Brain Economy-Aligned Graph Transformer

Stefano Vannoni[1,3], **Inês Won Sampaio**[1,2], **Eleonora Maggioni**[1,2], **Islem Rekik**[3]

[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
[2]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
[3]BASIRA Lab, Imperial-X (I-X) and Department of Computing, Imperial College London, London, United Kingdom

Abstract

Graph Neural Networks for brain connectomics treat every connection as topologically equivalent, ignoring brain economy, which balances the trade-off between the metabolic cost of maintaining a connection and the topological value it delivers to the network. We argue that this trade-off is a property of the channel (i.e., edge) between two brain regions, not of the regions themselves, and should therefore shape how regions interact during attention computation rather than being appended as a node feature or auxiliary loss. We introduce the *ecospace*, a two-dimensional coordinate system that characterizes each brain connection by its economy, the trade-off between connectivity strength and anatomical cost, and uses it to guide how brain regions attend to one another. We further introduce the *EcoSpace Rotary Encoding* (ESRE), an attention mechanism that injects the ecospace dimensions through asymmetric rotations of the query-key subspace, guaranteeing by mathematical identity that the attention score between two brain regions depends on the economy of the edge connecting them. Together, these two contributions form the **B**rain **G**raph **T**ransformer with **E**co**S**pace **R**otary **E**ncoding (**BGT-ESRE**). On brain graph classification tasks, BGT-ESRE achieves the highest AUC among the evaluated methods. Ablations suggest that the rotary mechanism and the economy-aligned graph construction both contribute to the performance. Our source code is available at `http://github.com/basiralab/BGT-ESRE`.

1 Introduction

Graph neural networks (GNNs) are powerful in analyzing functional brain networks derived from resting-state fMRI, demonstrating strong performance in multiple tasks, including brain disease classification Kan et al. [2022], Li et al. [2021], Cui et al. [2022] and biological sex prediction [Said et al., 2023]. Despite these advances, all existing architectures share a fundamental limitation: they treat every brain connection as topologically equivalent, aggregating neighbourhood information based solely on learned feature similarity. This design ignores the most basic organizing principle of biological neural systems, the *connectivity economy*: the trade-off between the metabolic cost of maintaining a connection and the topological value it delivers to the network [Bullmore and Sporns, 2012]. A model blind to this trade-off cannot distinguish a locally specialized sensory processor from a globally integrating connector hub, and cannot detect the systematic disruption of network economy that characterizes neurological and psychiatric disorders Van Den Heuvel and Sporns [2011], Crossley et al. [2014], de Lange et al. [2019].

Three converging lines of evidence motivate an economy-informed architecture. Bullmore and Sporns [2012] established that connectome organization emerges from an evolutionary optimization balancing wiring expenditure against topological value: long-range cortico-cortical projections, such as those connecting the precuneus to the lateral prefrontal cortex, are metabolically expensive to maintain yet deliver disproportionate topological value by bridging otherwise distant functional

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

modules and sustaining the short characteristic path lengths that enable rapid global integration [Van Den Heuvel and Sporns, 2011]. The same authors demonstrated that this economy trade-off predicts vulnerability to pathology in schizophrenia, Alzheimer's disease, and epilepsy. The connections that are disrupted earliest and most severely are precisely those that rank highest on both metabolic and topological measures [Van Den Heuvel et al., 2013, Crossley et al., 2014]. Goñi et al. [2013] explored the relationship between routing-based and diffusion-based communication measures across families of graph topologies, demonstrating their mathematical non-equivalence and showing that the two measures capture orthogonal aspects of network communication: diffusion efficiency quantifies local, random-walk-based spreading within densely connected communities, while routing efficiency quantifies global, shortest-path-based transmission across them. Seguin et al. [2023] further showed that communication measures derived from routing and diffusion-based flow model dimensions jointly predict functional connectivity (FC) and clinical phenotypes more accurately than either alone. Yet no existing brain network based GNN encodes these distinctions at the level of the attention mechanism, where they would have the most structural impact.

The core insight driving our approach is that the metabolic investment required to maintain a neural communication pathway is a property of the *channel* between two regions, not of the regions themselves. Characterizing each channel requires a coordinate system that jointly encodes its topological value and its metabolic expense, a space in which every brain network edge occupies a position reflecting both what it contributes to network communication and what it costs the brain to sustain. To this end, we introduce an **economy-aligned brain graph**, in which each edge weight is defined as the brain connectivity between two regions penalized by an exponential decay of the Euclidean distance between their atlas centroid coordinates (i.e., physical distance). This formulation directly operationalizes the Bullmore-Sporns economy model [Bullmore and Sporns, 2012]: a high edge weight requires both strong connectivity and low anatomical cost. Building on the economy-aligned graph, we introduce the **ecospace**: a two dimensional coordinate system in which each edge is characterized by a pair of topological measures computed on the economy-aligned brain connectivity adjacency matrix, rather than on the raw one, so that every coordinate inherits the economy principle by construction.

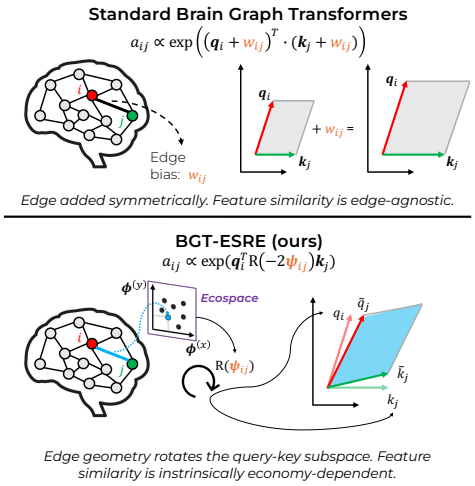


Figure 1: Standard brain graph transformers add the edge symmetrically to query and key (top); BGT-ESRE rotates them by the edge ecospace coordinate (bottom).

With the ecospace established as a principled coordinate system for brain network edges, three hypotheses govern how our proposed model operates.

Hypothesis 1. *Economy belongs to the connection, not the region.* The cost of maintaining a communication pathway between two brain regions is a property of the connection itself, not of either region individually. A model that encodes economy information at the level of individual regions, rather than at the level of the connection between them, will inevitably lose the information that matters most: how expensive and how valuable that specific link is. Economy should therefore shape how two regions communicate to each other inside the model, not be stored as a property of each region independently.

Hypothesis 2. *Weighting connections by their anatomical cost regularizes the economy of the network topology.* Topological measures computed on the economy-aligned brain graph encode qualitatively different information from the same measures computed on the raw brain graph. A robust connection that is metabolically efficient to maintain is systematically differentiated from an equally strong but metabolically expensive pathway. We argue that this differentiation encodes critical neurobiological information regarding the connection communication role within the broader network architecture. We therefore expect that a model trained on economy-aligned brain graph will learn representations that a model trained on raw brain connectivity cannot recover.

Hypothesis 3. *The mechanism by which economy acts on the attention computation determines whether the model can actually use it.* The most straightforward way to incorporate economy information into a graph transformer is to add it as a bias term to the attention score. We argue that this is insufficient since an additive bias can be ignored or overridden during training, and it does not change the fundamental way in which the model compares two brain regions to each other. We therefore seek a mechanism that embeds economy information within the region-to-region attention computation itself, rather than treating it as an external feature or additive bias (see Figure 1). We achieve this by drawing on Rotary Position Encoding (RoPE) [Su et al., 2024], which encodes the relative position of tokens in a sequence by rotating their internal representations before the attention dot product is computed. We propose an analogous mechanism for brain graphs: the ecospace coordinate of each connection is used to rotate the representations of the two regions it links, in opposite directions and by an amount reflecting the economy of that connection.

These three hypotheses together determine the architecture. We present the **Brain Graph Transformer with EcoSpace Rotary Encoding (BGT-ESRE)**, a novel graph transformer that embeds the brain economy principle directly into its attention mechanism and produces economy-aligned representations of each brain region's communication role. Our contributions are twofold. First, we introduce the ecospace, a two-dimensional coordinate system for brain connections in which each edge is characterized by a pair of topological measures computed on the economy-aligned brain connectivity matrix rather than on the raw connectivity one, so that every coordinate captures the economy principle. The choice of measure pair spans the segregation-integration spectrum [Seguin et al., 2023] and is studied systematically in the ablation. Second, we introduce an attention mechanism that projects the ecospace coordinate of each brain edge into a per-head rotation angle and applies it asymmetrically to the query and key vectors, so that the attention score between two brain regions depends on the economy of the channel connecting them. Finally, we evaluate BGT-ESRE across three experimental settings involving sex classification and clinical diagnosis classification, demonstrating its applicability to different brain-connectivity analysis tasks.

2 Related work

Graph transformers and brain-graph GNNs. Graph transformers extend the self-attention mechanism to non-sequential graph data by combining content-based attention with structural priors [Ying et al., 2021, Kreuzer et al., 2021b, Rampášek et al., 2022]. Graphormer [Ying et al., 2021] introduces spatial, edge, and centrality encodings as additive biases on attention logits; SAN [Kreuzer et al., 2021b] replaces them with learnable spectral positional encodings derived from the graph Laplacian. Concurrent work has extended RoPE to graphs by rotating node representations according to Laplacian spectral coordinates [Reid et al., 2025], a node-level mechanism that is complementary to our edge level, economy-aligned instantiation. A parallel line of work develops GNN architectures specifically for functional brain networks: BrainGNN [Li et al., 2021] introduces ROI-aware graph convolutions for interpretable fMRI analysis, BRAINNETTF [Kan et al., 2022] adapts the transformer to brain networks via a learnable orthonormal cluster readout, and ALTER [Yu et al., 2024] introduces long-range graph attention to capture distal functional dependencies.

Topology-aware GNNs and communication-based brain models. Beyond standard message-passing GNNs, two lines of work are particularly relevant to BGT-ESRE. The first incorporates explicit topological structure into graph learning through persistent homology, simplicial complexes, and sheaf neural networks Hofer et al. [2020], Bodnar et al. [2021, 2022], which characterize the global shape of a graph through filtration-based descriptors. The second models brain connectivity through explicit communication processes, routing efficiency Latora and Marchiori [2001], communicability Estrada and Hatano [2008], and network diffusion Raj et al. [2012], and uses these measures as hand-crafted predictors of FC and clinical phenotypes Goñi et al. [2013], Finn et al. [2015], Seguin et al. [2023].

3 Method

Let $G = (V, E, \mathbf{X}, \mathbf{A})$ be a brain graph for a single subject, where V is the set of N brain regions defined by an anatomical parcellation atlas, and $E \subseteq V \times V$ is the set of connections between regions. The node feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$ encodes the per-region input features, where F denotes the feature dimensionality: in the case of functional brain graphs, $\mathbf{X}$ contains the BOLD time-series of

each region ($F = T$ time points), whereas in the case of morphological brain graphs, $\mathbf{X}$ contains the morphological connectivity (MC) similarity profile of each region with respect to all other regions in the atlas ($F = N$). The binary adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is constructed differently depending on the network type: for functional brain graphs, $\mathbf{A}$ is obtained by proportional thresholding of the Pearson FC matrix, retaining only the top $\kappa\%$ strongest positive connections; for morphological brain graphs, no thresholding is applied and $\mathbf{A}$ is set to the fully connected adjacency, reflecting the fact that morphological similarity is defined between every pair of regions by construction. We hereafter use brain connectivity matrix BC to denote the weighted adjacency matrix of the brain graph, instantiated as the FC matrix for functional networks and as the MC matrix for morphological networks. The full model pipeline is illustrated in Figure 2.

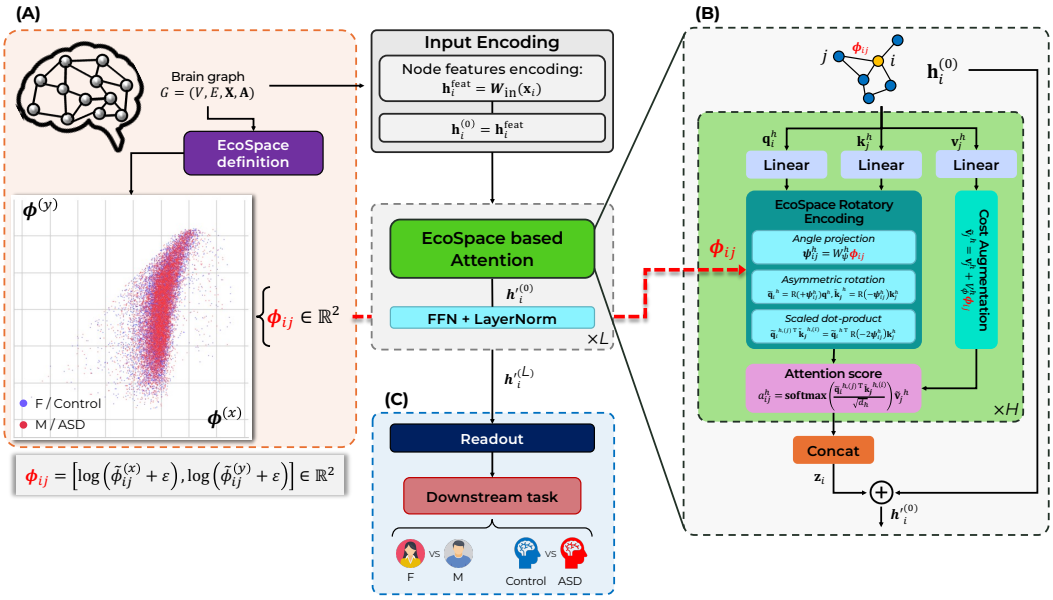


Figure 2: Proposed Brain Graph Transformer with EcoSpace Rotary Encoding (BGT-ESRE) model. **(A) Proposed ecospace:** two-dimensional, log-scaled topological coordinates are computed for each edge in the brain graph (each point in the ecospace represents a brain edge, colour-coded by class label). **(B) Ecospace based attention layer:** per-head query, key, and value are linearly projected; ESRE projects $\boldsymbol{\phi}_{ij}$ into a rotation angle $\boldsymbol{\psi}_{ij}^h$ that is applied with opposite signs to query and key, and value is augmented by a cost-augmentation branch. Heads are concatenated and added residually to $\mathbf{h}_i^{(0)}$. **(C) Downstream task:** the model is trained end-to-end on two distinct binary classification tasks.

3.1 Ecospace definition

A central design choice of BGT-ESRE is to encode the Bullmore-Sporns economy trade-off directly into the graph topology through an *economy-aligned adjacency matrix*:

$$[\mathbf{A}^{\text{eco}}]_{ij} = \text{BC}_{ij} \cdot \exp(-\lambda\, d_{ij}) \cdot A_{ij}, \tag{1}$$

where $d_{ij} = \|\mathbf{c}_i - \mathbf{c}_j\|_2$ is the Euclidean distance between the atlas centroid coordinates of regions i and j, and λ is a distance decay constant expressed in mm^{-1} (see Figure H1 in Appendix H). A pure ratio BC_{ij}/d_{ij} would assign high weights to weak short-range connections, conflating anatomically cheap connections with economically valuable ones, whereas the exponential penalty in Eq. 1 ensures that a high weight requires both strong connectivity and low anatomical cost simultaneously. By embedding the trade-off at the level of the edge weights, all graph-theoretic communication measures subsequently derived from $\mathbf{A}^{\text{eco}}$ inherit the economy principle by construction, without requiring any additional post-hoc cost correction.

For each subject we pre-compute, prior to training, a two-dimensional ecospace coordinate matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N \times N \times 2}$. The ecospace is defined by a pair of topological measures $(\phi^{(x)}, \phi^{(y)})$ selected from

five economy-aligned topological metrics, all computed on $\mathbf{A}^{\text{eco}}$: edge diffusion efficiency ($E^{\text{eco}}_{\text{diff}}$), edge routing efficiency ($E^{\text{eco}}_{\text{rout}}$), edge communicability ($C^{\text{eco}}_{\text{comm}}$), edge clustering coefficient ($C^{\text{eco}}_{\text{cc}}$), and edge betweenness centrality ($B^{\text{eco}}_{\text{bc}}$).

The mathematical formulation of each measure is reported in Appendix E. We define the ecospace coordinates of an edge (i, j) as:

$$\boldsymbol{\phi}_{ij} = \left[\log(\tilde{\phi}^{(x)}_{ij} + \varepsilon),\ \log(\tilde{\phi}^{(y)}_{ij} + \varepsilon)\right] \in \mathbb{R}^2, \tag{2}$$

where $\tilde{\phi}^{(k)}_{ij}$ denotes the per-subject min-max normalization to $[0, 1]$ of the k-th selected measure, applied prior to the log transformation to ensure both dimensions contribute equally to the angle projection regardless of their absolute scale for each edge (i, j), and $\varepsilon > 0$ is a small constant added for numerical stability. The ecospace coordinates are fixed during training and stored as edge attributes; only the entries corresponding to edges in E are used by the model.

3.2 EcoSpace Rotary Encoding (ESRE)

ESRE encodes the economy-aligned geometry of each brain connection directly into the query-key dot product through asymmetric rotations of the embedding subspace. The full attention block is depicted in Figure 2(B); we describe its four computational steps below for a single attention head h and edge (i, j), with per head dimension $d_h = d/H$.

Step 1 – Linear projections. For each attention head h, the input query, key, and value projections are computed directly from the layer's input embeddings $\mathbf{h}^{(\ell)}$:

$$\mathbf{q}^h_i = W^h_Q\, \mathbf{h}^{(\ell)}_i, \quad \mathbf{k}^h_j = W^h_K\, \mathbf{h}^{(\ell)}_j, \quad \mathbf{v}^h_j = W^h_V\, \mathbf{h}^{(\ell)}_j, \tag{3}$$

where $W^h_Q, W^h_K, W^h_V \in \mathbb{R}^{d_h \times d}$ are learned head-specific projection matrices, i denotes the receiving node (query) and j the sending node (key, value). No edge information enters at this stage, keeping the ESRE modification strictly localized to the rotary mechanism.

Step 2 – Angle projection. The ecospace coordinate $\boldsymbol{\phi}_{ij} \in \mathbb{R}^2$ is mapped to a $d_h/2$-dimensional rotation-angle vector via a head-specific learned projection:

$$\boldsymbol{\psi}^h_{ij} = W^h_\psi\, \boldsymbol{\phi}_{ij} \in \mathbb{R}^{d_h/2}, \tag{4}$$

where $W^h_\psi \in \mathbb{R}^{d_h/2 \times 2}$ is the sole trainable parameter connecting the ecospace to the attention operator, allowing each head to learn an independent profile over the two ecospace dimensions.

Step 3 – Asymmetric block-diagonal rotation. The angle vector $\boldsymbol{\psi}^h_{ij}$ defines a block-diagonal rotation matrix of $d_h/2$ independent 2×2 blocks:

$$\mathbf{R}(\boldsymbol{\psi}^h_{ij}) = \bigoplus_{k=1}^{d_h/2} \begin{pmatrix} \cos\psi^{h,k}_{ij} & -\sin\psi^{h,k}_{ij} \\ \sin\psi^{h,k}_{ij} & \cos\psi^{h,k}_{ij} \end{pmatrix} \tag{5}$$

The rotation is applied *asymmetrically*: the query by $+\boldsymbol{\psi}^h_{ij}$ and the key by $-\boldsymbol{\psi}^h_{ij}$:

$$\tilde{\mathbf{q}}^{h,(j)}_i = \mathbf{R}(+\boldsymbol{\psi}^h_{ij})\, \mathbf{q}^h_i, \qquad \tilde{\mathbf{k}}^{h,(i)}_j = \mathbf{R}(-\boldsymbol{\psi}^h_{ij})\, \mathbf{k}^h_j \tag{6}$$

The opposite-sign convention is essential. Applying the same rotation to both vectors cancels in the dot product, $[\mathbf{R}(+\boldsymbol{\psi})\mathbf{q}]^\top[\mathbf{R}(+\boldsymbol{\psi})\mathbf{k}] = \mathbf{q}^\top\mathbf{k}$, which would leave the score unchanged and reduce ESRE to standard attention. The opposite-sign rotation instead exploits the composition identity $\mathbf{R}(+\boldsymbol{\psi})^\top\mathbf{R}(-\boldsymbol{\psi}) = \mathbf{R}(-2\boldsymbol{\psi})$, yielding:

$$\tilde{\mathbf{q}}^{h,(j)\top}_i \tilde{\mathbf{k}}^{h,(i)}_j = \mathbf{q}^{h\top}_i\, \mathbf{R}(-2\boldsymbol{\psi}^h_{ij})\, \mathbf{k}^h_j \tag{7}$$

The attention score therefore depends on $\boldsymbol{\phi}_{ij}$ by mathematical identity: feature similarity between two regions is characterized in terms of the economy-aligned geometry of the channel connecting them, rather than as a fixed property of the regions themselves. Because $\mathbf{R}$ is orthogonal, $\|\mathbf{R}\mathbf{x}\| = \|\mathbf{x}\|$, so the projected feature vectors retain their norms and the attention temperature does not vary systematically with $\|\boldsymbol{\phi}_{ij}\|$, in contrast with additive encoding schemes, where the bias magnitude entangles with the content magnitude. The block-diagonal structure admits an efficient element-wise

implementation in $\mathcal{O}(|E| \cdot d_h)$ without explicit matrix construction. The resulting attention weights are:

$$a_{ij}^h = \frac{\exp\left(\tilde{\mathbf{q}}_i^{h,(j)\top} \tilde{\mathbf{k}}_j^{h,(i)} / \sqrt{d_h}\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\tilde{\mathbf{q}}_i^{h,(k)\top} \tilde{\mathbf{k}}_k^{h,(i)} / \sqrt{d_h}\right)} \tag{8}$$

A formal proof of permutation equivariance is provided in Appendix A.

Step 4 – Ecospace conditioned value augmentation. The value vector $\mathbf{v}_j^h$ is edge-agnostic: region j transmits the same message regardless of which economy channel connects it to region i. To let message content adapt to the channel, we augment the value with the projected ecospace coordinate, gated by a learnable scalar γ initialized to zero:

$$\tilde{\mathbf{v}}_j^{h,(i)} = \mathbf{v}_j^h + \tanh(\gamma)\, V_\phi^h\, \boldsymbol{\phi}_{ij}, \tag{9}$$

with $V_\phi^h \in \mathbb{R}^{d_h \times 2}$. The gate ensures the correction is inactive at initialization and grows only as training demands it. Information is then aggregated using the rotated attention weights:

$$\mathbf{z}_i^h = \sum_{j \in \mathcal{N}(i)} a_{ij}^h\, \tilde{\mathbf{v}}_j^{h,(i)}, \tag{10}$$

and the multi-head output is formed by concatenating and projecting the heads, $\mathbf{z}_i = W_O[\mathbf{z}_i^1 \| \cdots \| \mathbf{z}_i^H]$. The layer update then follows the standard pre-norm residual scheme with a Gaussian Error Linear Unit (GELU) Feed-Forward Network (FFN):

$$\mathbf{h}_i' = \mathbf{h}_i^{(\ell)} + \mathbf{z}_i, \qquad \mathbf{h}_i^{(\ell+1)} = \mathbf{h}_i' + \mathrm{FFN}(\mathrm{LN}(\mathbf{h}_i')) \tag{11}$$

3.3 Input encoding, readout, and training objective

For functional brain graphs, each node feature $\mathbf{x}_i \in \mathbb{R}^T$ consists of the BOLD time-series of region i, which is encoded using a grouped 1D-CNN with `groups` $= N$ throughout all convolutional layers, ensuring that each brain region is processed by an independent set of temporal filters with no cross-region mixing; an early downsampling stage reduces the sequence length before the deeper layers. For morphological brain graphs, each node feature $\mathbf{x}_i \in \mathbb{R}^N$ consists of the MC of region i with respect to all other regions, which is projected into the model dimension via a simple linear layer. In both cases, the encoder outputs one feature vector per region $\mathbf{h}_i^{(0)} \in \mathbb{R}^d$, which serves as the initial node embedding fed into the stack of L ecospace-based attention layers. After the final layer, node embeddings are aggregated via mean pooling to obtain a graph-level representation $\mathbf{h}_G = \frac{1}{N}\sum_{i=1}^N \mathbf{h}_i^{(L)} \in \mathbb{R}^d$, which is then processed by a two-layer MLP classifier with GELU activation and dropout to yield the predicted class distribution $\hat{\mathbf{y}}$. The model is trained end-to-end using the standard cross entropy loss:

$$\mathcal{L} = -\frac{1}{|\mathcal{B}|} \sum_{G \in \mathcal{B}} \mathbf{y}_G^\top \log \hat{\mathbf{y}}_G, \tag{12}$$

where $\mathbf{y}_G$ is the one-hot ground-truth label, $\hat{\mathbf{y}}_G$ is the predicted class distribution, and $\mathcal{B}$ denotes a mini-batch of subjects.

4 Experiments

4.1 Experimental setup

Datasets and preprocessing. We evaluated BGT-ESRE on three neuroimaging datasets spanning functional and morphological brain networks from two cohorts, across two downstream tasks: sex classification and autism spectrum disorder (ASD) classification. For sex classification, we used resting-state fMRI of 748 healthy participants (405 female, 343 male) from the Human Connectome Project Young Adult cohort (HCP-YA) [Van Essen et al., 2013]. For ASD classification, we used two datasets from the Autism Brain Imaging Data Exchange I (ABIDE I) [Di Martino et al., 2014]: ABIDE I FC, comprising resting-state fMRI of 884 participants (476 neurotypical controls and 408

patients with ASD), obtained through the Preprocessed Connectomes Project [Craddock et al., 2013], and ABIDE I MC comprising morphological connectomes of 341 participants (155 neurotypical controls and 186 patients with ASD). Full preprocessing details are provided in Appendix B.

Baselines. The selected baselines are divided into three distinct categories to ensure a comprehensive comparative analysis. The first category comprises classical machine learning methods, specifically Support Vector Machine (SVM) and Random Forest. The second category consists of general-purpose graph transformers (not explicitly designed for brain networks), including Graphormer [Ying et al., 2021], SAN [Kreuzer et al., 2021a], and GATv2 [Brody et al., 2021]. The third category encompasses brain graph-specific neural networks tailored for connectomics, namely BRAINNETTF [Kan et al., 2022], ALTER [Yu et al., 2024], and BrainGNN Li et al. [2021]. Note that for the deep learning baselines, the original code shared by the respective authors was utilized and adapted for this comparative analysis. Full details of the baseline training and comparison protocol are provided in Appendix D.

Evaluation measures. We utilize standard machine learning and diagnostic specific metrics to evaluate the performance of the proposed method. These include classification Accuracy (ACC), Area Under the Receiver Operating Characteristic Curve (AUC), Sensitivity (SEN), and Specificity (SPE). In the experimental results, we report the mean and standard deviation across a 5-fold cross-validation. Full model implementation details and dataset-specific hyperparameters are reported in Appendix C.

4.2 Performances comparison

Results. Table 1 summarizes the classification performances across the three datasets. On HCP-YA, BGT-ESRE achieves the best performance among all evaluated methods, with ACC 0.769±0.045 and AUC 0.895±0.012, showing a clear improvement over both general-purpose graph transformers and brain graph-specific baselines. On ABIDE I MC, BGT-ESRE achieves the highest AUC (0.633±0.044) and matches the highest ACC (0.578±0.059), indicating a more modest but still competitive gain in the clinical classification setting. On ABIDE I FC, BGT-ESRE also obtains the highest AUC (0.697±0.034), while achieving ACC comparable to the strongest baseline. Taken together, these results demonstrate that encoding the brain economy principle directly into the attention mechanism yields consistent gains across both functional and morphological brain graph classification tasks.

Table 1: Performance comparison across the three datasets. HCP-YA is used for sex classification, whereas ABIDE I MC and ABIDE I FC are used for ASD diagnosis classification. Results are reported as mean (std) across the five folds of a *k*-fold cross-validation. **Bold** indicates the highest result per column.

Category	Method	HCP-YA – Sex classification		ABIDE I MC – ASD Diagnosis		ABIDE I FC – ASD Diagnosis	
		ACC	AUC	ACC	AUC	ACC	AUC
ML Methods	SVM	0.568 (0.030)	0.558 (0.025)	0.566 (0.039)	0.531 (0.087)	0.533 (0.002)	0.472 (0.054)
	Random Forest	0.583 (0.019)	0.597 (0.030)	0.545 (0.061)	0.513 (0.073)	0.551 (0.038)	0.542 (0.041)
Graph Transformer Models	Graphormer	0.604 (0.030)	0.665 (0.044)	0.578 (0.059)	0.616 (0.075)	0.527 (0.033)	0.547 (0.014)
	SAN	0.517 (0.039)	0.535 (0.046)	0.537 (0.062)	0.599 (0.055)	0.458 (0.044)	0.447 (0.052)
	GATv2	0.525 (0.033)	0.530 (0.063)	0.551 (0.055)	0.572 (0.080)	0.523 (0.031)	0.540 (0.056)
Brain Graph Neural Networks	BRAINNETTF	0.666 (0.189)	0.789 (0.154)	0.540 (0.046)	0.601 (0.210)	0.652 (0.037)	0.687 (0.035)
	BrainGNN	0.642 (0.046)	0.710 (0.052)	0.558 (0.051)	0.607 (0.067)	0.623 (0.041)	0.665 (0.043)
	ALTER	0.591 (0.078)	0.571 (0.125)	0.507 (0.086)	0.503 (0.099)	0.606 (0.036)	0.651 (0.039)
Our Model	**BGT-ESRE**	**0.769 (0.045)**	**0.895 (0.012)**	**0.578 (0.059)**	**0.633 (0.044)**	**0.652 (0.040)**	**0.697 (0.034)**

4.3 Ablation study

Table 2 isolates the contribution of each component. Replacing the rotary encoding with additive injection of the same ecospace coordinates into the query and key embeddings (w/o Rotation) produces the largest drop. Replacing the economy-aligned adjacency with a standard connectivity-weighted one (FC and MC) while retaining the rotary mechanism yields a consistent additional drop across both tasks. A smaller decay constant $\lambda = 1/\bar{d}$ further degrades performance relative to $\lambda = 0.1$, indicating that the choice of decay constant is consequential. The ecospace configuration

Table 2: Ablation study of the BGT-ESRE architecture across the three datasets. We compare the full model against variants without ecospace rotation, with raw connectivity-weighted ecospace, and with an alternative distance-decay parameter. Results are reported as mean (std) across the five folds of a k-fold cross-validation. **Bold** indicates the highest result per column.

Category	Model variant	HCP-YA		ABIDE I MC		ABIDE I FC	
		ACC	AUC	ACC	AUC	ACC	AUC
BGT-ESRE (Ablations)	w/o Rotation	0.729 (0.013)	0.843 (0.024)	0.528 (0.036)	0.505 (0.106)	0.613 (0.018)	0.690 (0.033)
	Raw connectivity w/ Rotation	0.749 (0.063)	0.873 (0.023)	0.531 (0.069)	0.553 (0.110)	0.633 (0.022)	0.686 (0.035)
	$\lambda = 1/\bar{d}$ w/ Rotation	0.707 (0.045)	0.868 (0.018)	0.533 (0.097)	0.579 (0.096)	0.630 (0.024)	0.692 (0.034)
	Full Model	**0.769 (0.045)**	**0.895 (0.012)**	**0.578 (0.059)**	**0.633 (0.044)**	**0.652 (0.040)**	**0.697 (0.034)**

analysis (Table G1, Appendix G) shows that performances are robust across all ten measure pairs, with the highest results coming from configurations pairing a local segregation measure with a global integration measure; communicability appears among the top configurations on both datasets.

5 Discussion

Economy-aligned attention improves brain graph classification. The results of the ablation study in Table 2 supports the contribution of both main components of BGT-ESRE: the economy-aligned adjacency matrix and the asymmetric rotary attention mechanism. Removing ESRE or replacing the economy-aligned ecospace with raw connectivity consistently reduces performance, indicating that the observed gains are not simply due to increased model capacity or the availability of additional edge information. Rather, the results suggest that channel-level economy information is more effectively used when it modulates the query-key comparison through asymmetric rotation, supporting ***Hypothesis 1*** and ***Hypothesis 3***. The economy-aligned adjacency also contributes independently, supporting ***Hypothesis 2***: penalizing connectivity by anatomical distance provides an edge-coordinate system that captures information not fully recovered from raw connectivity alone.

Ecospace geometry and the segregation-integration balance. The ecospace configuration analysis reveals that the pairs yielding the highest results consistently combine a segregation-related measure with an integration-related measure, consistent with the routing-diffusion morphospace of Goñi et al. [2013] and the taxonomy of Seguin et al. [2023]; the pairwise ecospace visualizations and subject-level distributions supporting this interpretation are provided in Appendix G, Figure G1-G4. Communicability appears among the top configurations on all three datasets and both modalities, suggesting that broad-band multi-walk integration captures a dimension of economy-efficient communication that single-path measures do not. This is consistent with the sensitivity of communicability to redundant parallel pathways [Estrada and Hatano, 2008], a structural property likely to differ systematically between neurotypical and clinical populations. Together, these patterns suggest that the segregation-integration trade-off is a robust organizing principle for the ecospace across brain connectivity modalities and classification tasks.

Economy-aligned attention reflects known group differences in connectome organization. Figure 3 projects group-level differences in learned attention weights onto the ecospace. For the two functional datasets, HCP-YA and ABIDE I FC (Figure 3 A,C), attention differences are visualized in the same diffusion–communicability ecospace. In HCP-YA (Figure 3 A), male-biased attention is more visible toward edges with higher diffusion efficiency and communicability, whereas female-biased attention appears more evident toward lower-communicability regions. This qualitative pattern is compatible with previous reports of sex-related differences in functional network organization [Ingalhalikar et al., 2014, Satterthwaite et al., 2015]. In ABIDE I FC (Figure 3 C), ASD-biased attention is more associated with edges showing higher diffusion efficiency and communicability, whereas control-biased attention is more associated with lower values of these measures. This suggests that the classifier relies more strongly on ASD-related functional connections embedded in communication-efficient and redundantly connected regions of the economy-aligned graph. This pattern is consistent with resting-state fMRI studies showing that ASD is not characterized by a uniform connectivity alteration, but by heterogeneous patterns of both hyperconnectivity and hypoconnectivity across networks and subjects [Ilioska et al., 2023, Hull et al., 2017]. For ABIDE I MC (Figure 3 B), attention differences are instead visualized in the betweenness-communicability ecospace. ASD-biased attention is more apparent among edges with lower edge betweenness

centrality, whereas neurotypical-biased attention extends toward edges with higher betweenness. This pattern is qualitatively consistent with the view that ASD is associated with atypical integration and altered long-range connectivity [Just et al., 2004, Geschwind and Levitt, 2007].
This suggests that the model captures different aspects of economy-aligned communication depending on the input modality: diffusion and communicability for FC, and routing-related betweenness for MC. Overall, these attention maps indicate that BGT-ESRE uses economy-aligned edge coordinates when forming task-specific graph representations. However, these patterns should be interpreted cautiously, because attention weights do not by themselves establish causal or mechanistic neurobiological effects. Therefore, we interpret the results as hypothesis-generating evidence that the model exploits economy-aligned functional and morphological communication properties.

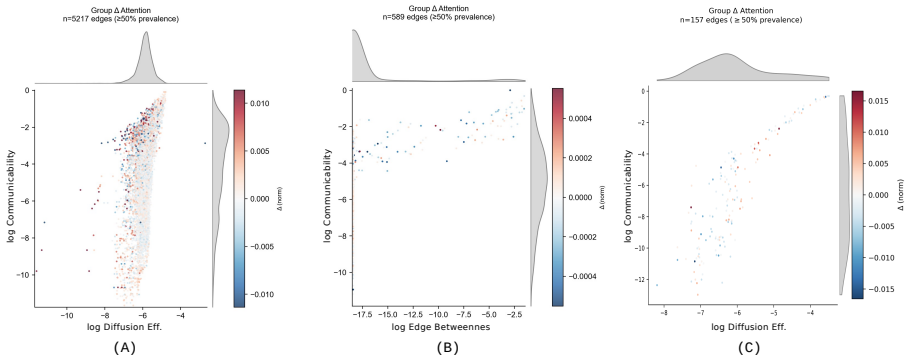


Figure 3: Group difference in attention weights (Δ attention) projected onto the ecospace for HCP-YA (**A**, $E^{\text{eco}}_{\text{diff}}/C^{\text{eco}}_{\text{comm}}$), ABIDE I MC (**B**, $B^{\text{eco}}_{\text{bc}}/C^{\text{eco}}_{\text{comm}}$), and ABIDE I FC (**C**, $E^{\text{eco}}_{\text{diff}}/C^{\text{eco}}_{\text{comm}}$). Each point represents a single brain edge, restricted to edges present in at least 50% of subjects in the respective dataset (prevalence $\geq$ 50%); colour encodes the normalized difference in mean attention weight between groups (positive = male/ASD, negative = female/Control).

Limitations The economy-aligned adjacency matrix uses the Euclidean distance between atlas centroid coordinates as a proxy for axonal wiring cost, which is a geometric approximation of the metabolic cost that the Bullmore-Sporns framework originally formulates in terms of white-matter tract length [Bullmore and Sporns, 2012]. Replacing this proxy with subject-specific tract-length priors derived from diffusion MRI tractography would tighten the correspondence between the architectural inductive bias and its neurobiological motivation, and represents the most natural direction for future work.

A BGT-ESRE perspective and broader impact. Beyond population-level classification, the economy-aligned representation at the core of BGT-ESRE opens a principled path toward modeling higher-order cognitive processes. A growing body of evidence demonstrates that the brain's cost-value trade-off is not merely a wiring constraint but a functional one: economy-efficient network organization directly predicts individual differences in executive function, memory, complex reasoning, and social cognition Zhou et al. [2022], and genetic propensity for cognitive ability is itself associated with weaker wiring-cost penalties and more topologically diverse connectomes Monaghan et al. [2025]. More broadly, the metabolic cost of sustaining a connection shapes the brain's capacity for higher-level computation, with parsimonious communication processes underpinning the most cognitively demanding behaviours Bullmore and Sporns [2012], Jamadar et al. [2025]. Metacognition, the brain's capacity to monitor and evaluate its own internal states, relies specifically on frontoparietal and precuneus connectivity Hu et al. [2025], networks whose economy-aligned communication geometry BGT-ESRE is well-positioned to characterize. By embedding the cost-value trade-off directly into the attention mechanism, our model provides a framework for investigating how economy-constrained communication channels support or constrain such higher-order processes, moving connectome-based modeling beyond diagnostic classification toward a principled geometry of cognitive function.

6 Conclusion

We introduced BGT-ESRE, a brain-graph transformer that embeds the brain economy principle directly into its attention mechanism via two independent contributions: the **ecospace** and the

EcoSpace Rotary Encoding (ESRE). By guaranteeing that attention scores depend on the economy of the channel connecting two regions, the model learns attention profiles consistent with known group differences in connectome organization and outperforms all tested baselines on both sex classification and ASD diagnosis. These results suggest that the cost-value trade-off is not merely a wiring constraint but a genuine source of discriminative signal for brain-graph learning. More broadly, by grounding attention in the economy of neural communication, BGT-ESRE provides a principled framework for investigating how metabolic constraints shape higher-order cognitive processes moving connectome-based modeling beyond diagnostic classification toward a geometry of cognitive function.

References

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural information processing systems*, 34:2625–2640, 2021.

Cristian Bodnar, Francesco Di Giovanni, Benjamin Chamberlain, Pietro Lio, and Michael Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns. *Advances in Neural Information Processing Systems*, 35:18527–18541, 2022.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349, 2012.

Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7(27):5, 2013.

Nicolas A Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, and Carl Yang. Braingb: a benchmark for brain network analysis with graph neural networks. *IEEE transactions on medical imaging*, 42(2):493–506, 2022.

Siemon C de Lange, Lianne H Scholtens, Alzheimer's Disease Neuroimaging Initiative, Leonard H van den Berg, Marco P Boks, Marco Bozzali, Wiepke Cahn, Udo Dannlowski, Sarah Durston, Elbert Geuze, et al. Shared vulnerability for connectome alterations across psychiatric and neurological brain disorders. *Nature human behaviour*, 3(9):988–998, 2019.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Nico UF Dosenbach, Binyam Nardos, Alexander L Cohen, Damien A Fair, Jonathan D Power, Jessica A Church, Steven M Nelson, Gagan S Wig, Alecia C Vogel, Christina N Lessov-Schlaggar, et al. Prediction of individual brain maturity using fmri. *Science*, 329(5997):1358–1361, 2010.

Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 77(3):036111, 2008.

Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015.

Jean-Philippe Fortin, Drew Parker, Birkan Tunç, Takanori Watanabe, Mark A Elliott, Kosha Ruparel, David R Roalf, Theodore D Satterthwaite, Ruben C Gur, Raquel E Gur, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage*, 161:149–170, 2017.

Daniel H Geschwind and Pat Levitt. Autism spectrum disorders: developmental disconnection syndromes. *Current opinion in neurobiology*, 17(1):103–111, 2007.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

Joaquín Goñi, Andrea Avena-Koenigsberger, Nieves Velez de Mendizabal, Martijn P van den Heuvel, Richard F Betzel, and Olaf Sporns. Exploring the morphospace of communication efficiency in complex networks. *PLoS One*, 8(3):e58070, 2013.

Christoph Hofer, Florian Graf, Bastian Rieck, Marc Niethammer, and Roland Kwitt. Graph filtration learning. In *International Conference on Machine Learning*, pages 4314–4323. PMLR, 2020.

Ruocen Hu, Meng Yu, Liangfang Li, Hui He, Sihan Wei, Junji Ma, Yue Gu, and Zhengjia Dai. The specificity of metacognition questionnaire-30 subdimensions: Findings from connectome-based predictive modeling. *Depression and Anxiety*, 2025(1):5581270, 2025.

Jocelyn V Hull, Lisa B Dokovna, Zachary J Jacokes, Carinna M Torgerson, Andrei Irimia, and John Darrell Van Horn. Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry*, 7:205, 2017.

Iva Ilioska, Marianne Oldehinkel, Alberto Llera, Sidhant Chopra, Tristan Looden, Roselyne Chauvin, Daan Van Rooij, Dorothea L Floris, Julian Tillmann, Carolin Moessnang, et al. Connectome-wide mega-analysis reveals robust patterns of atypical functional connectivity in autism. *Biological psychiatry*, 94(1):29–39, 2023.

Madhura Ingalhalikar, Alex Smith, Drew Parker, Theodore D Satterthwaite, Mark A Elliott, Kosha Ruparel, Hakon Hakonarson, Raquel E Gur, Ruben C Gur, and Ragini Verma. Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2):823–828, 2014.

Sharna D Jamadar, Anna Behler, Hamish Deery, and Michael Breakspear. The metabolic costs of cognition. *Trends in Cognitive Sciences*, 29(6):541–555, 2025.

Marcel Adam Just, Vladimir L Cherkassky, Timothy A Keller, and Nancy J Minshew. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127(8):1811–1821, 2004.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

Devin Kreuzer, Dominique Beaini, Will Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. *Advances in neural information processing systems*, 34:21618–21629, 2021a.

Devin Kreuzer, Dominique Beaini, Will Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. *Advances in neural information processing systems*, 34:21618–21629, 2021b.

Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical image analysis*, 74:102233, 2021.

Alicja Monaghan, Danyal Akarca, and Duncan E Astle. Brain wiring economics, network organisation and population-level genomics. *Imaging Neuroscience*, 3:IMAG–a, 2025.

W Penny, K Friston, J Ashburner, S Kiebel, and T Nichols. Statistical parametric mapping: The analysis of functional brain images.; 2007. doi: 10.1016. Technical report, B978-0-12-372560-8. X5000-1.

Ashish Raj, Amy Kuceyeski, and Michael Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in neural information processing systems*, 35:14501–14515, 2022.

Isaac Reid, Arijit Sehanobish, Cederik Höfs, Bruno Mlodozeniec, Leonhard Vulpius, Federico Barbero, Adrian Weller, Krzysztof Choromanski, Richard E Turner, and Petar Veličković. Rotary position encodings for graphs. *arXiv preprint arXiv:2509.22259*, 2025.

Anwar Said, Roza Bayrak, Tyler Derr, Mudassir Shabbir, Daniel Moyer, Catie Chang, and Xenofon Koutsoukos. Neurograph: Benchmarks for graph machine learning in brain connectomics. *Advances in Neural Information Processing Systems*, 36:6509–6531, 2023.

Theodore D Satterthwaite, Daniel H Wolf, David R Roalf, Kosha Ruparel, Guray Erus, Simon Vandekar, Efstathios D Gennatas, Mark A Elliott, Alex Smith, Hakon Hakonarson, et al. Linked sex differences in cognition and functional connectivity in youth. *Cerebral cortex*, 25(9):2383–2394, 2015.

Caio Seguin, Olaf Sporns, and Andrew Zalesky. Brain network communication: concepts, models and applications. *Nature reviews neuroscience*, 24(9):557–574, 2023.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

Martijn P Van Den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *The Journal of neuroscience*, 31(44):15775–15786, 2011.

Martijn P Van Den Heuvel, Olaf Sporns, Guusje Collin, Thomas Scheewe, René CW Mandl, Wiepke Cahn, Joaquín Goñi, Hilleke E Hulshoff Pol, and René S Kahn. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA psychiatry*, 70(8), 2013.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393 (6684):440–442, 1998.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

Shuo Yu, Shan Jin, Ming Li, Tabinda Sarwar, and Feng Xia. Long-range brain graph transformer. *Advances in Neural Information Processing Systems*, 37:24472–24495, 2024.

Dale Zhou, Christopher W Lynn, Zaixu Cui, Rastko Ciric, Graham L Baum, Tyler M Moore, David R Roalf, John A Detre, Ruben C Gur, Raquel E Gur, et al. Efficient coding in the economics of human brain connectomics. *Network Neuroscience*, 6(1):234–274, 2022.

Appendices

A Permutation equivariance of BGT-ESRE

A function f over graphs is permutation equivariant if for any permutation matrix $\mathbf{P}$:

$$f(\mathbf{PX},\, \mathbf{PAP}^\top,\, \pi(\mathbf{\Phi})) = \mathbf{P}\, f(\mathbf{X}, \mathbf{A}, \mathbf{\Phi}) \tag{13}$$

where $\pi(\mathbf{\Phi})$ denotes ecospace edge attributes relabelled as $\boldsymbol{\phi}_{\pi(i)\pi(j)} = \boldsymbol{\phi}_{ij}$. We verify the two core components of BGT-ESRE.

Ecospace coordinates. The economy-weighted adjacency $\mathbf{A}^{\mathrm{eco}}$ is symmetric by construction since $\mathrm{BC}_{ij} = \mathrm{BC}_{ji}$ and $d_{ij} = d_{ji}$. All five ecospace metrics are symmetric functions of this symmetric adjacency: betweenness centrality sums over unordered source-target pairs, routing efficiency derives from undirected shortest paths, diffusion efficiency from a symmetric transition matrix, communicability from the symmetric matrix exponential $[\exp(\mathbf{A}^{\mathrm{eco}})]_{ij}$, and clustering coefficient from common neighbours. Therefore $\boldsymbol{\phi}_{ij} = \boldsymbol{\phi}_{ji}$ for all $(i, j) \in E$, and the ecospace coordinates are well-defined, permutation-invariant edge attributes.

ESRE attention. The linear projections W_Q, W_K, W_V are node-independent and equivariant. The rotation angle $\boldsymbol{\psi}_{\pi(i)\pi(j)} = W_\psi^h \boldsymbol{\phi}_{\pi(i)\pi(j)} = W_\psi^h \boldsymbol{\phi}_{ij} = \boldsymbol{\psi}_{ij}$ is invariant to node relabelling for a given edge, since W_ψ^h is shared across all edges and $\boldsymbol{\phi}$ transforms equivariantly with the edge set. The pre-softmax logit therefore satisfies $e_{\pi(i)\pi(j)} = e_{ij}$, and since the neighbourhood of $\pi(i)$ in the permuted graph is $\{\pi(j) : j \in \mathcal{N}(i)\}$, the softmax over that neighbourhood gives $a_{\pi(i)\pi(j)} = a_{ij}$. The scatter aggregation over $\mathcal{N}(\pi(i))$ then produces $\mathbf{z}_{\pi(i)} = [\mathbf{PZ}]_{\pi(i)}$. Combined with node-independent output projection and residual, each ESRE layer is permutation equivariant.

B Datasets and preprocessing details

HCP-YA. Resting-state fMRI data of 748 healthy subjects were obtained from the Human Connectome Project Young Adult dataset [Van Essen et al., 2013]. Data underwent the HCP minimal preprocessing pipeline [Glasser et al., 2013] supplemented with SPM8 [Penny et al.], including slice-timing and motion correction, spatial registration to standard templates, and confound regression to remove trends, motion parameters, and white matter/cerebrospinal fluid signals. The resulting regional time courses were detrended, band-pass filtered (0.01–0.1 Hz), and z-scored. Regional signals were extracted using the Glasser multimodal parcellation ($N = 360$) with 19 subcortical parcels, for a total of 379 regions of interests (ROIs) [Glasser et al., 2016], while subject-level $N \times N$ symmetric FC matrices were constructed via Pearson correlation. This dataset was used for sex classification (female vs. male).

ABIDE I FC. Resting-state fMRI data of 884 subjects, including individuals with ASD and neurotypical controls, were obtained from the Autism Brain Imaging Data Exchange (ABIDE) dataset [Di Martino et al., 2014] through the Preprocessed Connectomes Project [Craddock et al., 2013]. Data were preprocessed using the C-PAC pipeline, including motion correction, spatial normalization, nuisance-signal regression, detrending, band-pass filtering (0.01–0.1 Hz), and z-scoring. Regional time courses were extracted using the Dosenbach atlas ($N = 160$) [Dosenbach et al., 2010], and subject-level $N \times N$ symmetric FC matrices were constructed via Pearson correlation. Since ABIDE is a multi-site dataset, ComBat harmonization [Fortin et al., 2017] was applied to reduce site-related distributional shifts. Site was treated as the batch variable, while sex, age, and diagnosis were included as covariates to preserve biologically and clinically relevant variation. This dataset was used for ASD diagnosis classification.

ABIDE I MC. Morphological brain network data from 341 subjects (155 neurotypical controls and 186 individuals with ASD) were obtained from the ABIDE I public dataset [Di Martino et al., 2014]. For each hemisphere, the cortical surface was reconstructed from T1-weighted MRI using the FreeSurfer pipeline and parcellated into 35 regions of interest using the Desikan-Killiany atlas [Desikan et al., 2006]. The morphological brain network was derived from mean curvature by computing the pairwise absolute difference in this cortical measure between all pairs of regions, yielding a symmetric $N \times N$ MC matrix per subject. This dataset was used for ASD diagnosis classification.

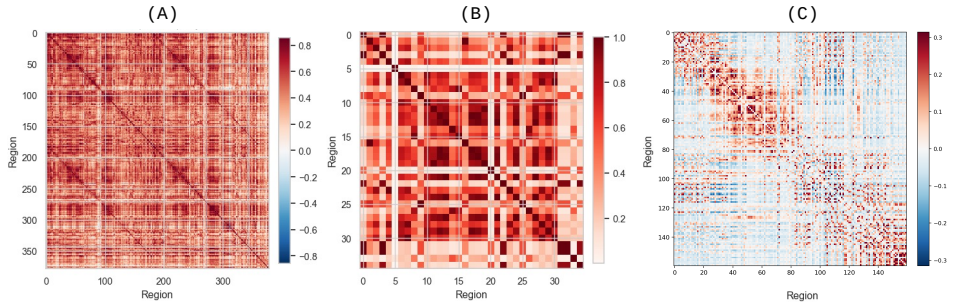


Figure B1: Representative brain connectivity matrices for a single random subject taken from each of the three datasets used. (**A**) FC matrix for HCP-YA ($N = 379$ regions), constructed via Pearson correlation of regional BOLD time-series; colour encodes correlation strength. (**B**) MC similarity matrix for ABIDE I MC ($N = 35$ regions), obtained by applying a Gaussian kernel to the pairwise absolute difference in mean curvature; colour encodes morphological similarity. (**C**) FC matrix for ABIDE I FC ($N = 160$ regions), constructed via Pearson correlation of regional BOLD time-series; colour encodes correlation strength.

Brain connectivity matrix preprocessing. The economy-aligned adjacency formulation requires a similarity-based connectivity matrix in which higher weights reflect stronger and cheaper connections. For the HCP-YA and ABIDE I FC datasets, Pearson FC matrices already satisfy this requirement, since higher values directly encode stronger functional coupling between regions; therefore, no additional transformation was needed before constructing the economy-aligned adjacency matrix. For ABIDE I MC, the MC matrix encodes pairwise dissimilarity, so it was converted to a similarity matrix via a Gaussian kernel:

$$[\mathbf{MC}^{\text{sim}}]_{ij} = \exp\left(-\frac{\bar{\text{MC}}_{ij}}{\sigma}\right), \tag{14}$$

where $\sigma = \bar{\text{MC}}$ is the mean pairwise dissimilarity across all subjects and edges. This ensures that an edge at the mean dissimilarity receives a similarity weight of $e^{-1} \approx 0.368$, placing the kernel inflection point at the centre of the empirical distribution, analogously to the choice $\lambda = 1/\bar{d}$ for the distance decay constant. The resulting $\mathbf{MC}^{\text{sim}}$ was used in place of the raw dissimilarity matrix in all subsequent computations. Representative examples of FC and MC matrices are shown in Figure B1.

C Model implementation details and hyperparameters

This appendix reports the main implementation choices and hyperparameters used for the three experimental settings: HCP-YA, ABIDE I FC, and ABIDE I MC. We report the hyperparameters required to reproduce the model configuration, graph construction, optimization procedure, and model-selection protocol. All experiments were performed on a single NVIDIA RTX 4090 GPU.

HCP-YA. For HCP-YA, the input graph was constructed from Pearson FC matrix. Negative FC values were discarded, the diagonal was set to zero, and the binary adjacency matrix was obtained by proportional thresholding, retaining the strongest $\kappa\% = 15$ of positive edges. Each node was represented by its regional BOLD time series and encoded by the temporal encoder before entering the ecospace-based attention blocks.

ABIDE I FC. The ABIDE I FC experiment followed the same graph-construction setting used for HCP-YA. Each subject was represented by a Pearson FC matrix extracted from resting-state fMRI data. Negative FC values were clipped to zero, the diagonal was removed, and the graph adjacency was obtained by retaining the strongest $\kappa\% = 15$ of positive FC edges. As in the HCP-YA setting, regional BOLD time series were used as node features and encoded through the temporal encoder before entering the ecospace-based attention blocks. Since ABIDE is a multi-site dataset, ComBat harmonization was applied before model training, with site treated as the batch variable and sex, age, and diagnosis included as covariates.

ABIDE I MC. For ABIDE I MC, the input graph was constructed from cortical MC matrix derived from mean curvature. Since the original MC matrix encodes pairwise dissimilarity, it was converted into a similarity matrix using the Gaussian kernel described in Appendix B. The resulting similarity

matrix was used as the BC matrix for the economy-aligned adjacency construction and as the node-feature matrix, with each node represented by its MC to all other regions.

Table C1: Main hyperparameters used across the three datasets.

Hyperparameter	HCP-YA	ABIDE I MC	ABIDE I FC
Distance decay constant λ	0.1	0.1	0.1
Model dimension d	128	64	32
Number of layers L	4	3	2
Attention heads H	8	4	4
Batch size	16	16	8
Optimizer	AdamW	AdamW	AdamW
Learning rate	10^{-3}	10^{-3}	10^{-4}
Weight decay	10^{-3}	10^{-3}	3×10^{-2}
LR scheduler	ReduceLROnPlateau	ReduceLROnPlateau	ReduceLROnPlateau
Scheduler factor	0.5	0.5	0.5
Scheduler patience	10	10	5
Early stopping patience	15	15	15
Selection metric	Validation AUC	Validation AUC	Validation AUC
Cross-validation	5-fold	5-fold	5-fold

D Baseline training and comparison protocol

To ensure a fair comparison between BGT-ESRE and the baseline models, all architectures were evaluated under a matched experimental protocol. Specifically, models were trained and tested using the same cross-validation folds, the same train/validation partitions within each fold, and the same evaluation metrics. For neural baselines, including Graphormer, SAN, GATv2, BRAINNETTF, BrainGNN, and ALTER, we used the original author implementations whenever available, adapting only the dataset interface, input formatting, and training wrapper required for compatibility with the present connectomics datasets. The core architectural components of each baseline were preserved. This protocol was designed to ensure that performance differences reflect the contribution of the proposed model, rather than differences in data splits, model-selection procedures, or evaluation criteria. Table D1 summarizes the common training and evaluation protocol used across all methods.

Table D1: Baseline fairness protocol. All methods were evaluated using the same cross-validation folds, validation metric, and held-out test partitions. Hyperparameters were selected using only training and validation data within each fold.

Model	Same folds	Same validation metric	Hyperparameter search	Early stopping
SVM	Yes	Yes	Yes	N/A
Random Forest	Yes	Yes	Yes	N/A
Graphormer	Yes	Yes	Yes	Yes
SAN	Yes	Yes	Yes	Yes
GATv2	Yes	Yes	Yes	Yes
BRAINNETTF	Yes	Yes	Yes	Yes
BrainGNN	Yes	Yes	Yes	Yes
ALTER	Yes	Yes	Yes	Yes
BGT-ESRE	Yes	Yes	Yes	Yes

E Economy-aligned graph metrics

Economy-aligned edge diffusion efficiency ($E_{\text{diff}}^{\text{eco}}$) The row-normalized transition matrix $P_{ij} = [\mathbf{A}^{\text{eco}}]_{ij}/s_i^{\text{eco}}$, where $s_i^{\text{eco}} = \sum_k [\mathbf{A}^{\text{eco}}]_{ik}$, defines a random walk whose stationary distribution $\pi_i = s_i^{\text{eco}} / \sum_k s_k^{\text{eco}}$ reflects economy-weighted node centrality. The fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1}$, where every row of $\mathbf{W}$ equals $\boldsymbol{\pi}$, yields the mean first-passage time $t_{ij} = (Z_{jj} - Z_{ij})/\pi_j$, and edge diffusion efficiency is $E_{\text{diff}}^{\text{eco}} = 1/(t_{ij} + \varepsilon)$. A high value indicates rapid random-walk spreading between i and j through economy efficient pathways, characteristic of short-range intra-module connections with strong BC and low anatomical cost [Seguin et al., 2023].

Economy-aligned edge routing efficiency ($E_{\text{rout}}^{\text{eco}}$) Edge weights are inverted to path lengths, $L_{ij} = 1/([\mathbf{A}^{\text{eco}}]_{ij} + \varepsilon)$, so that connections with high economy (high BC and low distance) correspond to short paths. All pairs shortest-path distances D_{ij} are obtained via Floyd-Warshall, and routing

efficiency is $E^{\text{eco}}_{\text{rout}} = 1/(D_{ij} + \varepsilon)$. A high value indicates efficient shortest-path communication between i and j through high-economy routes, characteristic of long-range inter-module connections that provide valuable routing shortcuts at justifiable metabolic cost [Latora and Marchiori, 2001]. Goñi et al. [Goñi et al., 2013] demonstrate that E_{diff} and E_{rout} are mathematically non-equivalent, establishing the routing-diffusion morphospace that motivates the present generalization.

Economy-aligned edge communicability ($C^{\text{eco}}_{\text{comm}}$) The communicability between nodes i and j is computed via the matrix exponential of the economy-aligned adjacency matrix:

$$C^{\text{eco}}_{\text{comm}} = [\exp(\mathbf{A}^{\text{eco}})]_{ij} = \sum_{k=0}^{\infty} \frac{([\mathbf{A}^{\text{eco}}]^k)_{ij}}{k!}. \tag{15}$$

By summing over walks of all possible lengths, penalized by the inverse factorial of the walk length, communicability captures broad-band integration across the network. A high value indicates that i and j are connected by a dense web of redundant, economy-efficient parallel pathways, reflecting robust communication channels that are highly resilient to the disruption of any single shortest path [Estrada and Hatano, 2008].

Economy-aligned edge clustering coefficient ($C^{\text{eco}}_{\text{cc}}$) For each edge $(i, j) \in E$:

$$C^{\text{eco}}_{\text{cc}} = \frac{\sum_k [\mathbf{A}^{\text{eco}}]_{ik} \, [\mathbf{A}^{\text{eco}}]_{jk}}{\min(s^{\text{eco}}_i, \, s^{\text{eco}}_j) + \varepsilon}. \tag{16}$$

A high value indicates that i and j share many strongly connected, anatomically cheap common neighbours, the hallmark of intra-module edges embedded in dense, triangulated local structure [Watts and Strogatz, 1998].

Economy-aligned edge betweenness centrality ($B^{\text{eco}}_{\text{bc}}$)

$$B^{\text{eco}}_{\text{bc}} = \sum_{s \neq t} \frac{\sigma_{st}(i, j)}{\sigma_{st}}, \tag{17}$$

where σ_{st} is the total number of economy-weighted shortest paths from s to t, and $\sigma_{st}(i, j)$ is the number passing through edge (i, j). A high value identifies critical long-range routing bridges through which a large fraction of economy-efficient inter-regional communication must flow [Girvan and Newman, 2002].

F Extended experimental results

Tables F1 and F2 report the complete experimental results. Both tables extends the main comparison and the ablation study results by including sensitivity and specificity.

Table F1: Complete performance comparison across all datasets. HCP-YA is used for sex classification, whereas ABIDE I MC and ABIDE I FC are used for ASD diagnosis classification. Results are reported as mean (std) across the five folds of a 5-fold cross validation. **Bold** indicates the highest result per column.

Category	Method	HCP-YA – Sex classification				ABIDE I MC – ASD classification				ABIDE I FC – ASD classification			
		ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.
ML Methods	SVM	0.568 (0.030)	0.558 (0.025)	0.289 (0.052)	0.805 (0.078)	0.566 (0.039)	0.531 (0.087)	0.252 (0.043)	0.827 (0.084)	0.533 (0.002)	0.472 (0.054)	0.006 (0.012)	**0.995 (0.010)**
	Random Forest	0.583 (0.019)	0.597 (0.030)	0.396 (0.030)	0.741 (0.055)	0.545 (0.061)	0.513 (0.073)	0.335 (0.063)	0.720 (0.079)	0.551 (0.038)	0.542 (0.041)	0.221 (0.051)	0.840 (0.075)
Graph Transformer Models	Graphormer	0.604 (0.030)	0.665 (0.044)	0.682 (0.124)	0.538 (0.106)	**0.578 (0.059)**	0.616 (0.075)	0.239 (0.217)	**0.860 (0.118)**	0.527 (0.033)	0.547 (0.014)	0.321 (0.358)	0.704 (0.371)
	SAN	0.517 (0.039)	0.535 (0.046)	0.411 (0.185)	0.607 (0.187)	0.537 (0.052)	0.599 (0.055)	**0.480 (0.112)**	0.585 (0.094)	0.458 (0.044)	0.447 (0.052)	0.595 (0.279)	0.340 (0.295)
	GATv2	0.525 (0.033)	0.530 (0.063)	0.510 (0.071)	0.538 (0.068)	0.551 (0.055)	0.572 (0.080)	0.423 (0.094)	0.660 (0.087)	0.523 (0.031)	0.540 (0.056)	0.500 (0.079)	0.543 (0.073)
Brain Graph Neural Networks	BRAINNETTF	0.666 (0.189)	0.789 (0.154)	0.568 (0.465)	0.748 (0.380)	0.540 (0.046)	0.601 (0.068)	0.430 (0.118)	0.632 (0.101)	**0.652 (0.037)**	0.687 (0.035)	**0.600 (0.091)**	0.696 (0.090)
	BrainGNN	0.642 (0.046)	0.710 (0.052)	0.610 (0.104)	0.669 (0.087)	0.558 (0.051)	0.607 (0.067)	0.410 (0.158)	0.681 (0.121)	0.623 (0.041)	0.665 (0.043)	0.548 (0.083)	0.688 (0.072)
	ALTER	0.591 (0.078)	0.571 (0.125)	0.249 (0.271)	**0.882 (0.103)**	0.507 (0.086)	0.503 (0.099)	0.432 (0.136)	0.570 (0.061)	0.606 (0.036)	0.651 (0.039)	0.529 (0.034)	0.672 (0.044)
Our Model	**BGT-ESRE**	**0.769 (0.045)**	**0.895 (0.012)**	**0.725 (0.211)**	0.805 (0.160)	**0.578 (0.059)**	**0.633 (0.044)**	0.439 (0.244)	0.695 (0.140)	**0.652 (0.040)**	**0.697 (0.034)**	0.498 (0.076)	0.784 (0.017)

Table F2: Complete ablation study of the BGT-ESRE architecture across the three datasets. We compare the full model against variants without ecospace rotation, with raw connectivity weighted ecospace, and with an alternative distance-decay parameter. Results are reported as mean (std) across the five folds of a 5-fold cross-validation. **Bold** indicates the highest result per column.

Model variant	HCP-YA – Sex classification				ABIDE I MC – ASD classification				ABIDE I FC – ASD classification			
	ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.
w/o Rotation	0.729 (0.013)	0.843 (0.024)	0.642 (0.175)	0.803 (0.141)	0.528 (0.036)	0.505 (0.106)	0.200 (0.400)	**0.800 (0.400)**	0.613 (0.018)	0.690 (0.033)	0.343 (0.130)	**0.844 (0.087)**
Raw connectivity w/ Rotation	0.749 (0.063)	0.873 (0.023)	0.694 (0.133)	0.795 (0.222)	0.531 (0.069)	0.553 (0.110)	**0.477 (0.331)**	0.575 (0.322)	0.633 (0.022)	0.686 (0.035)	**0.549 (0.010)**	0.706 (0.038)
$\lambda = 1/\bar{d}$ w/ Rotation	0.707 (0.045)	0.868 (0.018)	**0.865 (0.128)**	0.573 (0.179)	0.533 (0.097)	0.579 (0.096)	0.361 (0.151)	0.677 (0.213)	0.630 (0.024)	0.692 (0.034)	0.503 (0.069)	0.740 (0.052)
Full Model	**0.769 (0.045)**	**0.895 (0.012)**	0.725 (0.211)	**0.805 (0.160)**	**0.578 (0.059)**	**0.633 (0.044)**	0.439 (0.244)	0.695 (0.140)	**0.652 (0.040)**	**0.697 (0.034)**	0.498 (0.076)	0.784 (0.017)

G Ecospace configurations analysis

Figures G1–G6 and Table G1 jointly characterize the geometry of the ecospace and its relationship to classification performance across all ten measure pairs. Figures G1, G3, and G5 show the pairwise edge-level ecospace for HCP-YA, ABIDE I MC, and ABIDE I FC, respectively, where each point represents a single brain edge averaged over subjects in which it exists (prevalence $\geq 50\%$). Figures G2, G4, and G6 provide the complementary subject-level views, where each point corresponds to one subject and is obtained by averaging the edge-level ecospace coordinates across all prevalent edges in that subject's brain network.

On HCP-YA, the subject-level ecospace shows the clearest separation for pairs combining segregation- and integration-related measures, most notably $E^{\text{eco}}_{\text{diff}}/C^{\text{eco}}_{\text{comm}}$ and $E^{\text{eco}}_{\text{diff}}/E^{\text{eco}}_{\text{rout}}$. This pattern is consistent with Table G1, where the highest AUC values are obtained by measure pairs spanning diffusion, routing, and communicability. In contrast, all $C^{\text{eco}}_{\text{cc}}$ containing pairs achieve lower HCP-YA AUC values, suggesting that edge clustering coefficient contributes less discriminative information than communication-based measures in this setting. On ABIDE I MC, the separation between neurotypical control subjects and patients with ASD is subtler, consistent with ASD diagnosis being a harder classification problem than sex prediction. Nevertheless, performance is highest when communicability is included in the ecospace, with $C^{\text{eco}}_{\text{comm}}/B^{\text{eco}}_{\text{bc}}$ achieving the highest AUC and $E^{\text{eco}}_{\text{rout}}/C^{\text{eco}}_{\text{comm}}$ achieving the highest ACC. This suggests that broad-band multi-walk integration captures relevant morphological differences between neurotypical and ASD connectomes. The ABIDE I FC results in Figures G5 and G6 further support the relevance of diffusion-communicability structure in FC. In Table G1, $E^{\text{eco}}_{\text{diff}}/C^{\text{eco}}_{\text{comm}}$ achieves the highest ACC and AUC, while $E^{\text{eco}}_{\text{diff}}/E^{\text{eco}}_{\text{rout}}$ achieves the highest sensitivity. The performance spread across ecospace pairs is relatively small, indicating that several economy-aligned communication coordinates provide useful signal, but the strongest overall profile is obtained when diffusion efficiency is paired with communicability.

Comparing the ecospace variants with the raw FC-MC weighted ablation provides additional support for ***Hypothesis 2***. On HCP-YA, nine out of ten economy-aligned ecospace variants exceed the raw-connectivity variant in AUC. On ABIDE I MC, all ten variants improve over the raw-connectivity variant in AUC. On ABIDE I FC, the full economy-aligned variants show a smaller but consistent advantage in ACC/AUC over the raw-connectivity variant, with $E^{\text{eco}}_{\text{diff}}/C^{\text{eco}}_{\text{comm}}$ yielding the highest overall performances. These results indicate that distance-penalized economy coordinates encode information not fully captured by raw connectivity alone, although the size of this advantage depends on the dataset and modality.

Taken together, the ecospace analysis shows that BGT-ESRE is relatively robust to the specific choice of measure pair, while yields highest results when the selected coordinates capture complementary aspects of brain communication. In particular, diffusion, routing, and communicability emerge as the most informative dimensions across datasets, consistent with the view that local spreading, shortest-path communication, and multi-walk integration capture distinct properties of connectome organization. The consistent advantage of the full ESRE variants over the non-rotary ablation further supports ***Hypothesis 1*** and ***Hypothesis 3***: economy information is most useful when represented as an edge-level property and used to condition the interaction between brain regions, rather than being introduced only through additive node-level injection.

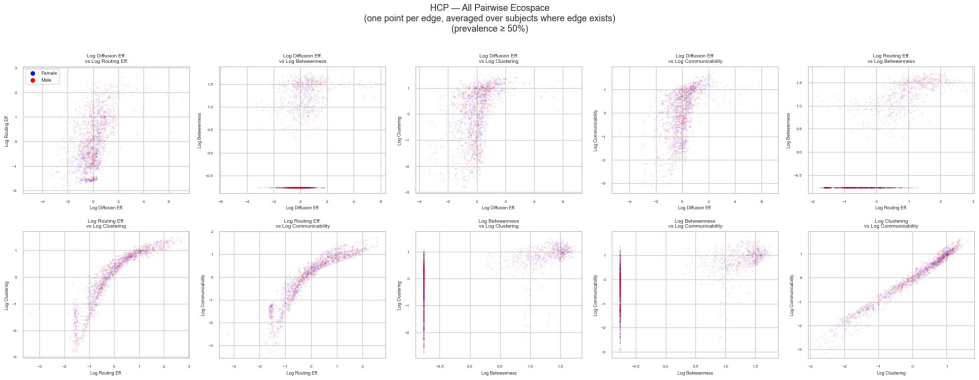


Figure G1: Pairwise ecospace of communication geometry for the HCP-YA dataset (one point per edge, averaged over subjects where the edge exists, prevalence $\geq 50\%$), computed with $\lambda = 0.1$. Each scatter plot shows the joint distribution of two economy-aligned measures across all shared edges.

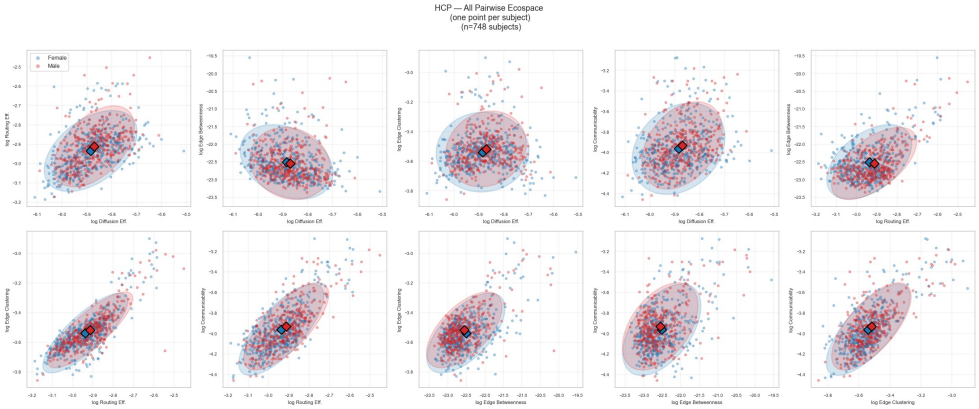


Figure G2: Pairwise ecospace of communication geometry for the HCP-YA dataset (one point per subject, $n = 748$), computed with $\lambda = 0.1$. Each point represents a single subject, coloured by biological sex (female/male), showing the subject-level distribution of economy aligned edge measures across all pairwise combinations.

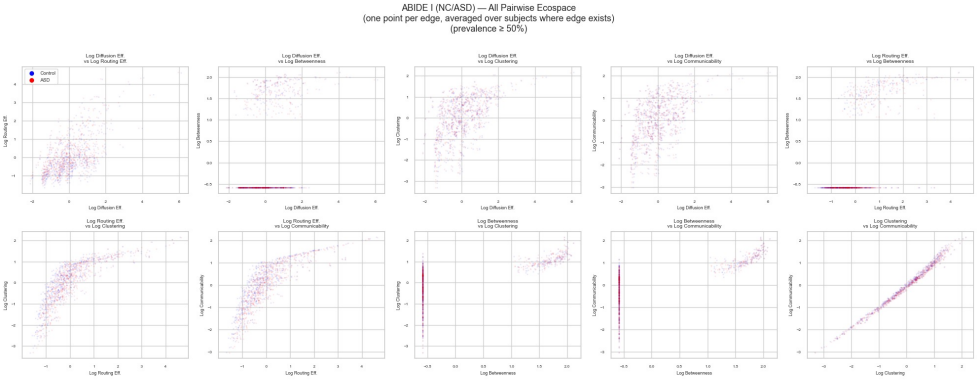


Figure G3: Pairwise ecospace of communication geometry for the ABIDE I MC dataset (one point per edge, averaged over subjects where the edge exists, prevalence $\geq 50\%$), computed with $\lambda = 0.1$. Each scatter plot shows the joint distribution of two economy-aligned measures across all shared edges.

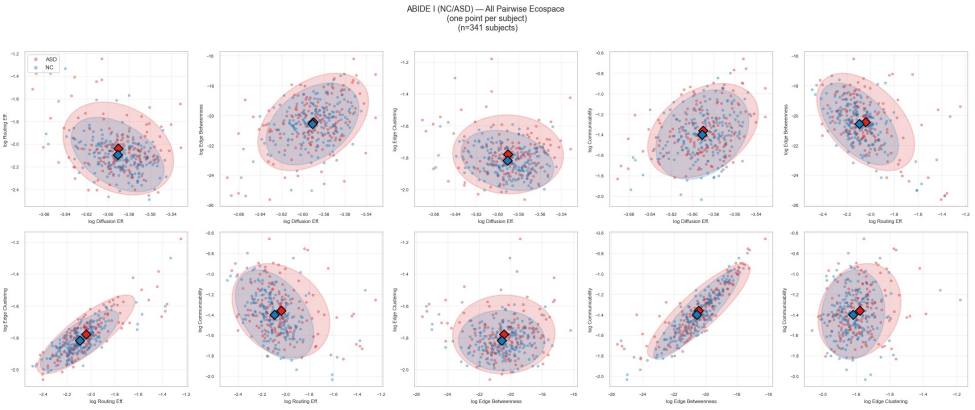


Figure G4: Pairwise ecospace of communication geometry for the ABIDE I MC dataset (one point per subject, $n = 341$), computed with $\lambda = 0.1$. Each point represents a single subject, coloured by diagnostic label (Control/ASD), showing the subject-level distribution of economy-aligned edge measures across all pairwise combinations.

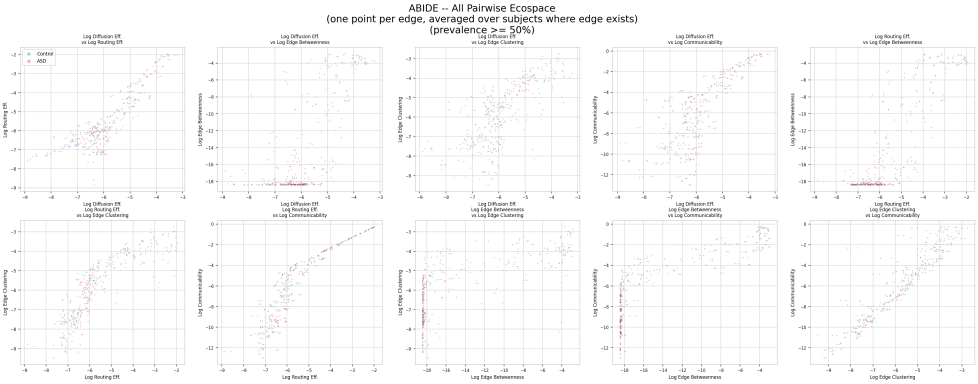


Figure G5: Pairwise ecospace of communication geometry for the ABIDE I FC dataset (one point per edge, averaged over subjects where the edge exists, prevalence $\geq 50\%$), computed with $\lambda = 0.1$. Each scatter plot shows the joint distribution of two economy-aligned measures across all shared edges.

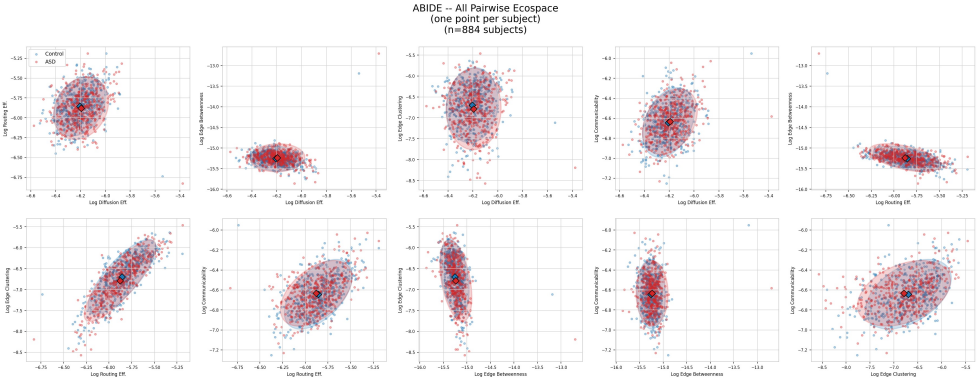


Figure G6: Pairwise ecospace of communication geometry for the ABIDE I FC dataset (one point per subject, $n = 884$), computed with $\lambda = 0.1$. Each point represents a single subject, coloured by diagnostic label (Control/ASD), showing the subject-level distribution of economy-aligned edge measures across all pairwise combinations.

Table G1: Performance comparison of ten distinct ecospace variants across the three datasets. Results are reported as mean (std) across the five folds of a 5-fold cross-validation. **Bold** indicates the highest result per column.

Category	Edge Types	HCP-YA – Sex classification				ABIDE I MC – ASD classification				ABIDE I FC – ASD classification			
		ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.	ACC	AUC	Sen.	Spe.
Ablation	w/o Rotation	0.729 (0.013)	0.843 (0.024)	0.642 (0.175)	0.803 (0.141)	0.528 (0.036)	0.505 (0.106)	0.200 (0.400)	0.800 (0.400)	0.613 (0.018)	0.690 (0.033)	0.343 (0.130)	**0.844 (0.087)**
	Raw connectivity w/ Rotation	0.749 (0.063)	0.873 (0.023)	0.694 (0.133)	0.795 (0.222)	0.531 (0.069)	0.553 (0.110)	**0.477 (0.331)**	0.575 (0.322)	0.633 (0.022)	0.686 (0.035)	0.549 (0.010)	0.706 (0.038)
BGT-ESRE ($\lambda = 0.1$)	$E^{\text{eco}}_{\text{diff}}$ / $E^{\text{eco}}_{\text{rout}}$	0.762 (0.097)	0.893 (0.015)	0.599 (0.273)	**0.899 (0.070)**	0.542 (0.046)	0.567 (0.092)	0.329 (0.302)	0.721 (0.281)	0.624 (0.037)	0.688 (0.034)	**0.586 (0.058)**	0.657 (0.098)
	$E^{\text{eco}}_{\text{diff}}$ / $C^{\text{eco}}_{\text{comm}}$	0.769 (0.045)	**0.895 (0.012)**	0.725 (0.211)	0.805 (0.160)	0.557 (0.046)	0.614 (0.065)	0.445 (0.281)	0.650 (0.279)	**0.652 (0.040)**	**0.697 (0.034)**	0.498 (0.076)	0.784 (0.017)
	$E^{\text{eco}}_{\text{diff}}$ / $B^{\text{eco}}_{\text{bc}}$	**0.773 (0.021)**	0.886 (0.013)	0.676 (0.120)	0.854 (0.102)	0.563 (0.055)	0.612 (0.108)	0.239 (0.285)	0.835 (0.183)	0.609 (0.045)	0.668 (0.045)	0.488 (0.141)	0.712 (0.106)
	$E^{\text{eco}}_{\text{rout}}$ / $C^{\text{eco}}_{\text{comm}}$	0.735 (0.071)	0.885 (0.020)	0.602 (0.281)	0.847 (0.137)	**0.589 (0.059)**	0.629 (0.047)	0.348 (0.296)	0.791 (0.146)	0.627 (0.020)	0.691 (0.035)	0.500 (0.070)	0.735 (0.054)
	$E^{\text{eco}}_{\text{rout}}$ / $B^{\text{eco}}_{\text{bc}}$	0.755 (0.029)	0.873 (0.021)	**0.822 (0.149)**	0.699 (0.173)	0.572 (0.066)	0.579 (0.098)	0.226 (0.208)	0.861 (0.151)	0.627 (0.052)	0.678 (0.046)	0.493 (0.082)	0.742 (0.077)
	$C^{\text{eco}}_{\text{cc}}$ / $E^{\text{eco}}_{\text{rout}}$	0.728 (0.096)	0.887 (0.016)	0.667 (0.301)	0.780 (0.214)	0.528 (0.099)	0.569 (0.106)	0.310 (0.133)	0.710 (0.238)	0.636 (0.022)	0.687 (0.036)	0.556 (0.037)	0.704 (0.027)
	$C^{\text{eco}}_{\text{cc}}$ / $E^{\text{eco}}_{\text{diff}}$	0.733 (0.085)	0.883 (0.016)	0.636 (0.298)	0.815 (0.178)	0.572 (0.053)	0.595 (0.069)	0.258 (0.239)	0.835 (0.145)	0.623 (0.022)	0.691 (0.035)	0.498 (0.068)	0.731 (0.042)
	$C^{\text{eco}}_{\text{cc}}$ / $C^{\text{eco}}_{\text{comm}}$	0.739 (0.078)	0.871 (0.028)	0.731 (0.204)	0.746 (0.226)	0.578 (0.032)	0.622 (0.074)	0.181 (0.167)	**0.910 (0.111)**	0.635 (0.036)	0.671 (0.049)	0.505 (0.090)	0.746 (0.070)
	$C^{\text{eco}}_{\text{cc}}$ / $B^{\text{eco}}_{\text{bc}}$	0.746 (0.037)	0.885 (0.016)	0.595 (0.196)	0.874 (0.131)	0.581 (0.052)	0.610 (0.076)	0.355 (0.301)	0.770 (0.170)	0.628 (0.051)	0.678 (0.045)	0.510 (0.101)	0.729 (0.091)
	$C^{\text{eco}}_{\text{comm}}$ / $B^{\text{eco}}_{\text{bc}}$	0.742 (0.093)	0.889 (0.012)	0.599 (0.329)	0.862 (0.122)	0.578 (0.059)	**0.633 (0.044)**	0.439 (0.244)	0.695 (0.140)	0.630 (0.025)	0.686 (0.035)	0.554 (0.065)	0.695 (0.031)

H Economy-aligned edge weight formulation

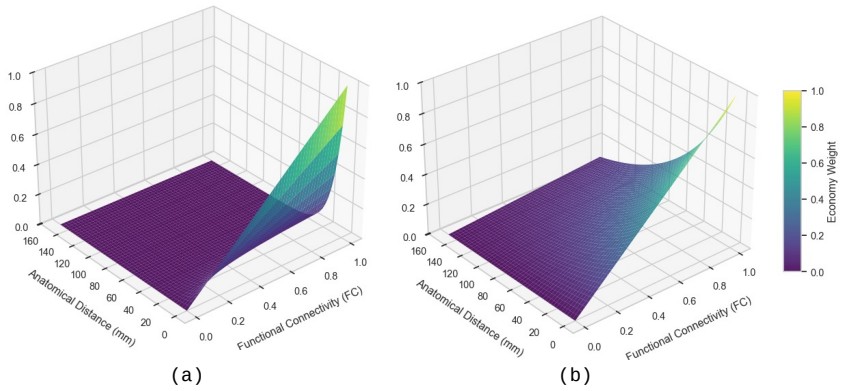


Figure H1: The economy-aligned edge formulation evaluated across distinct distance decay constants: (**a**) $\lambda = 0.1$, and (**b**) $\lambda = 1/\bar{d}$, where $\bar{d}$ denotes the mean anatomical distance among all regions.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect our paper's contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Our limitations are discussed in Section 5 (Discussions and Conclusions).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Theoretical results are reported and proved in Section 3 (Method).

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We detailed our training and evaluation protocols in the experimental setup section and in the appendices.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We used a public available dataset and the code is available on a public repository on GitHub.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We detailed our training and evaluation protocols in the experimental setup section and in the appendices.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All results are the average of 5 random runs on test sets with the standard deviation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The computer resources can be referred to in the experimental setup section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our research is conducted in compliance with the NeurIPS Code of Ethics and our experiments do not involve human subjects and potential harmful consequences for society.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This is not applicable as it is foundational basic research.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper properly credits the creators or original owners of the assets used such as code, data, and models; we also have correctly cited the relevant literature.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code is available in a public repository on GitHub.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper relies on previously collected, public neuroimaging and transcriptomic datasets. No new data gathering was conducted.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper relies on previously collected, public neuroimaging and transcriptomic datasets. No new data gathering was conducted.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: We only use LLMs for polishing and formatting this paper, without affecting the research methods.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.