPI-GNN [33]) that directly map the connectivity input $\mathbf{X}^{(0)}$ to the predictive output.

3 Experiments

3.1 Dataset and Input Construction

Our in-house dataset contains 105 participants diagnosed with ASD (58 female; age = 14.21 ± 2.06 years). FMRI data were acquired using either a Siemens 3T Trio scanner with a 12-channel head coil or a Siemens 3T Prisma scanner with a 20-channel head coil. The data were preprocessed using the standard fMRIprep pipeline [6] and are used to compute the connectivity matrix $\mathbf{X}_m^{(0)}$, with ROIs defined according to the Brainnetome parcellation ($N = 246$) [8]. The geometric information $\mathbf{g}_{ij} \in \mathbb{R}^3$ in (7) is the midpoint coordinate of the end-nodes i and j.

The behavioral profiles $\mathbf{p}_m \in \mathbb{R}^{60}$ are the responses of the Sensory Profile 2 for Adolescents and Adults [2]. It queries a broad spectrum of sensory-related behaviors and are essential for demonstrating the utility of our method. This questionnaire consists of 60 statements that probe different sensory experiences. These responses are grouped into six sensory-type sub-scores (auditory, taste, etc.; range: 0–50 each) and four sensory-pattern sub-scores (low registration, sensory avoiding, etc.). We focus on the "Low Registration" experience, where higher scores indicate reduced perception and response to sensory input. Using this sub-score, we define three behavioral subgroups (high/normal/low registration) for Eq. (2) and Eq. (9).

We consider three downstream prediction tasks. The first task is to predict the vector of six sub-scores for the different sensory types. The second task is to predict the Autism Diagnostic Observation Schedule (ADOS) cumulative score for each patient (range: 0-30) [20]. The third task is to predict the Social Responsiveness Scale (SRS) total score (range: 0-100) [20].

Model	Geodesic ↓	Frobenius ↓	Separation ↑	λ_1 Deviation ↓
BISCoT	**112.48±9.28**	**65.55±4.04**	**0.77±0.12**	**0.35±0.07**
GIN+TopK	181.09 ± 2.97*	65.64 ± 4.62	0.59 ± 0.07*	0.37 ± 0.09
GIN+SAGPool	180.55 ± 0.98*	67.94 ± 5.79*	0.60 ± 0.11*	0.42 ± 0.13*
GIN+EdgePool	179.51 ± 2.22*	79.16 ± 3.13*	0.56 ± 0.10*	0.61 ± 0.07*
GraphSAGE+TopK	179.12 ± 1.64*	66.89 ± 4.66	0.52 ± 0.14*	0.41 ± 0.10*

Table 1. Reconstruction performances of BISCoT and baseline models. Evaluation metrics are geodesic distance (Eq. (8)), Frobenious distance, subgroup separation accuracy, and percentage deviation of the largest eigenvalue. The asterisk * indicates statistically worse performance ($p < 0.05$) when compared with BISCoT.

3.2 Experimental Results

Reconstruction Performance: Table 1 compares the reconstruction performance of BISCoT against four ablation baselines using a repeated 5-fold cross validation setup. We have reported the test set metrics to avoid information leakage during training. As seen, BISCoT achieves uniformly better performance, almost all of which is statistically significant. Notably, our model significantly reduces the geodesic distance between the input and template connectivity matrices. This trend indicates that our model is more effective at capturing the intrinsic structure of the network. The lower deviation of the largest eigenvalue, which is related to network topology, further supports this conclusion.

Downstream Prediction Tasks: We use the same cross validation setup to demonstrate how the learned representations support downstream prediction tasks. Namely, within each fold, the learned representations $\widetilde{\mathbf{X}}$ from the *training data* are used as input to train a small predictor network. We then apply both BISCoT and the downstream predictor to the *testing data* to evaluate performance. This ensures *no information leakage between training and testing data.*

Table 2 compares the performance of BISCoT and baseline models across three prediction tasks (1) sensory category, (2) ADOS, and (3) SRS. As expected, our our model shows a significant improvement in sensory category score prediction, as the embeddings are constructed from sensory behavioral profiles. However, the ADOS and SRS evaluations focus on different behavioral traits are provide a fair comparison across all methods. Even so, BISCoT outperforms the baseline models by achieving lower and more stable metric values. Overall, these results demonstrate that our model not only reconstructs brain networks more reliably but also retains more informative features for downstream tasks.

Qualitative Interpretation of the Learned Templates: We compute subgroup specific templates by aggregating ROIs and edges that are retained by node and edge pooling, respectively, in at least 85% of the sub-group patients. In contrast, standard pooling methods yield inconsistent node sets even among "similar" subjects. These templates are illustrated in Fig. 2. As seen, the high

Model	Sensory Category		ADOS		SRS	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
BISCoT	**4.53±0.48**	**5.80±0.59**	**3.58±0.83**	**4.77±0.69**	**13.43±1.29**	**16.19±0.92**
GIN+TopK	7.15±0.82*	9.20±1.26*	4.10±1.08	5.38±1.43	16.81±1.87*	20.93±2.50*
GIN+SAGPool	6.79±0.75*	9.00±0.82*	4.27±0.51*	5.36±0.81*	23.93±2.87*	27.78±3.01*
GIN+EdgePool	6.90±0.83*	9.10±0.91*	4.62±1.44*	5.69±1.77*	23.47±4.94*	27.59±0.93*
GraphSAGE+TopK	6.74±0.55*	8.63±0.76*	4.72±0.47*	5.92±0.59*	25.03±2.84*	28.91±2.89*
BrainGNN	6.25±0.78*	7.97±0.99*	3.77±0.82	4.80±0.97	16.34±1.92*	20.97±1.98*
MGCN	5.14±0.64*	6.54±0.79*	4.20±0.69	5.09±0.85	14.01±1.54	18.24±1.81
BPI-GNN	19.19±1.31*	20.71±9.21*	7.21±0.71*	8.44±3.65*	18.45±4.46*	26.54±9.89*

Table 2. Downstream tasks performances of the proposed and baseline models. Evaluation metrics are mean absolute error (MAE) and root mean squared error (RMSE). The asterisk * indicates statistically worse performance ($p < 0.05$) compared to the best performing model in highlighted in bold.

registration subgroup consistently picks regions in frontolimbic and insular circuits, which reflect heightened emotional reactivity [7] and interoceptive awareness [25]. The normal registration subgroup shows more balanced involvement of temporal-limbic regions, such as the amygdala, hippocampus, and entorhinal cortex, which indicates typical emotional and sensory responsiveness [21]. The low registration subgroup emphasizes sensorimotor and parietal regions, while showing less direct limbic or insular engagement, which suggests reduced sensitivity to emotional and interoceptive stimuli. In terms of edge connections, all three groups display consistent involvement of the fusiform gyrus (FuG), which processes body representations, integrates multisensory inputs for perception, and may support interoceptive learning [23, 10, 24]. In addition, the high registration group exhibits a higher density of edges stemming from the inferior temporal gyrus (ITG) and parahippocampal gyrus (PhG), which highly correlate to the severity of autism symptoms [14], especially the repetitive behaviors [18].

4 Conclusion

We introduced BISCoT, a novel graph information compression framework to address key challenges in analyzing resting-state fMRI connectivity for complex and heterogeneous populations like ASD. By integrating a state-of-the-art GIN architecture with a subgroup-informed pooling module, our approach learns a sparse yet discriminative representation that maintains interpretability across different behavioral characteristics. Experimental results demonstrate that our framework not only outperforms existing GNN-based baselines in reconstruction accuracy and predictive tasks but also provides insights into the neurobiological signatures of distinct sensory registration patterns in ASD. Future work will focus on extending this methodology to larger and more diverse clinical cohorts, as well as exploring more gradated behavioral subtypes.

Acknowledgments. This work was supported by the National Institutes of Health awards R01 HD108790 (PI Venkataraman) and R01 EB029977 (PI Caffo) and the National Science Foundation CAREER award 1845430 (PI Venkataraman).

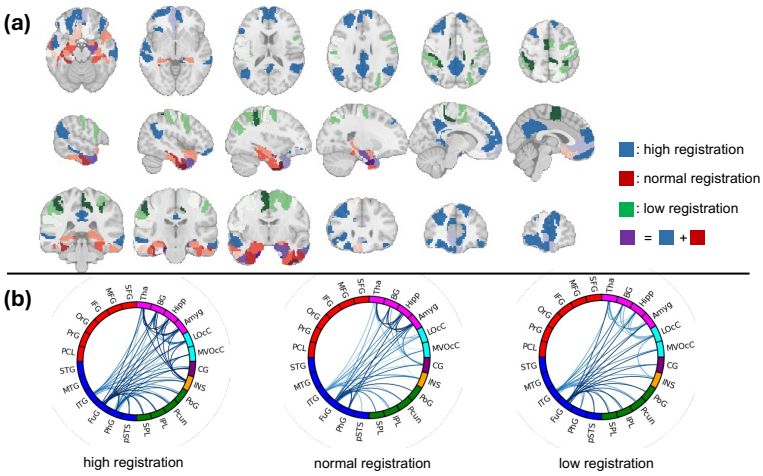

Fig. 2. Visualization of the common templates learned by the proposed model. **(a)** The most commonly selected regions **(b)** The most commonly selected connections. Darker colors indicate higher selection frequency.

Disclosure of Interests. The authors declare no conflict of interest.

References

1. Chen, Z., Varkanitsa, M., Ishwar, P., Konrad, J., Betke, M., Kiran, S., Venkataraman, A.: A lesion-aware edge-based graph neural network for predicting language ability in patients with post-stroke aphasia. In: International Workshop on Machine Learning in Clinical Neuroimaging. pp. 91–101. Springer (2025)
2. Crane, L., Goddard, L., Pring, L.: Sensory processing in adults with autism spectrum disorders. Autism **13**(3), 215–228 (2009)
3. Diehl, F., Brunner, T., Le, M.T., Knoll, A.: Towards graph pooling by edge contraction. In: ICML 2019 workshop on learning and reasoning with graph-structured data (2019)
4. Dsouza, N.S., Nebel, M.B., Crocetti, D., Robinson, J., Mostofsky, S., Venkataraman, A.: M-GCN: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations. In: Medical Imaging with Deep Learning. pp. 119–130. PMLR (2021)
5. Dvornek, N.C., Li, X., Zhuang, J., Ventola, P., Duncan, J.S.: Demographic-guided attention in recurrent neural networks for modeling neuropathophysiological heterogeneity. In: International Workshop on Machine Learning in Medical Imaging. pp. 363–372. Springer (2020)
6. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al.: fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods **16**(1), 111–116 (2019)
7. Etkin, A., Egner, T., Kalisch, R.: Emotional processing in anterior cingulate and medial prefrontal cortex. Trends in cognitive sciences **15**(2), 85–93 (2011)

8. Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al.: The human brainnetome atlas: a new brain atlas based on connectional architecture. Cerebral cortex **26**(8), 3508–3526 (2016)
9. Gao, H., Ji, S.: Graph u-nets. In: International Conference on Machine Learning. pp. 2083–2092. PMLR (2019)
10. García-Cordero, I., Sedeño, L., De La Fuente, L., Slachevsky, A., Forno, G., Klein, F., Lillo, P., Ferrari, J., Rodriguez, C., Bustin, J., et al.: Feeling, learning from and being aware of inner states: interoceptive dimensions in neurodegeneration and stroke. Philosophical Transactions of the Royal Society B: Biological Sciences **371**(1708), 20160006 (2016)
11. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: International Conference on Machine Learning. pp. 1263–1272. PMLR (2017)
12. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Advances in Neural Information Processing Systems **30** (2017)
13. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in Neural Information Processing Systems **33**, 18661–18673 (2020)
14. Kim, D., Lee, J.Y., Jeong, B.C., Ahn, J.H., Kim, J.I., Lee, E.S., Kim, H., Lee, H.J., Han, C.E.: Overconnectivity of the right Heschl's and inferior temporal gyrus correlates with symptom severity in preschoolers with autism spectrum disorder. Autism Research **14**(11), 2314–2329 (2021)
15. Lee, J., Lee, I., Kang, J.: Self-attention graph pooling. In: International Conference on Machine Learning. pp. 3734–3743. PMLR (2019)
16. Lin, C., Sun, G.J., Bulusu, K.C., Dry, J.R., Hernandez, M.: Graph neural networks including sparse interpretability. arXiv preprint arXiv:2007.00119 (2020)
17. Mitchell, R.L., Elliott, R., Woodruff, P.W.: fMRI and cognitive dysfunction in schizophrenia. Trends in cognitive sciences **5**(2), 71–81 (2001)
18. Monk, C.S., Peltier, S.J., Wiggins, J.L., Weng, S.J., Carrasco, M., Risi, S., Lord, C.: Abnormalities of intrinsic functional connectivity in autism spectrum disorders. Neuroimage **47**(2), 764–772 (2009)
19. Moon, H.S., Mahzarnia, A., Stout, J., Anderson, R.J., Han, Z.Y., Tremblay, J.T., Badea, C.T., Badea, A.: Feature attention graph neural network for estimating brain age and identifying important neural connections in mouse models of genetic risk for Alzheimer's disease. Imaging Neuroscience **2**, 1–22 (2024)
20. Payakachat, N., Tilford, J.M., Kovacs, E., Kuhlthau, K.: Autism spectrum disorders: a review of measures for clinical, health services and cost–effectiveness applications. Expert review of pharmacoeconomics & outcomes research **12**(4), 485–503 (2012)
21. Phelps, E.A.: Emotion and cognition: insights from studies of the human amygdala. Annu. Rev. Psychol. **57**(1), 27–53 (2006)
22. Santana, C.P., de Carvalho, E.A., Rodrigues, I.D., Bastos, G.S., de Souza, A.D., de Brito, L.L.: rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis. Scientific reports **12**(1), 6030 (2022)
23. Scheliga, S., Kellermann, T., Lampert, A., Rolke, R., Spehr, M., Habel, U.: Neural correlates of multisensory integration in the human brain: an ALE meta-analysis. Reviews in the Neurosciences **34**(2), 223–245 (2023)
24. Schwarzlose, R.F., Baker, C.I., Kanwisher, N.: Separate face and body selectivity on the fusiform gyrus. Journal of Neuroscience **25**(47), 11055–11059 (2005)

25. Simmons, W.K., Avery, J.A., Barcalow, J.C., Bodurka, J., Drevets, W.C., Bellgowan, P.: Keeping the body in mind: insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. Human brain mapping **34**(11), 2944–2958 (2013)
26. Tanveer, M., Richhariya, B., Khan, R.U., Rashid, A.H., Khanna, P., Prasad, M., Lin, C.T.: Machine learning techniques for the diagnosis of Alzheimer's disease: A review. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **16**(1s), 1–35 (2020)
27. Uddin, L.Q., Supekar, K., Menon, V.: Reconceptualizing functional brain connectivity in autism from a developmental perspective. Frontiers in human neuroscience **7**, 458 (2013)
28. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A.C.: A general statistical analysis for fMRI data. Neuroimage **15**(1), 1–15 (2002)
29. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 (2018)
30. Ye, Y., Ji, S.: Sparse graph attention networks. IEEE Transactions on Knowledge and Data Engineering **35**(1), 905–916 (2021)
31. Zhang, J., Guo, Y., Zhou, L., Wang, L., Wu, W., Shen, D.: Constructing hierarchical attentive functional brain networks for early AD diagnosis. Medical Image Analysis **94**, 103137 (2024)
32. Zhao, K., Duka, B., Xie, H., Oathes, D.J., Calhoun, V., Zhang, Y.: A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. Neuroimage **246**, 118774 (2022)
33. Zheng, K., Yu, S., Chen, L., Dang, L., Chen, B.: BPI-GNN: Interpretable brain network-based psychiatric diagnosis and subtyping. NeuroImage **292**, 120594 (2024)