TOPOSCANPE: TOPOLOGICAL POSITIONAL ENCODINGS FOR BRAIN NETWORK TRANSFORMERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Brain regions can have similar functional connectivity strengths while playing different roles within the connectome. To capture these differences, we introduce *TopoScanPE*, a positional encoding that represents each region of interest (ROI) through the topology of its local neighborhood. For every ROI we extract its local neighborhood from a sparse connectivity graph and apply Topo-Scan (Uddin et al., 2026), a sliding-window filtration over complementary geometric node functions that summarizes each slice by the topology of its clique complex. A shared encoder maps the resulting Topo-Scan (Uddin et al., 2026) sequences into a fixed-width code that is concatenated to the ROI token. Experiments on ABIDE and ADNI show that TopoScanPE demonstrates that it is competitive with state-of-the-art brain graph transformers and graph neural networks.

1 INTRODUCTION

Network-based representations of neuroimaging data enable the topology of functional brain organization to be examined quantitatively (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010; Fornito et al., 2016). Under this formulation, the brain is parcellated into regions of interest (ROIs) according to an anatomical or functional atlas, and each ROI is represented as a node in a graph. Edges encode estimated relationships between regional signals. For functional magnetic resonance imaging (fMRI), these relationships are commonly derived by extracting the blood oxygenation level-dependent (BOLD) signal from each ROI and computing pairwise Pearson correlations, which yields a functional connectivity (FC) matrix (Smith et al., 2011; Simpson et al., 2013). Analyses of these networks have identified condition-related differences in connectivity and topological organization, including altered network structure in autism spectrum disorder (Rudie et al., 2013), Alzheimer's disease (Stam et al., 2007), and schizophrenia (Liu et al., 2008). Such findings motivate computational models that can identify distributed connectivity patterns associated with diagnostic status and other clinically relevant outcomes. Graph neural networks were among the first deep learning models applied to this problem (Li et al., 2021; Cui et al., 2023), and transformers have since been widely adopted (Kan et al., 2022b; Yu et al., 2024; Yang et al., 2025): each ROI is represented by its connectivity profile, and self-attention learns which interactions between regions are informative for the diagnosis at hand.

Self-attention, however, is permutation-equivariant (Vaswani et al., 2017). It cannot tell which region a token represents on its own, and graph transformers consequently rely on positional encodings to inject structural identity into their inputs (Dwivedi et al., 2022; Rampášek et al., 2022). Brain network transformers usually take two routes. Some omit explicit encodings entirely and let each region's connection profile serve as its token and implicit identity (Kan et al., 2022b), while others derive positional encodings from the same FC matrix that already forms the tokens. Either way, the structural signal available to the model reflects how strongly a region is connected, but says little about *how its neighborhood is organized*: whether its neighbors form a tightly clustered module, whether the region bridges otherwise separate groups, or whether its connections enclose cycles that no shorter path completes. Such organizational properties are precisely what graph-theoretical studies of the brain have found to be disrupted in disease, from the loss of small-world organization in Alzheimer's disease (Supekar et al., 2008) to altered hub and module structure more broadly (Bullmore & Sporns, 2009). Two regions with comparable connection strengths can sit in very differently organized neighborhoods, and an encoding blind to that difference withholds from the model the very information that separates them.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Topological data analysis provides a mathematical framework for characterizing these structural patterns. Persistent homology and related tools summarize connected components, cycles and higher-order cavities across scales, and have been used to characterize functional brain networks (Lee et al., 2012; Sizemore et al., 2018). In deep learning, however, topological summaries are usually computed at the graph level and fused with a learned representation near the readout, as an additional, graph-level feature. This design discards where in the network a topological feature arises, and asks a weak side signal to compete with a strong backbone at the final layer. Topo-Scan (Uddin et al., 2026) proposes a different summary: rather than a single filtration, it scans a node or edge function with sliding windows and records the topology of each slice, producing an ordered sequence of topological descriptors that a transformer can process directly.

Based on this idea, we introduce TOPOSCANPE, a topological positional encoding for brain network transformers. For every ROI, we extract its local neighborhood from a sparse connectivity graph and apply Topo-Scan with two complementary geometric filtrations, Forman-Ricci curvature and the heat kernel signature, describing each slice by the Betti numbers of its clique complex together with its size. A small encoder shared across regions maps each resulting sequence to a fixed-width code, which is concatenated to the region's token in place of a conventional positional encoding. Topology thus enters the model at its input, where it can shape attention between regions, rather than at its output, where it can only adjust a prediction. The encoding leaves the backbone, objective and training protocol unchanged.

Our contributions are as follows:

- We introduce TOPOSCANPE, a positional encoding that characterizes each region of a brain network by the topology of its local connectivity rather than by the strength of its correlations.
- We adapt Topo-Scan to region neighborhoods with two complementary filtrations, Forman-Ricci curvature and heat kernel signature.
- We evaluate TOPOSCANPE on ABIDE and ADNI, demonstrating strong performance across brain network classification tasks.

2 RELATED WORK

Brain network analysis. Network neuroscience uses graph representations to characterize the organization of structural and functional brain connectivity (Bullmore & Sporns, 2009; Fornito et al., 2016). Classical approaches examine small-world structure, modular organisation, hubs and rich-club organisation, and network motifs (Bassett & Bullmore, 2006; Rubinov & Sporns, 2010; Meunier et al., 2009; van den Heuvel & Sporns, 2011; Sporns & Kötter, 2004). Community detection and comparisons with null models provide further tools for studying how these properties vary across subjects and conditions (Garcia et al., 2018; Váša & Mišić, 2022). Predictive modeling builds on this foundation by learning directly from connectivity. Connectome-based predictive modeling (Shen et al., 2017) and kernel regression (He et al., 2020) operate on vectorized edge weights and remain strong baselines; He et al. (2020) found that deep networks did not outperform kernel regression for predicting behavior from functional connectivity. BrainNetCNN (Kawahara et al., 2017) applies edge-to-edge and edge-to-node convolutions to connectivity matrices. More recently, BQN (Yang et al., 2025) replaces message passing with quadratic operations on the connectivity matrix, and BrainNetMLP (Hou et al., 2025) shows that a pure MLP over spatial and spectral connectivity features matches more complex architectures. Surveys of graph learning in neuroimaging (Ahmedt-Aristizabal et al., 2021; Bessadok et al., 2023) and recent benchmarks (Cui et al., 2023; Xu et al., 2023) document the range of tasks, representations, and evaluation practices in this area.

GNNs for brain networks. Message-passing GNNs combine node features with graph structure to learn representations (Gilmer et al., 2017; Kipf & Welling, 2017b; Hamilton et al., 2017). Brain-specific designs adapt their aggregation and training objectives to neuroimaging data. BrainGNN (Li et al., 2021) introduces ROI-aware convolutions; GroupINN (Yan et al., 2019) learns groups of regions; and FBNetGen (Kan et al., 2022a) jointly learns functional networks and a predictor. IBGNN (Cui et al., 2022) targets interpretable disorder-specific connections, while A-GCL (Zhang et al., 2023a) uses adversarial graph contrastive learning. MVS-GCN (Wen et al., 2022) combines

multiple views of connectivity with a structural prior, and STAGIN (Kim et al., 2021) models dynamic brain networks through spatiotemporal attention.

Brain transformers BrainNetTF (Kan et al., 2022b) shows that connection profiles provide useful positional information when ROI ordering is consistent across subjects. Its transformer learns interactions among these profiles and pools the resulting representations through orthonormal clustering. Subsequent work organizes attention around groups of regions: Com-BrainTF (Bannadabhavi et al., 2023) applies local attention within predefined functional communities before global attention across them, and THC (Dai et al., 2023) learns a hierarchical clustering of regions jointly with classification. ALTER (Yu et al., 2024) supplements the profiles with correlation-weighted walk encodings to represent connectivity over multiple steps. Other brain transformers introduce longitudinal representations (Dong et al., 2023), biologically motivated structural components (Peng et al., 2025), or multiscale representations (Shehzad et al., 2025).

3 PRELIMINARIES

Brain network construction. A subject's brain is parcellated into N regions of interest (ROIs) according to a fixed atlas, and each ROI u is summarized by the mean BOLD signal of its voxels, a time series $\mathbf{t}_u \in \mathbb{R}^T$. Functional connectivity between two ROIs is measured by the Pearson correlation of their signals,

$$X_{uv} = \frac{\sum_{\tau=1}^{T} \big(t_u(\tau) - \bar{t}_u\big)\big(t_v(\tau) - \bar{t}_v\big)}{\sqrt{\sum_{\tau=1}^{T} \big(t_u(\tau) - \bar{t}_u\big)^2}\,\sqrt{\sum_{\tau=1}^{T} \big(t_v(\tau) - \bar{t}_v\big)^2}}, \tag{1}$$

where $\bar{t}_u$ is the temporal mean of $\mathbf{t}_u$. The resulting symmetric matrix $\mathbf{X} \in [-1, 1]^{N \times N}$ defines a weighted brain network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ with vertex set $\mathcal{V} = \{v_1, \ldots, v_N\}$, one vertex per ROI.

Clique complexes and Betti numbers. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, its clique complex $\widehat{\mathcal{G}}$ contains a k-simplex for every $(k+1)$-clique of $\mathcal{G}$. The Betti numbers of $\widehat{\mathcal{G}}$, computed over $\mathbb{Z}/2$, count its topological features: β_0 the connected components and β_1 the independent cycles not filled by triangles (Aktas et al., 2019).

Topo-Scan. Sublevel filtrations of a graph tend to saturate early: once most vertices enter at low thresholds, larger-scale structure goes unrecorded. Topo-Scan (Uddin et al., 2026) avoids this by slicing rather than accumulating. Let

$f\colon \mathcal{V} \to \mathbb{R}$ be a vertex function with thresholds $\alpha_0 = \min_v f(v) < \alpha_1 < \cdots < \alpha_T = \max_v f(v)$, and let m be a window width. For $i = 0, \ldots, T - m$, the i-th slice is the induced subgraph

$$\mathcal{G}_i = \mathcal{G}\big[\{\, v \in \mathcal{V} \mid \alpha_i \le f(v) \le \alpha_{i+m} \,\}\big], \tag{2}$$

and the sequence of clique complexes $\{\widehat{\mathcal{G}}_i\}$ is called a *slicing*.

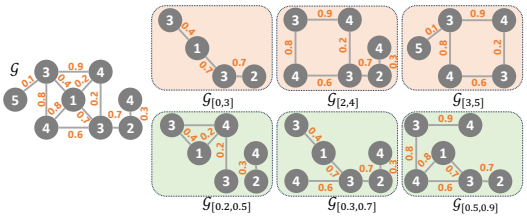


Figure 1: Topo-Scan decomposes a graph into sequential topological slices via node and edge filtrations. The top row shows node-based filtrations; the bottom row, edge-based ones. Adapted from (Uddin et al., 2026).

Transformers on graphs. For a token matrix $\mathbf{H} \in \mathbb{R}^{N \times d}$, a single attention head computes

$$\mathrm{Att}(\mathbf{H}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \tag{3}$$

where $\mathbf{Q} = \mathbf{H}\mathbf{W}_Q$, $\mathbf{K} = \mathbf{H}\mathbf{W}_K$, and $\mathbf{V} = \mathbf{H}\mathbf{W}_V$ are learned projections; d_k is the query and key dimension; and softmax is applied row-wise (Vaswani et al., 2017). Transformer layers combine multiple heads with a feed-forward network, residual connections, and layer normalization. Without positional information or an order-dependent mask, self-attention is permutation-equivariant: permuting the input tokens permutes the outputs in the same way. Graph transformers can incorporate structural information through token features or attention biases.

4 METHOD

TOPOSCANPE augments each ROI's connection profile with a positional encoding derived from the topology of its local neighborhood. For each ROI, we extract a small subgraph from a sparse connectivity graph (Section 4.1), apply Topo-Scan with two node functions to obtain its topological signature (Section 4.2), and map this signature to a positional encoding with a shared encoder (Section 4.3). Concatenating the encoding with the connection profile gives the ROI's token. Figure 3 summarizes the pipeline.

4.1 LOCAL NEIGHBORHOODS

We first sparsify $\mathbf{X}$ into an unweighted, undirected k-nearest-neighbor graph $\mathcal{G} = (\mathcal{V}, E)$ on positive correlations: $(u, v) \in E$ if $X_{uv} > 0$ and v is among the k strongest correlates of u, or vice versa. For each ROI u we take its two-hop neighborhood $\mathcal{N}_2(u) = \{u\} \cup \mathcal{N}(u) \cup \mathcal{N}(\mathcal{N}(u))$ and bound its size by a neighbor budget $B = \lfloor \rho N \rfloor$, so that the induced subgraph $G_u = G[\mathcal{V}_u]$ has at most $B + 1$ vertices. When $|\mathcal{N}_2(u)| > B + 1$, vertices are admitted in order of proximity: all one-hop neighbors first, then two-hop neighbors ranked by their degree within $\mathcal{N}_2(u)$, with ties broken deterministically.

Figure 2 illustrates the construction for ROI u with a budget of $B = 5$ neighbors. The two-hop ball $\mathcal{N}_2(u)$ contains three one-hop neighbors (ROIs 1–3) and five two-hop neighbors (ROIs 4–8) All one-hop neighbors are retained, and the two remaining slots are assigned to the two-hop neighbors with the most edges into the ball: ROIs 4 and 5, with three such edges each, are kept, whereas ROIs 6 and 8, with two, and ROI 7, with one, are discarded. The budget keeps each description local even when a strong global signal would otherwise let the two-hop ball span most of the brain.

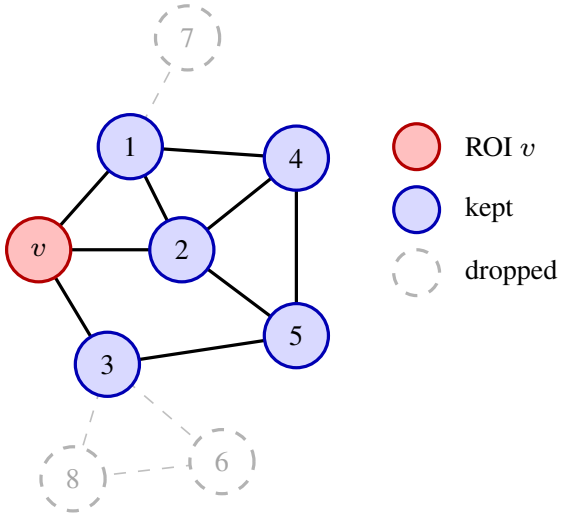


Figure 2: neighborhood extraction for ROI u with budget $B = 5$.

4.2 TOPO-SCAN OF A NEIGHBORHOOD

We summarize each neighborhood with Topo-Scan (Uddin et al., 2026), a sliding-window filtration. For a node function f on $\mathcal{G}_u$, let $\alpha_1 < \cdots < \alpha_Q$ be the distinct values among the quantiles of $\{f(v)\}_{v \in \mathcal{G}_u}$. For $i = 1, \ldots, Q - m = L$, where m the window width, the i-th slice is the subgraph induced by the vertices whose values fall in an overlapping window,

$$\mathcal{G}_u^{(i)} = \mathcal{G}_u[\{\, v : \alpha_i \le f(v) \le \alpha_{i+m} \,\}]\,. \tag{4}$$

Each slice is summarized by $\mathbf{s}^{(i)} = \big(\beta_0,\ \beta_1,\ |V^{(i)}|,\ |E^{(i)}|\big)$, where β_0 and β_1 are the Betti numbers of the clique complex of $\mathcal{G}_u^{(i)}$ over $\mathbb{Z}/2$.

Applying Topo-Scan with both filtrations yields, for every ROI, a topological signature $\mathbf{S}_u \in \mathbb{R}^{L \times 8}$: an ordered sequence of L slices, each described by four Forman-based and four heat-kernel-based quantities. Since the slice counts are heavy-tailed and differ in scale across features, each entry is log-transformed and standardized as

$$\tilde{S}_{u,i,j} = \frac{\log\big(1 + S_{u,i,j}\big) - \mu_{i,j}}{\sigma_{i,j}}, \tag{5}$$

where $\mu_{i,j}$ and $\sigma_{i,j}$ are the mean and standard deviation of $\log(1 + S_{u,i,j})$ for slice i and feature j, computed over all ROIs of the training subjects only.

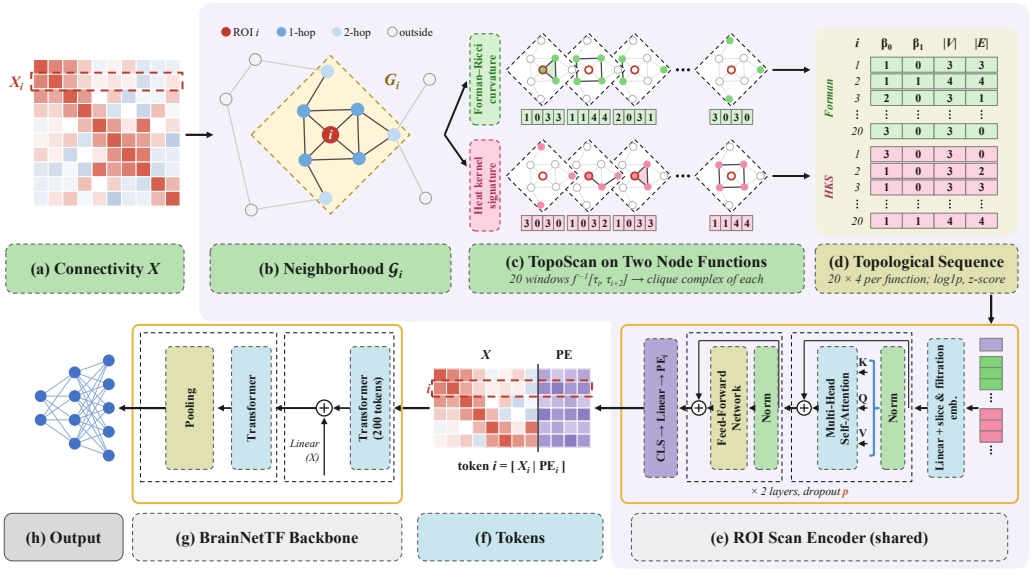


Figure 3: TopoScanPE. (a) Each subject is represented by a functional connectivity matrix $\mathbf{X}$, with one row per ROI. (b) For each ROI i, a two-hop neighborhood $\mathcal{G}_i$ is extracted from the k-nearest-neighbor graph. (c) Topo-Scan applies 20 sliding windows to two node functions: Forman–Ricci curvature and the heat kernel signature. (d) Each topological slice is represented by $(\beta_0, \beta_1, |V|, |E|)$, producing one descriptor sequence per node function. (e) A lightweight transformer shared across all ROIs maps these sequences to the positional encoding PE_i. (f) The encoding is concatenated with the connection profile $\mathbf{X}_i$ to form the token for ROI i. (g) The resulting tokens are processed by the unchanged BrainNetTF backbone and clustering readout. (h) The model produces the subject-level prediction.

4.3 The TopoScanPE encoder

The encoder turns the signature $\mathbf{S}_u \in \mathbb{R}^{L\times 8}$ of each ROI into a positional encoding $\mathbf{p}_u \in \mathbb{R}^{d_{\mathrm{pe}}}$. It is a small transformer whose weights are shared by all ROIs and subjects. We split $\mathbf{S}_u$ into its Forman and heat-kernel sequences, each of L slices described by four numbers $\mathbf{s}_{u,c}^{(i)} \in \mathbb{R}^4$, $c \in \{\mathrm{F}, \mathrm{H}\}$, and map every slice to a token

$$\mathbf{h}_{u,c,i} = \mathbf{W}_{\mathrm{in}}\, \mathbf{s}_{u,c}^{(i)} + \mathbf{e}_i + \mathbf{f}_c \in \mathbb{R}^d, \tag{6}$$

where $\mathbf{W}_{\mathrm{in}} \in \mathbb{R}^{d\times 4}$ is a shared linear embedding, $\mathbf{e}_i$ a learned embedding of the slice position, which attention alone cannot perceive, and $\mathbf{f}_c$ a learned embedding of the filtration. The $2L$ tokens are preceded by a learned summary token $\mathbf{c} \in \mathbb{R}^d$ (Devlin et al., 2019),

$$\mathbf{H}_u^{(0)} = \big[\, \mathbf{c};\ \mathbf{h}_{u,\mathrm{F},1}; \ldots; \mathbf{h}_{u,\mathrm{F},L};\ \mathbf{h}_{u,\mathrm{H},1}; \ldots; \mathbf{h}_{u,\mathrm{H},L} \,\big] \in \mathbb{R}^{(2L+1)\times d}, \tag{7}$$

and processed by a pre-norm transformer encoder, $\mathbf{H}_u = \mathrm{Enc}\big(\mathbf{H}_u^{(0)}\big)$, so that each slice can attend to the others along the same filtration and to those of the other filtration. The summary token carries no data of its own, so its output aggregates all slices with learned attention weights. The positional encoding is its linear projection,

$$\mathbf{p}_u = \mathbf{W}_{\mathrm{pe}}\, \mathbf{W}_{\mathrm{out}}\, \mathbf{H}_u[0], \tag{8}$$

with $\mathbf{W}_{\mathrm{out}} \in \mathbb{R}^{d_{\mathrm{tok}}\times d}$ and $\mathbf{W}_{\mathrm{pe}} \in \mathbb{R}^{d_{\mathrm{pe}}\times d_{\mathrm{tok}}}$. Since the encoder is shared, ROIs with identical signatures receive the same code.

4.4 Integration into a brain network transformer

The token matrix concatenates the connection profiles and the positional encodings,

$$\mathbf{Z}^{(0)} = \big[\, \mathbf{X} \,\|\, \mathrm{PE} \,\big] \in \mathbb{R}^{N\times(N+d_{\mathrm{PE}})}. \tag{9}$$

The backbone is BrainNetTF (Kan et al., 2022b): two attention layers followed by its orthonormal clustering readout. Writing Att for an attention block and Pool for the orthonormal cluster-assignment pooling (Xie et al., 2016),

$$\mathbf{Z}^{(1)} = \mathrm{Att}_1\big(\mathbf{Z}^{(0)}\big) + \mathbf{X}\mathbf{W}_{\mathrm{res}}, \tag{10}$$

$$\mathbf{Z}^{(2)} = \mathrm{Pool}\big(\mathrm{Att}_2\big(\mathbf{Z}^{(1)}\big)\big). \tag{11}$$

A linear projection of $\mathbf{x}_u$ is added as a residual after the first layer. The pooled representations are reduced per token, flattened, and passed through an MLP classifier. The encoding adds only the shared signature encoder to the backbone, and requires no change to its objective, attention mechanism or readout. It can therefore be attached to any transformer that represents ROIs as tokens.

5 Experiments

5.1 Experimental Setup

Datasets We evaluate TopoScanPE on two public resting-state fMRI datasets. The *Autism Brain Imaging Data Exchange* (ABIDE) (Di Martino et al., 2014) aggregates resting-state fMRI, anatomical and phenotypic data from 17 international sites, and is widely used to study functional connectivity alterations in autism spectrum disorder. We use the preprocessed release of the Preprocessed Connectomes Project [1], comprising 1,009 subjects: 516 with autism spectrum disorder (ASD) and 493 typically developing controls (TC). Brain regions are defined by the Craddock 200 atlas (Craddock et al., 2012), giving $N = 200$ ROIs.

The *Alzheimer's Disease Neuroimaging Initiative* (ADNI)[2] is a longitudinal, multi-site neuroimaging study of Alzheimer's disease; access to its images requires an approved data use application. We select resting-state scans. The resulting cohort contains 53 participants with Alzheimer's disease (AD) and 71 cognitively normal controls (CN). Following Yu et al. (2024), images are preprocessed with DPARSF (Yan & Zang, 2010). Brain regions are defined by the AAL atlas (Tzourio-Mazoyer et al., 2002), giving $N = 90$ ROIs.

For both datasets, ROI time series are obtained by averaging the BOLD signal within each region, and the functional connectivity matrix $\mathbf{X}$ is computed as the Pearson correlation between every pair of ROI time series (Section 3).

Baselines We compare TopoScanPE against fourteen baselines in three groups. The first group comprises graph neural networks: the general-purpose GCN (Kipf & Welling, 2017a) and GAT (Veličković et al., 2018), and four models designed for brain networks, BrainGNN (Li et al., 2021), BrainGB (Cui et al., 2023), FBNetGen (Kan et al., 2022a) and A-GCL (Zhang et al., 2023b). The second group comprises graph transformers: the general-purpose SAN (Kreuzer et al., 2021), Graphormer (Ying et al., 2021b) and GraphTrans (Wu et al., 2021b), and four brain-specific transformers, BrainNetTF (Kan et al., 2022b), ContrastPool (Xu et al., 2024b), ALTER (Yu et al., 2024) and BioBGT (Peng et al., 2025). The third group is BQN (Yang et al., 2025), which replaces message passing with quadratic neurons and is the current state of the art on both datasets. BrainNetTF shares our backbone and thus isolates the effect of the positional encoding.

Evaluation metrics We report four metrics that together cover discriminative power and the clinical costs of each type of error. Treating the patient group (ASD or AD) as positive and writing TP, TN, FP and FN for the counts of true positives, true negatives, false positives and false negatives, we use

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \qquad \mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \qquad \mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{12}$$

together with the area under the receiver operating characteristic curve (AUC). All metrics are reported as the mean and standard deviation over five runs.

[1] http://preprocessed-connectomes-project.org/abide/

[2] https://adni.loni.usc.edu/

Table 1: **Classification performance on ABIDE and ADNI.** Performance comparison with GNN-based, transformer-based and quadratic-based baselines. The best and second-best results in each column are shown in **bold** and underlined, respectively.

		ABIDE				ADNI			
Type	Model	AUC↑	ACC↑	SEN↑	SPE↑	AUC↑	ACC↑	SEN↑	SPE↑
GNN	GCN (Kipf & Welling, 2017b)	59.59±3.44	59.30±3.38	56.67±4.37	61.55±5.29	62.45±3.63	59.12±4.53	54.55±9.96	62.00±8.55
	GAT (Veličković et al., 2018)	60.43±3.88	60.10±4.13	59.26±5.51	62.89±8.03	62.00±2.88	58.75±2.86	53.20±7.50	64.79±7.65
	BrainGNN (Li et al., 2021)	64.42±3.57	63.09±1.35	65.65±2.88	60.67±3.68	61.81±1.58	58.72±3.14	52.88±9.70	62.70±4.26
	BrainGB (Cui et al., 2023)	70.32±3.66	65.12±3.90	67.01±10.00	60.07±8.53	66.44±3.33	63.70±4.65	60.73±8.25	64.67±9.07
	FBNetGen (Kan et al., 2022a)	74.55±3.77	67.09±3.37	64.71±9.85	69.61±9.30	67.05±2.16	63.26±1.38	66.79±6.93	61.31±9.65
	A-GCL (Zhang et al., 2023a)	73.86±2.91	71.04±2.40	71.42±3.03	70.95±3.19	68.15±3.33	67.08±4.57	65.45±8.39	68.33±4.36
GT	SAN (Kreuzer et al., 2021)	71.35±2.18	65.34±2.91	55.41±9.29	68.39±7.50	66.11±3.41	61.78±4.22	53.94±7.56	63.63±8.51
	Graphormer (Ying et al., 2021a)	63.91±4.05	61.88±6.85	66.30±9.98	55.74±11.00	60.69±5.26	55.75±3.18	60.18±11.36	47.75±13.53
	GraphTrans (Wu et al., 2021a)	60.13±6.73	57.83±4.71	65.70±10.30	49.77±11.52	61.41±3.65	58.60±5.41	65.57±6.05	54.37±3.42
	BrainNetTF (Kan et al., 2022b)	77.93±1.41	69.26±2.26	65.92±8.60	73.20±6.06	69.73±2.61	67.85±2.92	63.64±6.27	70.67±8.33
	ContrastPool (Xu et al., 2024a)	57.36±0.87	57.44±0.69	57.66±6.85	57.08±7.79	68.17±3.28	66.21±3.90	61.51±7.44	72.43±6.53
	ALTER (Yu et al., 2024)	77.99±2.21	70.10±2.26	72.84±7.40	67.68±5.81	71.86±3.13	66.92±3.93	**71.55±8.91**	64.00±6.80
	BioBGT (Peng et al., 2025)	69.96±1.18	69.70±2.90	67.04±3.41	72.02±4.67	63.16±3.74	62.27±3.23	63.97±7.88	60.55±6.71
Quadratic	BQN (Yang et al., 2025)	79.85±1.27	72.53±1.41	73.26±5.99	72.03±6.24	**74.18±3.34**	68.62±3.22	70.91±8.60	68.00±7.81
Ours	TOPOSCANPE	**81.96±1.04**	**75.47±0.81**	**74.99±6.58**	**76.34±5.58**	71.15±2.93	**72.31±1.72**	58.18±13.79	**82.67±12.11**

Implementation details. The experiments are performed on two machines: an NVIDIA RTX PRO 5000 Blackwell GPU (72GB) with an AMD Ryzen 9 5950X processor (16 cores, up to 4.90 GHz) and 64GB RAM, and a 12th Gen Intel Core i7-1270P vPro processor (E-cores up to 3.50 GHz, P-cores up to 4.80 GHz) with 32GB LPDDR5-6400MHz RAM. For all datasets, we randomly divide the data into training, validation and test sets with a ratio of 7:1:2 and report the mean and standard deviation over random splits. We use the Adam optimizer with a cosine learning-rate schedule, an initial learning rate of 10^{-4} and a weight decay of 10^{-4}, a batch size of 16, and 200 epochs, fix $\rho = 40\%$ of nodes, $m = 2$ and $L = 20$. Test metrics are reported at the epoch with the lowest validation loss, following the protocol of Yang et al. (2025). The code implementation is available at `https://anonymous.4open.science/r/BrainConnectivity-20B2/`.

5.2 RESULTS

TOPOSCANPE achieves the best performance across all four metrics on ABIDE datates (Table 1), reaching 81.96% AUC and 75.47% ACC, improving over the strongest baseline BQN by 2.11% and 2.94% points respectively, while also exhibiting the lowest standard across folds (1.04 AUC, 0.81 ACC). Its sensitivity and specificity are well balanced (74.99% vs. 76.34%), whereas several graph transformer baselines trade one for the other. On ADNI, TOPOSCANPE obtains the highest accuracy (72.31%) and specificity (82.67%), and remains competitive in AUC (71.15%), trailing BQN and ALTER.

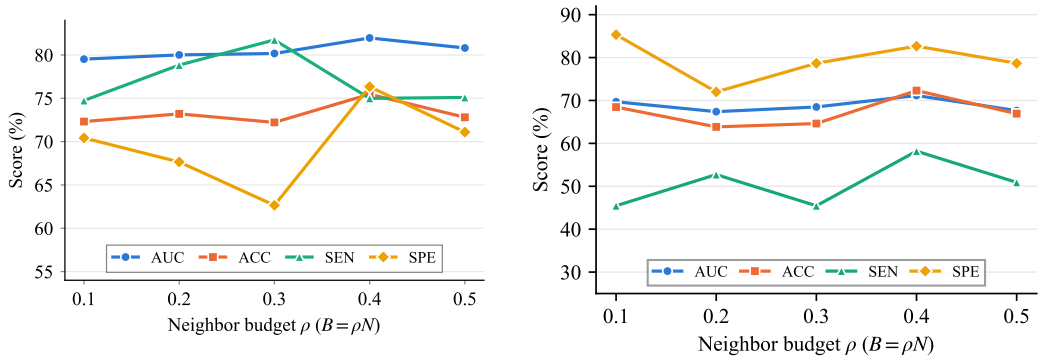
5.3 ABLATION STUDY.

We conduct **two ablation studies**, as follows.

Component Analysis Table 2 reports ablations of the positional encoding, the filtration, the slice features, and the backbone. The full TOPOSCANPE attains the highest AUC and ACC on both datasets, and removing any component degrades performance. The effects are most pronounced on ADNI: replacing the topological encoding with random Gaussian features lowers AUC by 3.76 points (71.15% $\rightarrow$ 67.39%), and restricting the filtration to Forman–Ricci curvature or the heat kernel signature alone lowers AUC by 4.36 and 3.03 points, respectively, indicating that the two filtrations capture complementary structure. Among slice features, combining Betti numbers (β_0, β_1) with slice sizes $(|V|, |E|)$ outperforms either in isolation on both datasets. Removing the residual projection $\mathbf{X}\mathbf{W}_{\text{res}}$ reduces AUC by 2.03 points on ABIDE and 4.48 points on ADNI, and markedly destabilizes sensitivity on ADNI ($36.36 \pm 35.79\%$), showing that the residual path is important for stable optimization. On ABIDE, the individual contributions are consistent but smaller (0.89–2.11 AUC), suggesting that the benefit of TOPOSCANPE there arises from the combination of components rather than from any single one.

Table 2: **Ablations study.** Contribution of each component of TopoScanPE on ABIDE and ADNI. Random PE replaces the Topo-Scan features with Gaussian noise of the same shape

	Variant	ABIDE				ADNI			
		AUC↑	ACC↑	SEN↑	SPE↑	AUC↑	ACC↑	SEN↑	SPE↑
	TopoScanPE (full)	**81.96±1.04**	**75.47±0.81**	74.99±6.58	**76.34±5.58**	**71.15±2.93**	**72.31±1.72**	**58.18±13.79**	**82.67±12.11**
	Random PE (Gaussian noise)	80.07±1.40	72.22±2.05	69.51±7.23	75.17±3.92	67.39±5.27	63.08±5.83	47.27±21.70	74.67±15.92
Filtration	Forman-Ricci only	80.72±1.54	73.30±1.65	74.92±9.37	72.14±7.28	66.79±5.61	65.38±7.69	43.64±33.65	81.33±18.50
	Heat kernel signature only	80.68±1.75	73.89±1.04	**75.03±3.90**	72.80±4.65	68.12±2.63	63.85±4.39	52.73±20.73	72.00±17.26
Slice features	Betti numbers only (β_0, β_1)	79.85±1.66	73.00±1.76	74.22±6.96	72.08±8.20	70.67±4.32	66.15±6.32	45.45±26.50	81.33±12.82
	Sizes only ($\lvert V\rvert, \lvert E\rvert$)	81.07±1.55	73.89±1.74	73.56±6.09	74.43±5.81	69.09±4.87	66.15±5.01	50.91±33.15	77.33±18.62
Backbone	Without residual $\mathbf{XW}_{\text{res}}$	79.93±3.51	73.30±2.84	74.21±7.05	72.41±7.16	66.67±4.98	63.08±5.16	36.36±35.79	82.67±19.78



(a) ABIDE (b) ADNI

Figure 4: Sensitivity to the node budget ρ ($B = \rho N$) at $L = 20$. Each point is the mean over five splits.

Sensitivity to the node budget Figure 4 shows how the node budget ρ affects performance on ABIDE and ADNI. For ABIDE (Figure 4a), AUC stays in a narrow band (79.5–81.9) and peaks at $\rho = 0.4$, as does accuracy (75.5). For small budgets, sensitivity and specificity move in opposite directions: from $\rho = 0.1$ to $\rho = 0.3$, sensitivity rises from 74.7 to 81.7 while specificity falls from 70.4 to 62.6, until the two balance again at $\rho = 0.4$ (75.0 and 76.4). On ADNI (Figure 4b), ρ changes accuracy more than AUC. AUC stays between 67.3 and 71.0, while accuracy drops to 63.8 at $\rho = 0.2$ and then recovers to a peak of 72.2 at $\rho = 0.4$. Specificity is high for every budget (72.0–85.3), but sensitivity is much lower (45.4–58.2), so minority-class recall is the main difficulty on ADNI. The budget $\rho = 0.4$ gives the highest sensitivity (58.2) while specificity stays high (82.6), which makes it the best setting for both AUC and accuracy on both datasets, and we use it as the default.

6 Conclusion

We introduced TopoScanPE, a positional encoding that describes each region of a brain network by the topology of its local neighborhood rather than by the strength of its correlations. For every ROI, Topo-Scan slices a bounded two-hop neighborhood along Forman-Ricci curvature and the heat kernel signature. It records the Betti numbers and size of each slice, and a small encoder shared across regions compresses the resulting sequences into a code that is concatenated to the region's connection profile. Topology therefore enters the model at its input, where it can shape attention between regions, and the backbone, objective and training protocol stay unchanged. On ABIDE, TopoScanPE achieves the best result on all four metrics, improving on BQN by 2.11 AUC and 2.94 accuracy points with the lowest variance across runs. On ADNI, it obtains the highest accuracy and specificity and remains competitive in AUC.

AI USE STATEMENT

In accordance with the ICLR policy on the use of large language models (LLMs), we disclose that LLMs (Claude and ChatGPT) were used in the preparation of this work in three ways. *Implementation*: LLMs assisted in cleaning and optimizing the code. All code was reviewed, tested and run by the authors, and all reported results were produced by the authors' own experiments. *Research support*: LLMs were used to search the literature and public data resources, and to suggest implementation and evaluation choices, such as the design of control experiments; the method and all design decisions were determined and validated by the authors. *Writing*: LLMs assisted in drafting and editing parts of the manuscript and in formatting references. The authors verified all text, claims, equations and citations, and take full responsibility for the content of this paper. LLMs are not authors of this work.

REFERENCES

David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. Graph-based deep learning for medical diagnosis and analysis: past, present and future. *Sensors*, 21:4758, 2021.

Mehmet Emin Aktas, Esra Akbas, and Ahmed El Fatmaoui. Persistence homology of networks: methods and applications. *Appl. Netw. Sci.*, 4(1):61:1–28, 2019. doi: 10.1007/S41109-019-0179-3. URL https://doi.org/10.1007/s41109-019-0179-3.

John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, 2007.

Anushree Bannadabhavi, Soojin Lee, Wenlong Deng, Rex Ying, and Xiaoxiao Li. Community-aware transformer for autism prediction in fMRI connectome. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, volume 14227 of *Lecture Notes in Computer Science*, pp. 287–297. Springer, Cham, 2023. doi: 10.1007/978-3-031-43993-3_28.

D. S. Bassett and E. Bullmore. Small-world brain networks. *The Neuroscientist*, 12(6):512–523, 2006.

Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5833–5848, 2023.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10:186–198, 2009.

R. Cameron Craddock, G. Andrew James, Paul E. Holtzheimer, III, Xiaoping P. Hu, and Helen S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, 2012.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *MICCAI*, 2022.

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, 2023. doi: 10.1109/TMI.2022.3218745.

Wei Dai, Hejie Cui, Xuan Kan, Ying Guo, Sanne van Rooij, and Carl Yang. Transformer-based hierarchical clustering for brain network analysis. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.

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.

Zijian Dong, Yilei Wu, Yu Xiao, Joanna Su Xian Chong, Yueming Jin, and Juan Helen Zhou. Beyond the snapshot: Brain tokenized graph transformer for longitudinal brain functional connectome embedding. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, pp. 348–357, 2023.

Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=wTTjnvGphYj.

Alex Fornito, Andrew Zalesky, and Edward T. Bullmore. *Fundamentals of brain network analysis*. Academic Press, 2016.

Karl J. Friston, Steven Williams, Robert Howard, Richard S. J. Frackowiak, and Robert Turner. Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35(3):346–355, 1996.

J. O. Garcia, A. Ashourvan, S. Muldoon, J. M. Vettel, and D. S. Bassett. Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proceedings of the IEEE*, 106(5):846–867, 2018.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, 2017.

Tong He, Ru Kong, Avram J. Holmes, Minh Nguyen, Mert R. Sabuncu, Simon B. Eickhoff, Danilo Bzdok, Jiashi Feng, and B. T. Thomas Yeo. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, 206:116276, 2020. doi: 10.1016/j.neuroimage.2019.116276.

Jiacheng Hou, Zhenjie Song, and Ercan Engin Kuruoglu. BrainNetMLP: An efficient and effective baseline for functional brain network classification. In *Efficient Medical Artificial Intelligence*, volume 16318 of *Lecture Notes in Computer Science*, pp. 11–20. Springer, 2025. doi: 10.1007/978-3-032-13961-0_2.

Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. FBNETGEN: Task-aware GNN-based fMRI analysis via functional brain network generation. In *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research*, pp. 618–637. PMLR, 2022a. URL https://proceedings.mlr.press/v172/kan22a.html.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. In *Proceedings of the 36th Conference on Neural Information Processing Systems*, volume 35 of *Advances in Neural Information Processing Systems*, pp. 25586–25599, 2022b. doi: 10.52202/068431-1855.

Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *NeurIPS*, 2021.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017a.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*, 2017b.

Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21618–21629, 2021.

Hyekyoung Lee, Hyejin Kang, Moo K. Chung, Bung-Nyun Kim, and Dong Soo Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging*, 31(12):2267–2277, 2012. doi: 10.1109/TMI.2012.2219590.

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.

Yong Liu, Meng Liang, Yuan Zhou, Yong He, Yihui Hao, Ming Song, Chunshui Yu, Haihong Liu, Zhening Liu, and Tianzi Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131: 945–961, 2008.

David Meunier, Renaud Lambiotte, Alex Fornito, Karen D. Ersche, and Edward T. Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3:37, 2009. doi: 10.3389/neuro.11.037.2009.

Ciyuan Peng, Yuelong Huang, Qichao Dong, Shuo Yu, Feng Xia, Chengqi Zhang, and Yaochu Jin. Biologically plausible brain graph transformer. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2502.08958.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 14501–14515, 2022.

Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52:1059–1069, 2010.

J. D. Rudie, J. A. Brown, D. Beck-Pancer, L. M. Hernandez, E. L. Dennis, P. M. Thompson, S. Y. Bookheimer, and M. Dapretto. Altered functional and structural brain network organization in autism. *NeuroImage: Clinical*, 2:79–94, 2013. doi: 10.1016/j.nicl.2012.11.006.

A. Shehzad, D. Zhang, S. Yu, S. Abid, D. K. Kumar, and F. Xia. Multiscale graph transformer for brain disorder diagnosis. *IEEE Transactions on Consumer Electronics*, 2025.

Xilin Shen, Emily S. Finn, Dustin Scheinost, Monica D. Rosenberg, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3):506–518, 2017. doi: 10.1038/nprot.2016.178.

Sean L Simpson, F DuBois Bowman, and Paul J Laurienti. Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. *Statistics Surveys*, 7: 1–36, 2013. doi: 10.1214/13-SS103.

Ann E. Sizemore, Chad Giusti, Ari Kahn, Jean M. Vettel, Richard F. Betzel, and Danielle S. Bassett. Cliques and cavities in the human connectome. *Journal of Computational Neuroscience*, 44(1): 115–145, 2018. doi: 10.1007/s10827-017-0672-6.

Stephen M. Smith, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, and Mark W. Woolrich. Network modelling methods for FMRI. *NeuroImage*, 54(2):875–891, 2011. doi: 10.1016/j.neuroimage.2010.08.063.

Olaf Sporns and Rolf Kötter. Motifs in brain networks. *PLoS Biology*, 2(11):e369, 2004. doi: 10.1371/journal.pbio.0020369.

Cornelis J Stam, BF Jones, G Nolte, M Breakspear, and Ph Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cerebral cortex*, 17:92–99, 2007.

Kaustubh Supekar, Vinod Menon, Daniel Rubin, Mark Musen, and Michael D. Greicius. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Computational Biology*, 4(6):e1000100, 2008. doi: 10.1371/journal.pcbi.1000100.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002.

Md Joshem Uddin, Astrit Tola, Cuneyt Gurcan Akcora, and Baris Coskunuzer. TopoFormer: Topology meets attention for graph learning. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=xanpQt9YZh.

Martijn P. van den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *Journal of Neuroscience*, 31(44):15775–15786, 2011. doi: 10.1523/JNEUROSCI.3539-11.2011.

František Váša and Bratislav Mišić. Null models in network neuroscience. *Nature Reviews Neuroscience*, 23(8):493–504, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *NeurIPS*, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Guangqi Wen, Peng Cao, Huiwen Bao, Wenju Yang, Tong Zheng, and Osmar Zaiane. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Computers in Biology and Medicine*, 142:105239, 2022.

Zhanghao Wu, Paras Jain, Matthew Wright, Azalia Mirhoseini, Joseph E Gonzalez, and Ion Stoica. Representing long-range context for graph neural networks with global attention. In *Proceedings of the 35th Neural Information Processing Systems*, pp. 13266–13279, 2021a.

Zhanghao Wu, Paras Jain, Matthew Wright, Azalia Mirhoseini, Joseph E. Gonzalez, and Ion Stoica. Representing long-range context for graph neural networks with global attention. In *Advances in Neural Information Processing Systems*, volume 34, 2021b.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International Conference on Machine Learning*, pp. 478–487, 2016.

Jiaxing Xu, Yunhan Yang, David Huang, Sophi Shilpa Gururajapathy, Yiping Ke, Miao Qiao, Alan Wang, Haribalan Kumar, Josh McGeown, and Eryn Kwon. Data-driven network neuroscience: On data collection and benchmark. In *Proceedings of the 37th Neural Information Processing Systems*, pp. 21841–21856, 2023.

Jiaxing Xu, Qingtian Bian, Xinhang Li, Aihu Zhang, Yiping Ke, Miao Qiao, Wei Zhang, Wei Khang Jeremy Sim, and Balázs Gulyás. Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, 43(9):3292–3305, 2024a. doi: 10.1109/TMI.2024.3392988.

Jiaxing Xu, Qingtian Bian, Xinhang Li, Aihu Zhang, Yiping Ke, Miao Qiao, Wei Zhang, Wei Khang Jeremy Sim, and Balázs Gulyás. Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, 2024b.

Chao-Gan Yan and Yu-Feng Zang. DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience*, 4:13, 2010.

Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. GroupINN: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *KDD*, 2019.

Liang Yang, Yuwei Liu, Jiaming Zhuo, Di Jin, Chuan Wang, Zhen Wang, and Xiaochun Cao. Do we really need message passing in brain network modeling? In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 70873–70887. PMLR, 2025. URL https://proceedings.mlr.press/v267/yang25r.html.

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? In *NeurIPS*, 2021a.

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? In *Advances in Neural Information Processing Systems*, volume 34, 2021b.

Shuo Yu, Shan Jin, Ming Li, Tabinda Sarwar, and Feng Xia. Long-range brain graph transformer. In *Advances in Neural Information Processing Systems*, volume 37, pp. 24472–24495, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/2bd3ffba268a2699c212a233ed2907f1-Abstract-Conference.html.

Shengjie Zhang, Xiang Chen, Xin Shen, Bohan Ren, Ziqi Yu, Haibo Yang, Xi Jiang, Dinggang Shen, Yuan Zhou, and Xiao-Yong Zhang. A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders. *Medical Image Analysis*, 90:102932, 2023a.

Shengjie Zhang, Xiang Chen, Xin Shen, Bohan Ren, Ziqi Yu, Haibo Yang, Xi Jiang, Dinggang Shen, Yuan Zhou, and Xiao-Yong Zhang. A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders. *Medical Image Analysis*, 90:102932, 2023b.

Appendix

A Additional Experimental Details

A.1 Data preprocessing

For ABIDE we use the preprocessed release of the Preprocessed Connectomes Project and define ROIs with the Craddock 200 atlas (Craddock et al., 2012), which yields $N = 200$ regions.

For ADNI, raw images are preprocessed with the Data Processing Assistant for Resting-State fMRI (DPARSF) (Yan & Zang, 2010), following the pipeline of Yu et al. (2024). The first 10 volumes of each scan are discarded to allow for signal equilibration, and slice timing is corrected using the middle slice as reference. Head motion is modelled with the Friston 24-parameter model (Friston et al., 1996), a brain mask is estimated automatically, and nuisance covariates are regressed out; all remaining options follow the DPARSF defaults. To account for anatomical variability across participants, functional images are normalised to MNI space with DARTEL (Ashburner, 2007), using each participant's T1-weighted image. ROIs are defined with the AAL atlas (Tzourio-Mazoyer et al., 2002) provided in DPARSF, from which we retain the 90 cerebral regions, yielding $N = 90$.

For both datasets, the ROI time series is the average BOLD signal over the voxels of each region, and the connectivity matrix $\mathbf{X}$ is the Pearson correlation between every pair of ROI time series.

A.2 Hyperparameters

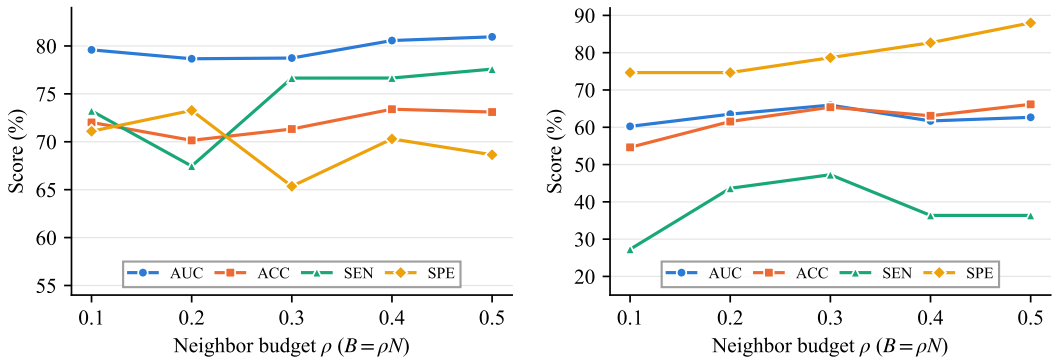
Table 3 lists all hyperparameters. The two datasets share the same configuration except for the number of ROIs, the pooled cluster count, and the dropout of the signature encoder, which is selected on validation loss and set to 0.1 on ABIDE and 0 on the smaller ADNI cohort.

Table 3: Hyperparameters of TopoScanPE on each dataset.

	Hyperparameter	ABIDE	ADNI
Data	Atlas / number of ROIs N	CC200 / 200	AAL / 90
	Train / validation / test split	7:1:2, stratified by site	
	Runs	5	
neighborhood	Nearest neighbors k	8	
	Node budget ρ (max. node number)	0.4 (80)	0.4 (36)
	Heat kernel diffusion time t	1	
Topo-Scan	Sequence length L	20	
	Window width m	2	
Signature encoder	Layers / heads	2 / 4	
	Hidden size d / feed-forward size	64 / 256	
	Dropout	0.1	0
	Summary projection d_{tok}	64	
	Positional encoding size d_{pe}	16	
Backbone	Layers / heads	2 / 4	
	Pooled clusters K	100	45
Training	Optimiser / weight decay	Adam / 10^{-4}	
	Learning rate	$10^{-4} \to 10^{-5}$, cosine	
	Batch size / epochs	16 / 200	
	Mixup	$\lambda \sim \mathrm{Beta}(1, 1)$	
	Loss	Cross-entropy	
	Model selection	Lowest validation loss	

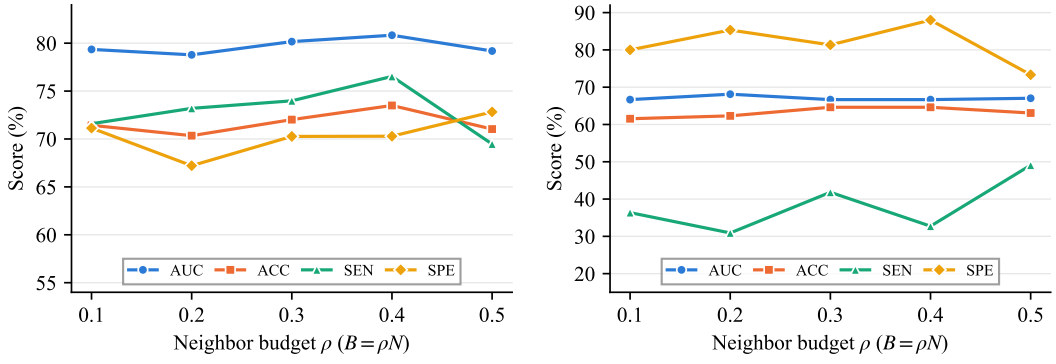
B Additional result

In this section we studied the sensitivity of TopoScanPE with different number of slices L.



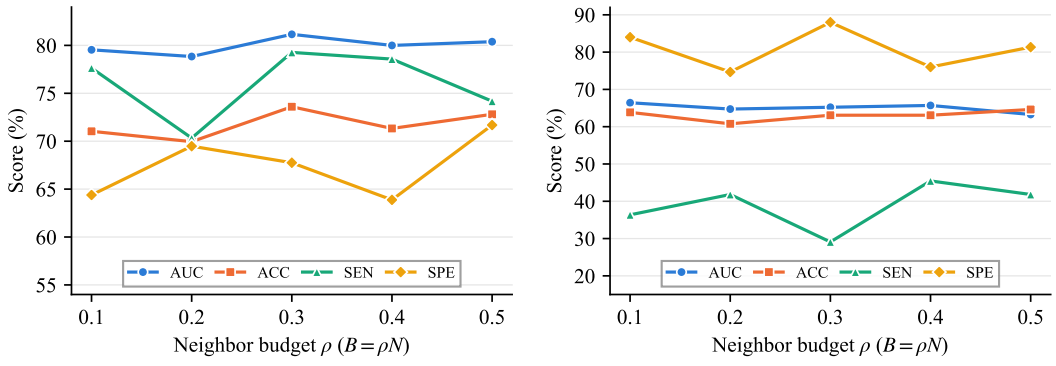
(a) ABIDE (b) ADNI

Figure 5: Sensitivity to the node budget ρ ($B = \rho N$) at $L = 5$. Each point is the mean over five splits.



(a) ABIDE (b) ADNI

Figure 6: Sensitivity to the node budget ρ ($B = \rho N$) at $L = 10$. Each point is the mean over five splits.



(a) ABIDE (b) ADNI

Figure 7: Sensitivity to the node budget ρ ($B = \rho N$) at $L = 15$. Each point is the mean over five splits.