Journal of Data-centric Machine Learning Research (2026) Submitted 2/26; Revised 10/26; Published

# Defining and Benchmarking a Data-Centric Design Space for Brain Graph Construction

**Qinwen Ge** QINWEN.GE@VANDERBILT.EDU
**Roza G. Bayrak** ROZA.G.BAYRAK@VANDERBILT.EDU
**Anwar Said** ANWAR.SAID@VANDERBILT.EDU
**Catie Chang** CATIE.CHANG@VANDERBILT.EDU
**Xenofon Koutsoukos** XENOFON.KOUTSOUKOS@VANDERBILT.EDU
**Tyler Derr** TYLER.DERR@VANDERBILT.EDU
*Department of Computer Science, Vanderbilt University, USA*

**Reviewed on OpenReview:** *https://openreview.net/forum?id=D6LS3f8VfM*

**Editor:** Yi Liu

## Abstract

The construction of brain graphs from functional Magnetic Resonance Imaging (fMRI) data plays a crucial role in enabling graph machine learning for neuroimaging. However, current practices often rely on rigid pipelines that overlook critical data-centric choices in how brain graphs are constructed. In this work, we adopt a Data-Centric AI perspective and systematically define and benchmark a data-centric design space for brain graph construction, constrasting with primarily model-centric prior work. We organize this design space into three stages: temporal signal processing, topology extraction, and graph featurization. Our contributions lie less in novel components and more in evaluating how combinations of existing and modified techniques influence downstream performance. Specifically, we study high-amplitude BOLD signal filtering, sparsification and unification strategies for connectivity, alternative correlation metrics, and multi-view node and edge features, such as incorporating lagged dynamics. Experiments on the HCP1200 and ABIDE datasets show that systematic, validation-guided exploration of data-centric configurations yields substantial accuracy gains over a standard fixed pipeline, while the best-performing construction varies across datasets and settings. These findings highlight the critical role of upstream data decisions and underscore the importance of systematically exploring the data-centric design space for graph-based neuroimaging. Our code is available at https://github.com/GeQinwen/DataCentricBrainGraphs.

**Keywords:** fMRI, GNNs, brain graphs, neuroimaging analysis, data-centric AI

## 1 Introduction

The human brain, as a complex network, can be effectively modeled as a graph to study neural activation patterns using neuroimaging. Functional Magnetic Resonance Imaging (fMRI) captures blood-oxygen-level-dependent (BOLD) signals over time, allowing analysis of brain activity via functional connectivity (FC). In these graphs, nodes correspond to regions of interest (ROIs) from a brain parcellation, and edges represent co-activation

©2026 Ge, Bayrak, Said, Chang, Koutsoukos, Derr.

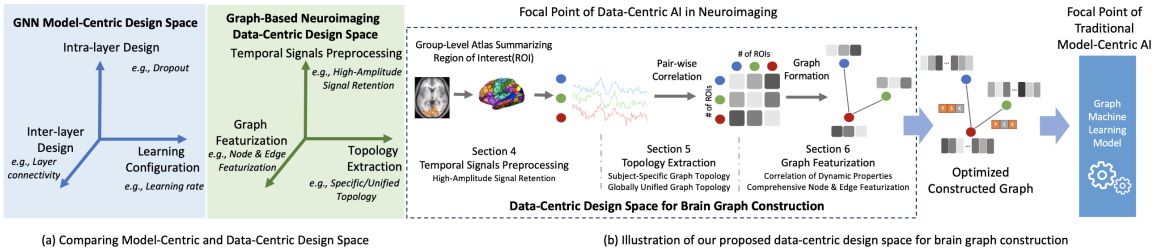



Figure 1: (a) Comparison between model-centric and data-centric design spaces. Traditional approaches focus on model configurations, while our work emphasizes upstream data decisions. (b) Overview of our proposed data-centric design space for brain graph construction. Instead of relying on a fixed pipeline, we systematically explore signal preprocessing, topology extraction, and graph featurization choices, transforming raw fMRI signals into optimized graph representations for GNNs.

relationships (Smith et al., 2013). This formulation has led to widespread adoption of graph-based approaches in neuroimaging (Luo et al., 2024; Du et al., 2024).

Graph machine learning has shown strong promise in analyzing brain graphs for tasks such as disease prediction, cognitive state decoding, and biomarker discovery (Wu et al., 2020; Kim et al., 2021; Said et al., 2023; Gadgil et al., 2020; Dahan et al., 2021; Li et al., 2021; Sivgin et al., 2024). However, most prior work focuses on model architectures and training objectives, while the upstream graph construction process—from raw fMRI signals to brain networks—remains underexplored within the graph machine learning community. These graphs are typically built using a fixed pipeline based on instantaneous correlations (e.g., Pearson), with limited attention to how preprocessing, connectivity definitions, or feature choices affect downstream performance.

Recent studies have shown that even modest variations in graph construction—such as incorporating dynamic correlations—can significantly impact predictive accuracy (Sivgin et al., 2024). This motivates a shift from model-centric innovation toward systematic study of data-centric decisions. In line with principles of Data-Centric AI (Zha et al., 2025; Jakubik et al., 2024; Jarrahi et al., 2022), we argue that the structure and quality of input graphs deserve closer scrutiny. This view is supported by broader AI research showing that improving data representations can yield better performance and faster development than tuning models alone (Ng, 2023). In graph-based neuroimaging, data-centric learning emphasizes how choices in signal processing, connectivity, and feature design—made during graph construction from raw fMRI—directly influence downstream model performance.

In this work, we reframe brain graph construction as the exploration of a structured *data-centric design space*, which organizes key decisions across three stages: (1) temporal signal processing, (2) topology extraction, and (3) graph featurization. As illustrated in Figure 1, this design space provides a flexible and structured alternative to the standard fixed pipeline, enabling systematic evaluation of data-centric choices between raw fMRI signals and graph learning models. Thus, rather than proposing new model architectures, we systematically benchmark individual and combined data-centric configurations of existing/modified techniques to assess their impact on neuroimaging graph classification.

We conduct extensive experiments on the HCP1200 and ABIDE datasets and show that systematic exploration of design choices—such as filtering high-amplitude BOLD fluctuations, applying alternative correlation metrics, unifying connectivity topologies, and encoding lagged dynamics—yields substantial classification gains over standard pipelines, with the best choices varying across datasets and settings. These findings highlight the role of upstream data decisions in shaping brain graph representations. Our benchmark and decomposition of brain graph construction offer a practical toolkit and conceptual foundation for future research in Auto-Data-Centric AI—where data construction pipelines, like model architectures, can be automatically explored and optimized.

Our contributions, framed around the proposed data-centric design space, advance understanding of how upstream data decisions influence graph-based neuroimaging:

- We define and benchmark a representative data-centric design space for brain graph construction, organized into three stages: temporal signal processing, topology extraction, and graph featurization. This framework highlights how upstream data-centric choices shape brain graphs for downstream learning.
- For temporal signal processing, we evaluate multiple high-amplitude signal retention strategies to assess their impact on co-activation patterns and connectivity formation.
- For topology extraction, we explore both subject-specific and globally unified topology frameworks, and systematically compare correlation metrics (e.g., Pearson, Kendall) for their robustness to outliers and nonlinear dependencies.
- For graph featurization, we incorporate not only static correlations but also dynamic lagged correlations into node features, and explore edge feature encodings to consolidate multiple functional interactions within a single graph representation.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents preliminaries, including notations and datasets. Sections 4–6 explore the data-centric design space across temporal signal processing, topology extraction, and graph featurization. Section 7 presents results, and Section 8 concludes. Also, we assess the reliability of each data-centric choice by measuring how often it outperforms the standard pipeline across datasets and ROI settings. We observe that no single design choice dominates across all settings, motivating a modular design space perspective where practitioners can systematically compare and compose data-centric options.

## 2 Related Work

### 2.1 Deep Learning for Neuroimaging

Deep learning's ability to learn highly complex and non-linear patterns from neuroimaging data has enabled learning individualized representations with more predictive power (Arbabshirani et al., 2017; Plis et al., 2014). Given its ability to detect abstract and complex patterns, DL has been applied in neuroimaging studies of psychiatric and neurological disorders, which are characterized by subtle and diffuse alterations (Vieira et al., 2017). Compared to standard machine learning methods, deep learning can learn more representative

information from the data (LeCun et al., 2015) to model the brain activity. Despite the advancements in deep learning for neuroimaging, these models heavily depend on the quality and structure of input data (Yan et al., 2022). This motivates our study to investigate data-centric AI approaches for optimizing neuroimaging data for improving downstream prediction performance.

### 2.2 Data-Centric AI on Graphs

Data-Centric AI focuses on systematically diagnosing and addressing real-world data issues through AI-driven methods (Zha et al., 2023). While AI research has primarily advanced model development, the growing demand for Data-Centric AI (Singh, 2023) arises from the fact that even state-of-the-art models underperform on noisy, uncurated data—illustrated by the "garbage in, garbage out" phenomenon.

Graph data, which models entities and relationships, requires a distinct approach due to its manually abstracted nature. Research in Data-Centric Graph Learning has explored how graph data and models interact (Yang et al., 2023; Wang et al., 2022). Given graph data's unique characteristics, we can analyze different stages of graph learning through a Data-Centric AI lens. This involves addressing key questions: how to construct, when and where to modify, and how to refine a graph. In this work, we adapt the principles of data-centric graph learning to the neuroimaging domain by analyzing how temporal 3D signals preprocessing, topology extraction, and graph featurization affect the construction of functional brain graphs from fMRI data.

### 2.3 Model-Centric Design Space for Graph Machine Learning

The high-impact design space paradigm introduced by GraphGym (You et al., 2020) enhances GNN development by systematically exploring architectural hyperparameters to identify optimal configurations across diverse tasks. Their work establishes a principled approach to model-centric optimization, ensuring generalizable performance. Inspired by this framework, we extend this paradigm into the realm of data-centric AI. Instead of tuning model hyperparameters, we conceptualize data-centric "hyperparameters"—a structured set of preprocessing strategies and feature engineering choices that critically shape model performance. By integrating the most effective proposed methods, we construct a generalizable and empirically robust data-centric workflow, optimizing the entire pipeline of deep learning in neuroimaging rather than just the model architecture. In addition to many other experiments, we ultimately empirically showcase the strength of data-centric design space exploration against fixed baselines in Figure 6.

## 3 Preliminaries

We first outline fMRI-to-graph preprocessing, then formally define the resulting brain graph.

**Standard fMRI to Graph Construction.** Constructing brain graphs from fMRI data typically follows a fixed pipeline that transforms voxel-level signals into region-level functional connectivity graphs. This pipeline involves preprocessing the fMRI signals, summarizing them within regions of interest (ROIs), computing pairwise correlations to obtain a functional connectivity (FC) matrix, and applying thresholding to define graph edges. A more detailed description of the standard pipeline is provided in Appendix B.

Table 1: Notations and corresponding descriptions.

| | Description | | Description |
|---|---|---|---|
| $N$ | Number of samples | $\mathbf{T}_r(\mathbf{Z}_r)$ | Temporal (Normalized) BOLD signal for ROI $r$ |
| $T$ | Number of fMRI time points | $\mu_r(\sigma_r)$ | Mean (Std. dev.) of the BOLD signal for ROI $r$ |
| $n$ | Number of nodes/ROIs | $v_i \in \mathcal{V}$ ($r_i \in \mathcal{R}$) | Node (ROI) $i$ in the node (ROI) set $\mathcal{V}(\mathcal{R})$ |
| $\mathcal{V}$ ($\mathcal{E}$) | Node (Edge) set | $\Theta(t; \mathbf{Z}_r; \theta, \gamma)$ | High-amplitude signal thresholding function |
| $d/d'$ | Node/edge feature dimension | $A$ | Adjacency matrix, $A = [a_{ij}] \in \{0,1\}^{\lvert\mathcal{V}\rvert \times \lvert\mathcal{V}\rvert}$ |
| $\mathbf{E}$ | Edge feature matrix, $\mathbb{R}^{\lvert\mathcal{E}\rvert \times d'}$ | $\mathbf{X}$ | Node feature matrix, $\mathbb{R}^{\lvert\mathcal{V}\rvert \times d}$ |
| $\theta(\gamma)$ | Threshold (Binarization) parameter for high-amplitude retention function $\Theta$ | | |

**Notations and Definitions.** Here we formally define the functional brain graph created from the standard fMRI to graph pipeline; Table 1 constains a summary of all notation.

**Definition 1** *The brain functional connectomics graph is defined as* $G = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{E})$*, where the node set* $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ *corresponds to regions of interest (ROIs) defined by a parcellation atlas, and the edge set* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ *represents functional connectivity (FC) above a defined threshold between pairs of ROIs. The matrix* $\mathbf{X} \in \mathbb{R}^{n \times d}$ *denotes the node feature matrix, where* $n$ *is the number of ROIs and* $d$ *is the feature dimension. The adjacency matrix* $\mathbf{A} \in \{0,1\}^{n \times n}$ *encodes the graph structure, with* $\mathbf{A}_{ij} = 1$ *indicating an edge between nodes* $v_i$ *and* $v_j$*, signifying a strong co-activation relationship between the corresponding ROIs. The edge feature matrix* $\mathbf{E} \in \mathbb{R}^{|\mathcal{E}| \times d'}$ *encodes additional multi-view connectivity information across multiple correlation measures.*

## 4 Temporal Signal Preprocessing

fMRI is a powerful non-invasive tool for studying brain function, widely used in Neuroscience, Psychology, and Psychiatry (Smith, 2004). However, fMRI data, derived from the BOLD signal, may be influenced by non-neural sources of variability (Power et al., 2017), necessitating careful preprocessing to optimize the quality of extracted features. Numerous techniques are available for denoising fMRI data (Caballero-Gaudes and Reynolds, 2017), which are crucial for further analysis in neuroimaging. For brain network analysis using fMRI data, such extensive preprocessing is essential before any analysis can be conducted (Smith et al., 2013). Nevertheless, even after common preprocessing steps, the BOLD signal often remains noisy, which can affect the accuracy of correlation calculations, a key step in graph construction.

*Motivating High-Amplitude Signal Retention.* We observed that the approach for calculating functional connectivity can be influenced by both high- and low-amplitude fluctuations. Motivated by this, previous neuroimaging studies have also explored related questions from a domain knowledge perspective. Previous studies on resting-state fMRI data have demonstrated that high-amplitude BOLD fluctuations may play a crucial role in defining brain functional connectivity, as they often correspond to meaningful co-activation patterns. Retaining primarily high-amplitude fluctuations has been shown to reduce data volume by up to 94% (Tagliazucchi et al., 2012), while preserving key functional networks (Liu and Duyn, 2013). Additionally, other work has shown that functional networks are driven by activity at only a few critical time points, using just 15% of the data to accurately replicate a resting-state functional network (Liu and Duyn, 2013). To investigate this, we explore several variants of high-amplitude signal retention as a preprocessing design choice.

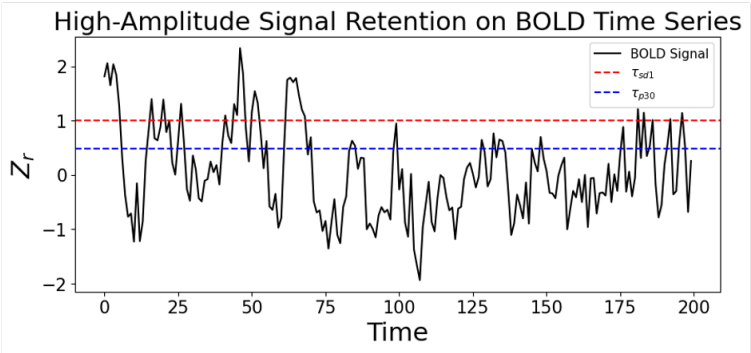



Figure 2: High-Amplitude Signal Retention Process.

### 4.1 High-Amplitude Signal Retention Process

Let $\mathcal{R}$ be the set of ROIs and $\mathcal{T} = \{1, 2, ..., T\}$ be the set of time points. For each ROI $r \in \mathcal{R}$, we define its temporal BOLD signal as $\mathbf{T}_r = [T_r(1), T_r(2), ..., T_r(T)] \in \mathbb{R}^T$, where $T_r(t)$ represents the BOLD signal value at time $t$ for ROI $r$.

We define the high-amplitude signal retention process as: $f_{\text{high-amplitude}} : \mathbb{R}^T \to \mathbb{R}^T$, which consists of two steps: normalization and thresholding.

The normalization begins with classical Z-score normalization, defined as follows:

$$\mathbf{Z}_r = [Z_r(1), ..., Z_r(T)], \text{ where } Z_r(t) = \frac{T_r(t) - \mu_r}{\sigma_r} \tag{1}$$

$$\mu_r = \frac{1}{T}\sum_{t=1}^{T} T_r(t), \quad \sigma_r = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(T_r(t) - \mu_r)^2} \tag{2}$$

where $\mu_r$ and $\sigma_r$ are the mean and standard deviation of ROI $r$'s signal $\mathbf{T}_r$, respectively.

After normalizing we apply a thresholding function, $\Theta(t, \mathbf{Z}_r; \theta, \gamma)$, that allows for multiple variants based on the threshold $\theta$ and a parameter $\gamma$, as follows:

$$\Theta(t, \mathbf{Z}_r; \theta, \gamma) = \begin{cases} Z_r(t), & \text{if } Z_r(t) \geq \theta \text{ and } \gamma = 0 \\ 1, & \text{if } Z_r(t) \geq \theta \text{ and } \gamma = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

This function allows for two distinct behaviors. When $\gamma = 0$, the function retains values greater than or equal to $\theta$, while setting all values below the threshold to zero. This is referred to as thresholding with retention and allows a continuous representation while emphasizing stronger signals. This setting ensures that high-magnitude fluctuations are preserved without introducing artificial saturation effects. On the other hand, when $\gamma = 1$, the function converts values greater than or equal to $\theta$ to 1, and all values below the threshold to 0 (i.e., a binarization based on $\theta$). This variant is called binary thresholding and enforces a strict thresholding mechanism that highlights only the most significant fluctuations. Notably, $\gamma = 1$ emphasizes positive high-amplitude events while preventing extreme values from disproportionately influencing the connectivity, but it discards differences in their original amplitudes.

We explore two common strategies to determine the threshold $\theta$. The first approach is the percentile-based method, where $\theta$ is defined as the $\alpha$-th percentile of $\mathbf{Z}_r$, denoted by $Q_\alpha(\mathbf{Z}_r)$, with $\alpha \in (0, 1)$. The second way is the standard deviation-based method, where $\theta$

is computed as $\beta\sigma$, with $\sigma$ representing the global standard deviation of $\mathbf{Z}_r$ across all ROI $r$ and $\beta > 0$ as a scaling factor.

The High-Amplitude Signal Retention process is visualized in Figure 2. In summary, this flexible thresholding framework allows for both continuous and binary transformations of the data, depending on the desired outcome. In this work, we fixed $\alpha = 30$ and $\beta = 1$ (i.e., top 30 percentile and 1 standard deviation, respectively). Thus, we can summarize our four high-amplitude signal retention variants as $\Theta_{\theta_{p30}\gamma_1}$, $\Theta_{\theta_{p30}\gamma_0}$, $\Theta_{\theta_{sd1}\gamma_0}$, $\Theta_{\theta_{sd1}\gamma_1}$.

## 5 Topology Extraction

Given our Data-Centric AI perspective, we treat the topology of the brain graph—not as a fixed byproduct of fMRI processing—but as a tunable design choice. While traditional pipelines extract topology from a correlation matrix using default heuristics, we investigate how alternative design decisions can shape the connectivity structure to better support graph-based learning. In this section, we explore two complementary directions: (1) modifying the correlation-based construction of subject-specific topologies, and (2) exploring a globally unified graph topology shared across subjects. These reflect different points in the design space and highlight how edge-level decisions impact downstream performance.

After extracting the signal data and averaging the time courses of ROIs based on a predefined atlas, correlation analysis is the most widely used approach for functional connectivity construction (i.e., topology extraction) (Du et al., 2024). This allows for the simple and direct computation of a subject-specific $n \times n$ correlation matrix (Smith et al., 2013).

However, a fixed correlation method may not always be optimal, and vary in ability to handle low-amplitude signal fluctuations (Smith et al., 2013) or oversimplify the complex patterns of brain activity (Ahmadi et al., 2023). Furthermore, subject-specific brain graph structures exhibit variability, as edges are determined individually for each subject. This motivates exploration of alternative topological designs within a data-centric framework.

### 5.1 Subject-Specific Brain Graph Topology

Here for each individual subject-specific brain we explore alternative correlation measures as design choices for defining functional connectivity, evaluating their potential to offer complementary perspectives in extracting topology for functional brain connectivity graphs.

#### 5.1.1 Motivation

Pearson Correlation Coefficient (PCC), while widely used in functional connectivity analysis for its simplicity, has notable limitations. It assumes linear relationships between time series, which may not capture the complex dependencies in brain activity (Ahmadi et al., 2023). Moreover, PCC's lack of robustness means outliers can introduce false correlations or mask existing ones (Wilcox and Rousselet, 2023). When dealing with non-normal distributions and outliers, researchers have suggested alternative measures like Spearman's rank correlation and Kendall's tau (Rousselet and Pernet, 2012). These metrics may offer advantages in handling non-linear dependencies and are shown to improve predictive performance (Ahmadi et al., 2023).

#### 5.1.2 Correlation-based Edge Construction

Based on the above, we first formally define the traditionally used Pearson Correlation Coefficient, and then also Spearman Correlation and Kendall's tau that we propose exploring to investigate their potential as being more robust (e.g., against outliers). The detailed definitions of three correlation coefficient can be found in Appendix A

We examined subject-specific topology through different correlation measures as design choices. While Pearson is most widely used, Spearman and Kendall offer robustness to nonlinearity and outliers, highlighting how simple substitutions in edge construction can affect brain graph quality.

### 5.2 Globally Unified Brain Graph Topology

The typical approach in brain graph construction is to define topology separately for each subject by thresholding their correlation matrix to retain only the strongest functional connections. This results in individualized graphs that vary across subjects. While widely adopted, the implications of such subject-specific variability for GNN-based learning—especially from a data-centric perspective—remain insufficiently understood.

#### 5.2.1 Motivation.

From a data-centric perspective, subject-specific brain graphs may underutilize GNN aggregation, as each node already encodes global connectivity through its features. Unlike typical graph tasks where neighborhoods carry new information, brain graphs derived from fMRI embed full co-activation profiles at the node level. While prior work has explored learning-based strategies to adapt graph structure during training (Cui et al., 2022b; Li et al., 2023), we instead consider a simple, data-driven alternative: constructing a shared, stable topology by identifying edges that consistently appear across subjects.

#### 5.2.2 Global Functional Co-activation Filtering.

We explore a data-driven design choice for constructing a globally consistent graph topology. This configuration assigns a shared graph topology to all subjects in the dataset by assigning a consistent adjacency matrix. The high-level idea of Global Functional Co-activation Filtering is visualized in Figure 3. Specifically, for each split, we define the set of training-subject brain graphs as $\mathcal{G} = \{G_1, G_2, \ldots, G_N\}$, where $N$ is the number of training subjects. Then, to construct a globally unified topology, we define a function $F$ that aggregates the adjacency matrices of these training-subject brain graphs and outputs a unified adjacency matrix: $F(\mathcal{G}) \rightarrow \mathbf{A}_{\text{unified}}$. This adjacency matrix is fixed within each split and shared by training, validation, and test subjects.

The entries of $\mathbf{A}_{\text{unified}}$ are determined by retaining the most frequently occurring edges across subjects. Specifically, we set a threshold $k_{\text{threshold}}$ to include only the top proportion (e.g., 5%) of the most consistently present connections:

$$A_{\text{unified},ij} = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} A_{i,ij} \geq k_{\text{threshold}} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

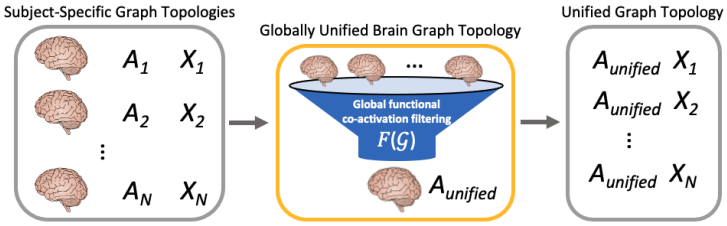



Figure 3: Globally unified brain graph topology process.

where $k_{\text{threshold}}$ is chosen such that the total number of edges in $\mathbf{A}_{\text{unified}}$ does not exceed the desired proportion of possible connections. Thus, the final unified graph is: $G_{\text{unified}} = (\mathcal{V}, \mathbf{A}_{\text{unified}}, \mathbf{X})$, where $\mathbf{X}$ represents the node features of individual subjects, and all subjects share the same unified adjacency matrix $\mathbf{A}_{\text{unified}}$.

This design choice emphasizes broadly co-activated regions to reduce variability and promote a more stable functional connectivity for GNNs. Consequently, it supports a meaningful aggregation mechanism in GNNs by focusing on generalized connectivity patterns; our findings show that even simple group-level topologies can deliver strong performance.

## 6 Advanced Graph Featurization

Featurization determines how functional connectivity is encoded in brain graphs and directly impacts what information GNNs can leverage. While prior work often treats static and dynamic functional connectivity separately, we view them as complementary design choices in a unified featurization space. We explore incorporating lagged temporal correlations into node features and propose edge feature encodings that capture multi-view functional relationships beyond the static Pearson graph.

### 6.1 Correlation of Dynamic Properties

The common approach to calculating functional connectivity (FC) relies on computing correlations between the time series of each pair of nodes (Said et al., 2023; Cui et al., 2022a; Li et al., 2021; Luo et al., 2024). As discussed in Section 5.1, previous studies typically use this $n \times n$ correlation matrix as the node features. While this approach is straightforward, using only the instantaneous (i.e., zero-lag) correlation may fail to capture temporally offset dependencies between regional BOLD signals (Marrelec et al., 2006), prompting work on dynamic properties of fMRI data and recent work has developed GNNs to model dynamic brain graphs (Kim et al., 2021; Gadgil et al., 2020; Zhu et al., 2024). Although advanced models enhance the ability of GNNs to capture temporal properties, they also introduce increased complexity, and can sometimes degrade performance (Said et al., 2023). Here we explore a data-centric design choice that incorporates lagged correlations into brain graph construction, aiming to improve sensitivity to temporal dependencies in fMRI signals.

From a data-centric perspective, the common approach of calculating only the instantaneous correlation may fail to capture crucial temporal dynamics, overlooking time-lagged interactions between brain regions. For example, two ROIs may show strong overall correlation, but if the activation of one lags behind the other, PCC would inaccurately suggest

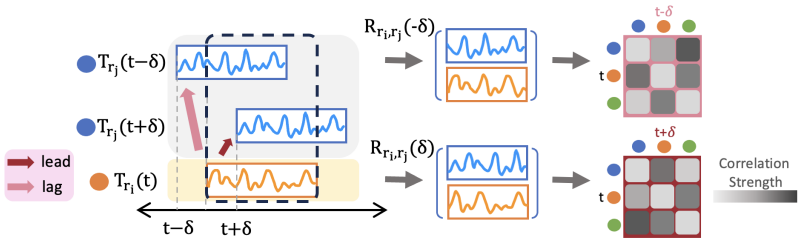



Figure 4: An illustration of lagged correlation computation.

a weak connection, missing valuable information about neural processes. By incorporating dynamic properties, we can enable brain graphs to better capture temporally offset dependencies between regional BOLD signals.

From a neuroimaging perspective,researchers have investigated temporal dependencies between brain regions (Smith et al., 2013; Du et al., 2024; Smith et al., 2011). Previous studies have demonstrated that dynamic inter-regional interactions can be captured by lagged correlation (Sivgin et al., 2024). Also, researchers have investigated the idea of Cross-Correlation (Fiecas et al., 2013; Hyde and Jesmanowicz, 2012), formalized as:

$$R_{r_i,r_j}(\delta) = \frac{\sum_t (\mathbf{T}_{r_i}(t) - \mu_{r_i})(\mathbf{T}_{r_j}(t+\delta) - \mu_{r_j})}{\sqrt{\sum_t (\mathbf{T}_{r_i}(t) - \mu_{r_i})^2}\sqrt{\sum_t (\mathbf{T}_{r_j}(t+\delta) - \mu_{r_j})^2}} \tag{5}$$

where $\mathbf{T}_{r_i}(t)$ and $\mathbf{T}_{r_j}(t+\delta)$ represent the BOLD signal values at time $t$ and $t+\delta$ for ROIs $r_i$ and $r_j$, respectively, and $\mu_{r_i}$ and $\mu_{r_j}$ denote their corresponding mean BOLD signals. Observed temporal offsets may also reflect hemodynamic variability, acquisition timing, and preprocessing effects.

As illustrated in Figure 4, we explore constructing brain graphs with lagged correlations as a design choice to capture leading/lagging relationships between regions. Specifically, we operationalize the cross-correlation in Eq. 5 by explicitly shifting the BOLD time series of one ROI relative to another and computing correlations at multiple temporal offsets. The left panel shows example time series from two ROIs, where one signal is shifted by a lag $\delta$ to align past or future activity across regions. As shown in Eq. 5, if $\delta$ is positive, it means that $\mathbf{T}_{r_i}(t)$ leads $\mathbf{T}_{r_j}(t)$, indicating that changes in ROI $r_i$ precede those in ROI $r_j$. Similarly, if $\delta$ is negative, it means $\mathbf{T}_{r_i}(t)$ lags $\mathbf{T}_{r_j}(t)$, i.e., changes in ROI $r_i$ follow those in $r_j$. This is repeated for all ROI pairs in the graph and enables the model to capture whether fluctuations in one ROI systematically precede or follow those in another, rather than assuming instantaneous synchronization. The resulting visualized feature matrix in the right panel is then the concatenation of the three correlation matrices. This procedure results in a graph representation that augments traditional functional connectivity with enhanced temporal features while remaining agnostic to the downstream GNN model architecture.

### 6.2 Comprehensive Node & Edge Featurization

The previous sections introduced multiple design dimensions for brain graph construction, each targeting different aspects of the data. Here, we investigate how combining these design choices—on signal selection, correlation measures, and temporal sensitivity—can improve node and edge featurization. These strategies represent different axes in the data-centric design space, and combining them allows us to evaluate richer representations through

expanded node and edge features. Since each of these strategies targets a different stage of brain network construction, here we can combine them to investigate a more comprehensive node/edge featurization. More specifically, we define the node and edges features as follows:

*Node features* are enriched by concatenating correlation matrices. Specifically, we construct the new node feature matrix as: $\mathbf{X} = \left[\mathbf{R}^{(1)}, \mathbf{R}^{(2)}, \ldots, \mathbf{R}^{(d')}\right] \in \mathbb{R}^{n \times d'n}$, where $\mathbf{R}^{(i)}$ represents the correlation matrix computed using the $i$-th correlation measure, including Pearson, Spearman, Kendall, and lagged correlations.

*Edge features* are typically not included in standard brain graph representations, or perhaps sometimes weights according to the correlation values. However, here we investigate whether encoding the multiple correlations as edge features can provide benefits beyond their contribution as node features. Specifically, here we retain the adjacency matrix from the Pearson correlation graph, which serves as the primary graph structure and then construct an edge feature matrix as: $\mathbf{E} \in \{0,1\}^{|E| \times d'}$, $\mathbf{E}_{(i,j),k} = \mathbb{I}\big(\mathbf{R}_{ij}^{(k)} > \rho\big)$, with $d'$ correlation measures and $\rho$ representing the correlation threshold to determine if edge (i,j) in correlation matrix $\mathbf{R}^{(k)}$ is encoded as 1/0 in $\mathbf{E}_{(i,j),k}$ according to the indicator function, $\mathbb{I}$. These binary edge attributes are used by the GNN layers in the edge-feature experiments.

## 7 Experiments

In this section, we systematically evaluate how different data-centric design choices affect downstream GNN performance on neuroimaging tasks. We conduct controlled experiments across our three dimensions—temporal signal processing, topology extraction, and graph featurization—to assess their individual and combined impact. Our findings demonstrate that structured exploration of this design space yields significant performance gains over standard fixed-pipeline approaches.

### 7.1 Experiment Settings

We implemented a unified GNN architecture across all datasets to maintain a consistent data-centric approach. For the HCP dataset, we used a three-layer GNN, while for the ABIDE dataset, we used a two-layer GNN, and both used sort pooling (Zhang et al., 2018) followed by a two-layer Multi-Layer Perceptron (MLP) (Said et al., 2023).

For the main result tables, we split the data into 70% training, 20% testing, and 10% validation. HCP-Task uses participant-disjoint partitions, with all scans from each participant assigned to the same partition and identical participant splits across ROI resolutions. Main-table accuracies are reported as means and standard deviations over five matched model seeds on a fixed data split; the separate robustness analyses use ten matched train/validation/test splits. Statistical comparisons use GCN results from the matched splits, with paired two-sided $t$-tests and 95% confidence intervals for mean paired accuracy differences. We apply Benjamini–Hochberg correction within each comparison family and assess significance at $q < 0.05$. Each model was trained for 100 epochs with a learning rate of 0.0005 for classification tasks. Throughout all experiments, we applied a dropout rate of 0.5, weight decay of 0.0005, and used 32 hidden dimensions for the GNN convolution layers and 64 hidden dimensions for the MLP layers. Cross-entropy was used as the loss function for all classification tasks.

Table 2: Average accuracy ranking of representative data-centric design-space components. Within each dataset–ROI–GNN setting, methods are ranked by mean test accuracy, with rank 1 denoting the highest accuracy and ties assigned the same minimum rank. To ensure common support when configurations are unavailable, rankings for each dataset are computed only over ROI–GNN settings completed by all methods shown in the table, yielding 8 settings for HCP-Rest, 6 for HCP-Task, and 6 for ABIDE. The reported value is the average rank across these settings.

| | | **Temporal Signal Processing** | | | | **Topology Extraction** | | **Graph Featurization** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dataset** | **Baseline** | $\Theta_{\theta_{p30}\gamma_0}$ | $\Theta_{\theta_{p30}\gamma_1}$ | $\Theta_{\theta_{sd1}\gamma_0}$ | $\Theta_{\theta_{sd1}\gamma_1}$ | **Spearman** | **Kendall** | **Lag=1** | **Lag=5** | **w/ Edge Feat.** | **E.F. Lag=1** | **E.F. Lag=5** |
| **Rest** | 8.6 | 7.4 | 5.6 | 9.2 | 5.0 | 8.0 | 4.4 | 10.1 | 2.5 | 9.4 | 5.1 | 2.2 |
| **Task** | 7.0 | 10.0 | 8.8 | 12.0 | 11.0 | 5.0 | 5.0 | 4.8 | 1.3 | 5.2 | 4.8 | 2.3 |
| **ABIDE** | 4.8 | 6.8 | 7.8 | 7.0 | 5.0 | 4.8 | 4.3 | 3.5 | 4.7 | 6.5 | 10.3 | 11.8 |

**Datasets.** Human Connectome Project (HCP) S1200 (Van Essen et al., 2013) and Autism Brain Imaging Data Exchange (ABIDE) dataset (Di Martino et al., 2014). For HCP, we use both resting-state (1,078 subjects) and task-based fMRI scans (>7,000 scans across 7 task types), framing the problems as gender classification (2-class) and task-type classification (7-class), respectively. For ABIDE, we use resting-state scans from 1,009 subjects with binary autism diagnosis labels. HCP and ABIDE were parcellated using the Schaefer and Craddock (CC200 and CC400) atlases, respectively, and underwent standard preprocessing pipelines; full details are provided in Appendix C.

### 7.2 Overall Results: Data-Centric Design Space

In this section, we synthesize empirical findings from our exploration of the data-centric design space, assessing how design dimensions—temporal signal preprocessing, topology extraction, and graph featurization—affect performance across datasets, models, and ROI settings. For comparisons affected by unavailable configurations, we average each component and its baseline over the same ROI–GNN settings completed by both within each dataset (common support). Next, we summarize key findings from these results:

- Table 7.2 compares the performances of each individual proposed component. We observe incorporating lagged correlation and advanced featurization generally achieve top ranks.
- Figure 5 provides an overall perspective of the benefits of our data-centric design space. Specifically, we present how often the different dimensions are able to outperform the baseline, i.e., just using the standard fixed fMRI to brain graph pipeline. These results are aggregated over all dataset and ROI settings. This experiment captures how consistently each design leverages data to gain improvements, offering additional insight into the reliability of specific strategies across contexts.
- For Figure 6, the baseline GNN and the complete data-centric configuration (graph construction and GNN architecture) are selected exclusively by mean validation accuracy over five matched model seeds. We report the test accuracy of the selected configurations. For HCP-Task, although the baseline already achieves high performance, the proposed data-centric strategies further enhance the results, confirming their complementary value. The validation-selected configurations improve test accuracy in seven of

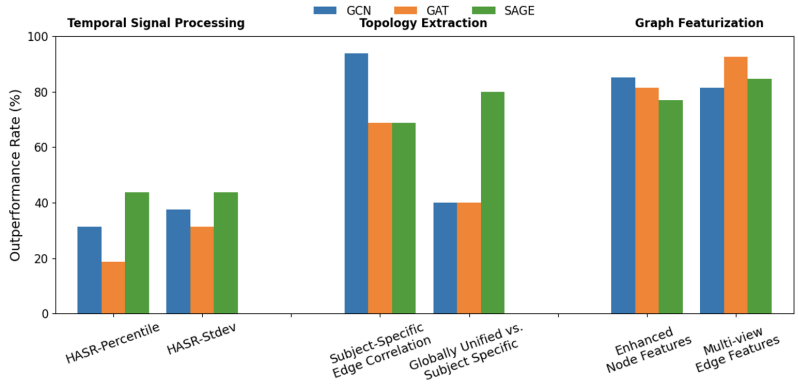



Figure 5: Outperformance rate of data-centric design choices across eight dataset–ROI settings: HCP-Rest and HCP-Task at 100/400/1000 ROIs, and ABIDE at 200/400 ROIs. For each design category and GNN, the rate is the fraction of available component–dataset–ROI comparisons in which mean test accuracy exceeds the baseline under the same dataset, ROI resolution, and GNN. Unavailable configurations are excluded from the denominator.

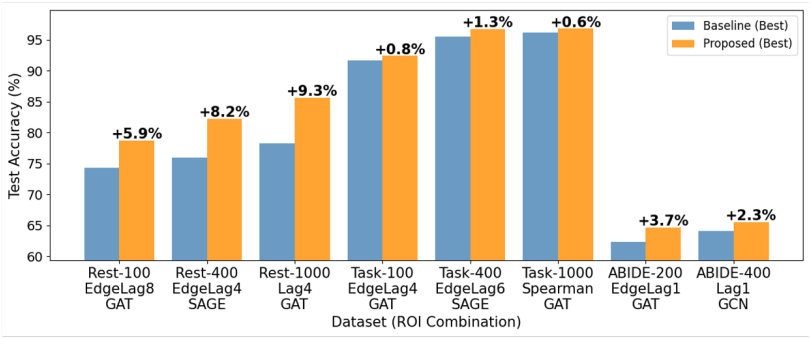



Figure 6: A comparison of the best baseline model using the fixed standard fMRI to graph construction pipeline (blue) as compared to systematic exploration within a data-centric design space (orange). Both the baseline GNN and the data-centric configuration are selected by mean validation accuracy, and their mean test accuracies are reported.

the eight evaluated settings, with a mean gain of 1.49 percentage points across all eight settings and a maximum gain of 4.17 percentage points.

- Lastly, we observe that no single strategy dominates across all datasets. This reinforces the value of a modular and flexible data-centric design space framework that enables systematic comparison and selection.

In the ten-split GCN robustness analyses, lag-5 improves over Pearson on HCP-Rest by 4.03 and 3.70 percentage points at 100 and 400 ROIs, respectively (95% CIs [1.30, 6.76] and [1.55, 5.86]; BH-adjusted $q = 0.039$ and $q = 0.033$). Under participant-disjoint HCP-Task evaluation, the corresponding gains are 0.86 and 1.21 percentage points (95% CIs [0.62, 1.10] and [0.62, 1.80]; $q = 0.0002$ and $q = 0.0036$). At 1000 ROIs, the corresponding HCP-Task gain is 0.64 percentage points (95% CI [0.13, 1.14]; BH-adjusted $q = 0.034$). Five of the nine HCP-Task comparisons are significant after BH correction (Table 13).

Table 3: Classification results on HCP-Rest and HCP-Task for various high-amplitude signal retention strategies.

| | | ROI 100 | | | | ROI 400 | | | | ROI 1000 | | | | Overall Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | Strategy | GCN | GAT | SAGE | GNN Avg. | GCN | GAT | SAGE | GNN Avg. | GCN | GAT | SAGE | GNN Avg. | across ROIs & GNNs |
| Rest | **Baseline** | 74.07±1.13 | 74.35±2.56 | 71.11±3.76 | 73.18 | 74.63±0.68 | 76.02±0.80 | 75.37±2.08 | 75.34 | 81.39±2.81 | 78.33±3.43 | 81.48±1.80 | 80.40 | 76.31 |
| | $\Theta_{\theta_{p30}\gamma_0}$ | 71.39±3.05 | 73.15±2.65 | 74.44±1.35 | 72.99 | 77.41±2.51 | 75.56±2.04 | 76.76±1.42 | 76.58 | 82.78±0.68 | 81.85±2.75 | 83.61±1.12 | 82.75 | 77.44 |
| | $\Theta_{\theta_{p30}\gamma_1}$ | 72.31±2.74 | 73.52±1.74 | 71.02±1.39 | 72.28 | 78.06±1.98 | 76.76±3.15 | 77.96±1.19 | 77.59 | 84.35±1.29 | 84.63±1.72 | 85.09±0.80 | 84.69 | 78.19 |
| | $\Theta_{\theta_{sd1}\gamma_0}$ | 69.91±1.01 | 72.41±1.62 | 72.13±1.81 | 71.48 | 76.30±1.15 | 75.37±1.72 | 76.57±2.35 | 76.08 | 81.39±1.07 | 82.96±1.53 | 81.85±2.18 | 82.07 | 76.54 |
| | $\Theta_{\theta_{sd1}\gamma_1}$ | 74.26±1.59 | 71.67±3.48 | 72.22±3.70 | 72.72 | 78.43±2.76 | 77.22±1.50 | 77.22±2.44 | 77.62 | 84.44±1.77 | 84.07±1.26 | 83.89±1.45 | 84.13 | 78.16 |
| Task | **Baseline** | 93.29±0.65 | 93.32±0.60 | 93.54±0.58 | 93.38 | 94.97±0.15 | 96.17±0.34 | 96.18±0.12 | 95.77 | 96.18±0.49 | 96.46±0.33 | 96.52±0.66 | 96.39 | 95.18 |
| | $\Theta_{\theta_{p30}\gamma_0}$ | 89.60±0.71 | 89.09±0.20 | 89.87±0.32 | 89.52 | 93.90±0.28 | 94.27±0.43 | 94.59±0.89 | 94.26 | 94.97±0.34 | 95.55±0.32 | 96.10±0.72 | 95.54 | 93.11 |
| | $\Theta_{\theta_{p30}\gamma_1}$ | 90.29±0.60 | 90.28±0.68 | 90.76±0.61 | 90.44 | 93.89±0.32 | 94.83±0.47 | 95.89±0.62 | 94.87 | 94.73±0.54 | 95.59±0.46 | 97.23±0.32 | 95.85 | 93.72 |
| | $\Theta_{\theta_{sd1}\gamma_0}$ | 89.45±0.56 | 88.97±0.62 | 89.68±0.57 | 89.37 | 93.70±0.64 | 94.21±0.70 | 95.13±0.24 | 94.35 | 94.55±0.32 | 94.71±1.40 | 95.94±0.54 | 95.07 | 92.93 |
| | $\Theta_{\theta_{sd1}\gamma_1}$ | 90.00±0.88 | 89.83±0.16 | 90.84±0.59 | 90.22 | 94.15±0.27 | 94.61±0.36 | 95.77±0.30 | 94.84 | 95.17±0.18 | 95.58±0.52 | 96.62±0.64 | 95.79 | 93.62 |

Table 4: ABIDE Classification results for various high-amplitude signal retention strategies.

| | ROI 200 | | | ROI 400 | | | Avg. across |
|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | GCN | GAT | SAGE | ROIs & GNNs |
| **Baseline** | 60.00 | 62.13 | 62.32 | 62.32 | 64.06 | 63.86 | 62.45 |
| $\Theta_{\theta_{p30}\gamma_0}$ | 61.74 | 62.13 | 62.22 | 61.64 | 63.09 | 57.87 | 61.45 |
| $\Theta_{\theta_{p30}\gamma_1}$ | 58.94 | 61.45 | 61.35 | 62.13 | 63.96 | 60.00 | 61.31 |
| $\Theta_{\theta_{sd1}\gamma_0}$ | 62.02 | 62.32 | 61.72 | 60.77 | 62.73 | 61.53 | 61.85 |
| $\Theta_{\theta_{sd1}\gamma_1}$ | 65.86 | 63.03 | 63.84 | 61.64 | 61.75 | 62.08 | 63.03 |

We also evaluated sensitivity to nominal graph densities of 2.5%, 5%, and 10%. For GCN on HCP-Rest at 400 ROIs, lag-5 achieves accuracies of 80.74%, 81.20%, and 80.74%, respectively, showing limited sensitivity to graph density in this setting.

### 7.3 Temporal Signal Processing

As discussed in Section 4, we treat high-amplitude signal retention as one axis of the data-centric design space for brain graph construction. Here, we examine multiple thresholding strategies—including percentile-based $\theta = Q_\alpha(\mathbf{Z}_r)$ and standard deviation-based $\theta = \beta\sigma$—to evaluate whether selectively retaining signal peaks improves the informativeness of functional connectivity patterns.

As shown in Table 3, this particular data-centric design choice, retaining only the top 30% of positive signal amplitudes, improves performance in several HCP-Rest settings but not consistently in HCP-Task. Another strategy we investigated involved setting the threshold at one standard deviation ($\theta = \sigma$), applying the function $\Theta_{\theta_{sd1}\gamma_0}$ to the z-normalized BOLD signal. Additionally, we experimented where the thresholded signal was transformed using $\Theta_{\theta,\gamma=1}$, mapping values above $\theta$ to 1 and otherwise 0. In this binary representation, 0 corresponds to sub-threshold signal components, while 1 represents high-amplitude peaks, further refining the feature selection process. Additional observations are as follows:

- In HCP-Rest, the data-centric choice of high-amplitude signal retention notably improves performance, especially at higher ROI resolutions (400 and 1000), where $\Theta_{\theta,\gamma=1}$ yields substantial gains. Performance benefits align with prior work suggesting high-resolution brain graphs capture finer-grained patterns (Said et al., 2023). Our signal retention further boosts this by focusing on positive high-amplitude fluctuations.

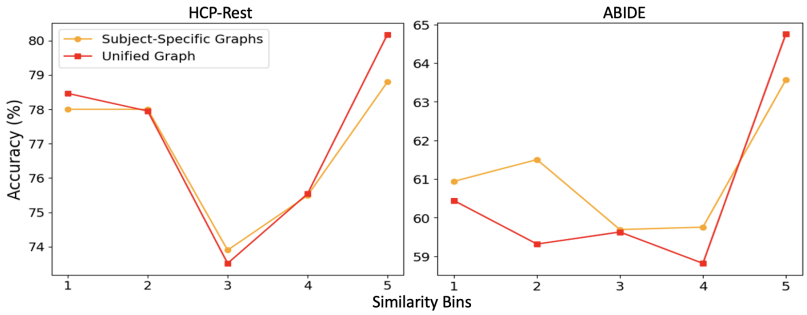



Figure 7: Comparing subject-specific and unified brain topologies on HCP-Rest and ABIDE.

- In HCP-Task, minimal improvement is observed, likely due to already strong task-driven co-activation patterns (Van Essen et al., 2013), suggesting limited room for further signal-level optimization.
- On ABIDE (Table 4), applying $\Theta_{\theta_{sd1}\gamma_1}$ improves overall GNN accuracy at ROI 200, with GCN rising from 60.00% to 65.86%, showing a gain in this evaluated setting.

Table 5: Classification results on HCP-Rest and HCP-Task datasets under different correlation strategies and ROI settings.

| Dataset | | ROI 100 | | | | ROI 400 | | | | ROI 1000 | | | | Overall Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GCN | GAT | SAGE | GNN Avg. | GCN | GAT | SAGE | GNN Avg. | GCN | GAT | SAGE | GNN Avg. | across ROIs & GNNs |
| Rest | **Baseline** | 74.07 ± 1.13 | 74.35 ± 2.56 | 71.11 ± 3.76 | 73.18 | 74.63 ± 0.68 | 76.02 ± 0.80 | 75.37 ± 2.08 | 75.34 | 81.39 ± 2.81 | 78.33 ± 3.43 | 81.48 ± 1.80 | 80.4 | 76.31 |
| | **Spearman** | 73.06 ± 1.48 | 72.69 ± 2.57 | 73.43 ± 2.69 | 73.06 | 77.69 ± 1.93 | 75.93 ± 2.11 | 76.20 ± 2.30 | 76.61 | 82.13 ± 1.23 | 79.17 ± 5.08 | 80.74 ± 1.19 | 80.68 | 76.78 |
| | **Kendall** | 74.54 ± 3.10 | 73.70 ± 3.02 | 75.09 ± 1.72 | 74.44 | 78.52 ± 1.42 | 77.13 ± 2.48 | 76.85 ± 2.23 | 77.5 | 84.07 ± 0.86 | 81.02 ± 3.60 | 82.31 ± 1.53 | 82.47 | 78.14 |
| Task | **Baseline** | 93.29 ± 0.65 | 93.32 ± 0.60 | 93.54 ± 0.58 | 93.38 | 94.97 ± 0.15 | 96.17 ± 0.34 | 96.18 ± 0.12 | 95.77 | 96.18 ± 0.49 | 96.46 ± 0.33 | 96.52 ± 0.66 | 96.39 | 95.18 |
| | **Spearman** | 93.08 ± 0.72 | 93.01 ± 0.39 | 93.99 ± 0.31 | 93.36 | 95.62 ± 0.61 | 96.25 ± 0.23 | 96.82 ± 0.18 | 96.23 | 95.41 ± 0.74 | 96.83 ± 0.40 | 96.66 ± 0.55 | 96.30 | 95.30 |
| | **Kendall** | 93.64 ± 1.10 | 93.64 ± 0.29 | 93.53 ± 0.54 | 93.60 | 95.51 ± 0.46 | 96.48 ± 0.31 | 96.47 ± 0.53 | 96.16 | 96.13 ± 0.48 | 96.60 ± 0.36 | 97.22 ± 0.31 | 96.65 | 95.47 |

### 7.4 Topology Extraction

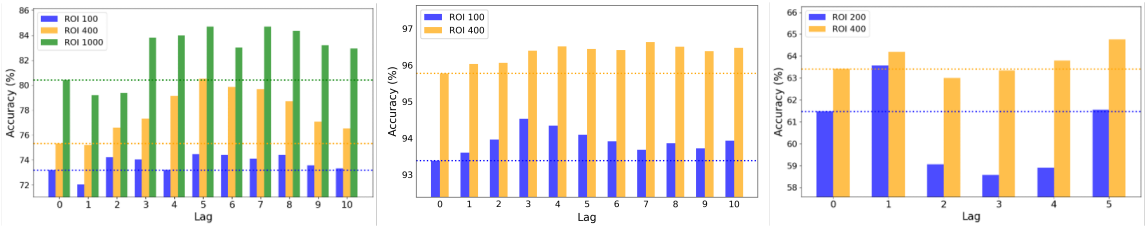
In Section 5.1, we explore the use of alternative correlation measures as data-centric design choices for brain graph topology construction. These rank-based metrics aim to better capture non-linear dependencies and reduce sensitivity to outliers compared to Pearson correlation. As shown in Table 5, the effect of replacing Pearson with Spearman or Kendall varies across datasets, ROI resolutions, and GNN architectures. We summarizes the findings here as:

- On HCP-Task, using alternative correlation coefficients yields overall improvements over Pearson, demonstrating the value of selecting appropriate correlation metrics as a data-centric design choice.
- In HCP-Rest, Kendall achieves the strongest GNN-averaged performance at 400 ROIs, suggesting its effectiveness in capturing informative patterns for functional connectivity in this setting. Spearman also outperforms Pearson at this resolution in the GNN-averaged results. Results on HCP-Task (Table 5) and ABIDE (Table 6) further show that the preferred correlation measure depends on the dataset and setting.

In addition, we explore a unified graph topology as a data-centric design choice aimed at enhancing the consistency of brain network representations across subjects. We construct

Table 6: Classification accuracy of GNNs on ABIDE dataset under different correlation strategies and ROI settings.

| | ROI 200 | | | ROI 400 | | | Avg. across |
|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | GCN | GAT | SAGE | ROIs & GNNs |
| **Baseline** | 60.00 | 62.13 | 62.32 | 62.32 | 64.06 | 63.86 | 62.45 |
| **Spearman** | 61.93 | 65.22 | 60.87 | 63.00 | 63.29 | 63.48 | 62.97 |
| **Kendall** | 62.13 | 63.09 | 61.06 | 63.77 | 63.77 | 63.38 | 62.87 |




(a) HCP-Rest (b) HCP-Task (c) ABIDE

Figure 8: Performance comparison using varying lag/lead correlations across the datasets HCP-Rest, HCP-Task, and ABIDE. HCP-Task results use participant-disjoint evaluation.

global unified graphs as discussed in Section 5.2 and then to evaluate its impact, we stratify test samples into bins based on the overlap between their subject-specific and unified graphs, using edge similarity scores. Due to data imbalance, the lower-similarity bins are combined for analysis. Since task-based fMRI reflects condition-specific co-activation patterns (Van Essen et al., 2013), we apply the unified graph only to resting-state datasets. Empirical results, shown in Figure 7, illustrate the effects of this design choice on downstream performance. We observe that in the HCP-Rest and ABIDE datasets, the unified topology yields performance gains particularly for samples with high structural similarity to the aggregated graph. This highlights that even simple data-centric strategies for enforcing structural consistency, e.g., shared topology aggregation, can enhance GNN performance.

### 7.5 Advanced Graph Featurization

As discussed in Section 6.1, we explore lagged correlation as a data-centric design choice to incorporate dynamic properties into brain graphs. Rather than relying solely on instantaneous co-activation between ROIs, we augment the node feature matrix by concatenating leading and lagging correlations, capturing temporal dependencies in functional connectivity. As shown in Figure 8, this strategy consistently improves classification performance across datasets. Additional details are provided in Appendix Tables 9 and 10, with further analysis of lag effects discussed below.

- The performance gains observed on HCP-Rest and HCP-Task suggest that temporal offsets capture temporally offset BOLD dependencies beyond what instantaneous correlation provides.

- In HCP-Rest and HCP-Task, lag values between 3 and 8 time points provide the most benefit across mental states, especially at higher ROI resolutions where finer-grained dynamics are better captured.
- In ABIDE, a lag of 1 yields the most improvement, likely due to its longer repetition time (TR $\approx$ 2s) (Di Martino et al., 2014) compared to HCP (TR $\approx$ 0.72s) (Van Essen et al., 2013). These findings support evaluating lag as a data-centric design parameter.

We also evaluate balanced accuracy and AUROC on ABIDE; balanced accuracy closely tracks raw accuracy, and full AUROC results are reported in Appendix Table 14. Additionally, for edge features, we encode multi-perspective connectivity by assigning each edge a binary vector indicating its presence in the Pearson, Spearman, or Kendall correlation graphs, resulting in an edge feature matrix of shape $|E| \times 3$. Incorporating these edge-level signals into the standard GNN pipeline yields performance improvements in several settings, as shown in Tables 11 and 12 in the Appendix D. To further expand the data-centric design space, we augment node features by concatenating correlation matrices computed from multiple correlation types (Pearson, Spearman, Kendall) and multiple temporal offsets (lag/lead), yielding an enriched node feature matrix of shape $n \times 5n$. These wider multi-view representations increase memory requirements and can cause out-of-memory failures at high ROI resolutions. These strategies systematically introduce diverse functional connectivity cues into the graph representation. As detailed in Appendix D, we observe that when combined with lagged correlation, the most beneficial lag window for HCP lies between 3 and 8 time points, reinforcing its value as a meaningful data-centric hyperparameter.

**Limitations.** Our evaluation is limited to HCP and ABIDE, so the preferred graph construction may differ for other populations, scanners, acquisition protocols, preprocessing pipelines, and atlases. HCP family relationships are not controlled, and the ABIDE evaluation is not site-disjoint. High-ROI configurations can be computationally demanding, and data-centric optimization may amplify dataset-specific structure. Further design-space exploration can consider graph density, negative edges, partial/precision correlation, atlas choice, nuisance regression, frequency filtering, and alternative dynamic-connectivity constructions.

## 8 Conclusion

In this work, we introduced a data-centric AI framework that systematically explores representative choices within the design space of brain graph construction from fMRI data. By disentangling and evaluating the effects of feature selection, topology extraction, and graph featurization, we demonstrate gains in downstream GNN classification, with the most effective choices varying across datasets and model settings. Rather than proposing a fixed method, our findings underscore the critical role of data processing choices in shaping brain graph quality and highlight the importance of incorporating these design decisions into the development of effective neuroimaging pipelines.

## Broader Impact Statement

This work systematizes brain-graph construction from fMRI as a data-centric design space and benchmarks how upstream choices affect downstream graph learning. Potential positive impacts include improving transparency and reproducibility in neuroimaging graph machine learning by making graph-construction decisions explicit and systematically comparable, which may help researchers build more reliable pipelines and reduce trial-and-error model tuning. Such improvements can support scientific studies on brain-state decoding and disease-related modeling, but this paper does not constitute a clinical diagnostic tool.

Potential negative impacts stem from misuse or overgeneralization. Because our benchmarks are conducted on specific public datasets and preprocessing assumptions, applying the same configurations to different populations, scanners, or acquisition protocols without validation could lead to unreliable conclusions, or encourage over-interpretation of correlational findings as individualized clinical or causal claims. Moreover, dataset demographic imbalances may cause certain data-centric choices to perform unevenly across subgroups, raising fairness concerns if adapted for high-stakes decisions. As brain imaging is sensitive human-derived data, stronger modeling pipelines may also amplify risks of privacy-invasive or discriminatory downstream uses.

To mitigate these risks, we emphasize clear scope limitations (research benchmarking rather than clinical deployment), transparent reporting of all design choices and selection procedures, and evaluation on held-out test sets. We release code and implementation details to facilitate auditing and replication, and we encourage future work to test cross-site generalization and subgroup robustness before any real-world adoption.

## Acknowledgments and Disclosure of Funding

This research is supported by the National Science Foundation (NSF) under grant numbers IIS2239881 and ECCS2325417.

## References

Hessam Ahmadi, Emad Fatemizadeh, and Ali Motie-Nasrabadi. A comparative study of correlation methods in functional connectivity analysis using fmri data of alzheimer's patients. *Journal of Biomedical Physics & Engineering*, 13(2):125, 2023.

Mohammad R Arbabshirani, Sergey Plis, Jing Sui, and Vince D Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145: 137–165, 2017.

César Caballero-Gaudes and Richard C Reynolds. Methods for cleaning the bold fmri signal. *Neuroimage*, 154:128–149, 2017.

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, 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, 2022a.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International conference on medical image computing and computer-assisted intervention*, pages 375–385. Springer, 2022b.

Simon Dahan, Logan ZJ Williams, Daniel Rueckert, and Emma C Robinson. Improving phenotype prediction using long-range spatio-temporal dynamics of functional connectivity. In *Machine Learning in Clinical Neuroimaging: 4th International Workshop, MLCN 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*, pages 145–154. Springer, 2021.

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.

Yuhui Du, Songke Fang, Xingyu He, and Vince D Calhoun. A survey of brain functional network extraction methods using fmri data. *Trends in Neurosciences*, 2024.

Mark Fiecas, Hernando Ombao, Dan Van Lunen, Richard Baumgartner, Alexandre Coimbra, and Dai Feng. Quantifying temporal correlations: A test–retest evaluation of functional connectivity in resting-state fmri. *NeuroImage*, 65, 2013.

Soham Gadgil, Qingyu Zhao, Adolf Pfefferbaum, Edith V Sullivan, Ehsan Adeli, and Kilian M Pohl. Spatio-temporal graph convolution for resting-state fmri analysis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*, pages 528–538. Springer, 2020.

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.

James S Hyde and Andrzej Jesmanowicz. Cross-correlation: an fmri signal-processing strategy. *NeuroImage*, 62(2):848–851, 2012.

Johannes Jakubik, Michael Vössing, Niklas Kühl, Jannis Walk, and Gerhard Satzger. Data-centric artificial intelligence. *Business & Information Systems Engineering*, pages 1–9, 2024.

Mohammad Hossein Jarrahi, Ali Memariani, and Shion Guha. The principles of data-centric ai (dcai). *arXiv preprint arXiv:2211.14611*, 2022.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. *Advances in Neural Information Processing Systems*, 34:4314–4327, 2021.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015.

Gaotang Li, Marlena Duda, Xiang Zhang, Danai Koutra, and Yujun Yan. Interpretable sparsification of brain graphs: Better practices and effective designs for graph neural networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1223–1234, 2023.

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.

Xiao Liu and Jeff H Duyn. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 110(11):4392–4397, 2013.

Xuexiong Luo, Jia Wu, Jian Yang, Shan Xue, Amin Beheshti, Quan Z Sheng, David McAlpine, Paul Sowman, Alexis Giral, and Philip S Yu. Graph neural networks for brain graph learning: A survey. *arXiv preprint arXiv:2406.02594*, 2024.

Guillaume Marrelec, Alexandre Krainik, Hugues Duffau, Mélanie Pélégrini-Issac, Stéphane Lehéricy, Julien Doyon, and Habib Benali. Partial correlation for functional brain interactivity investigation in functional mri. *Neuroimage*, 32(1):228–237, 2006.

Andrew Ng. Landing ai. `https://landing.ai/`, 2023. Retrieved February 8, 2025.

Sergey M Plis, Devon R Hjelm, Ruslan Salakhutdinov, Elena A Allen, Henry J Bockholt, Jeffrey D Long, Hans J Johnson, Jane S Paulsen, Jessica A Turner, and Vince D Calhoun. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, 8:229, 2014.

Jonathan D Power, Mark Plitt, Timothy O Laumann, and Alex Martin. Sources and implications of whole-brain fmri signals in humans. *Neuroimage*, 146:609–625, 2017.

Guillaume A Rousselet and Cyril R Pernet. Improving standards in brain-behavior correlation analyses. *Frontiers in human neuroscience*, 6:119, 2012.

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.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

Prerna Singh. Systematic review of data-centric approaches in artificial intelligence and machine learning. *Data Science and Management*, 6(3), 2023.

Irmak Sivgin, Hasan Atakan Bedel, Saban Ozturk, and Tolga Çukur. A plug-in graph neural network to boost temporal sensitivity in fmri analysis. *IEEE Journal of Biomedical and Health Informatics*, 2024.

Stephen M Smith. Overview of fmri analysis. *The British Journal of Radiology*, 77(suppl_2): S167–S175, 2004.

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), 2011.

Stephen M Smith, Diego Vidaurre, Christian F Beckmann, Matthew F Glasser, Mark Jenkinson, Karla L Miller, Thomas E Nichols, Emma C Robinson, Gholamreza Salimi-Khorshidi, Mark W Woolrich, et al. Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, 17(12):666–682, 2013.

Enzo Tagliazucchi, Pablo Balenzuela, Daniel Fraiman, and Dante R Chialvo. Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis. *Frontiers in physiology*, 3:15, 2012.

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.

Sandra Vieira, Walter HL Pinaya, and Andrea Mechelli. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. volume 74. Elsevier, 2017.

Yu Wang, Yuying Zhao, Yushun Dong, Huiyuan Chen, Jundong Li, and Tyler Derr. Improving fairness in graph neural networks via mitigating sensitive attribute leakage. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 1938–1948, 2022.

Rand R Wilcox and Guillaume A Rousselet. An updated guide to robust statistical methods in neuroscience. *Current Protocols*, 3(3):e719, 2023.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Weizheng Yan, Gang Qu, Wenxing Hu, Anees Abrol, Biao Cai, Chen Qiao, Sergey M Plis, Yu-Ping Wang, Jing Sui, and Vince D Calhoun. Deep learning in neuroimaging: Promises and challenges. *IEEE Signal Processing Magazine*, 39(2):87–98, 2022.

Cheng Yang, Deyu Bo, Jixi Liu, Yufei Peng, Boyu Chen, Haoran Dai, Ao Sun, Yue Yu, Yixin Xiao, Qi Zhang, et al. Data-centric graph learning: A survey. *arXiv preprint arXiv:2310.04987*, 2023.

Jiaxuan You, Zhitao Ying, and Jure Leskovec. Design space for graph neural networks. *Advances in Neural Information Processing Systems*, 33:17009–17021, 2020.

Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, and Xia Hu. Data-centric ai: Perspectives and challenges. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, pages 945–948. SIAM, 2023.

Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. Data-centric artificial intelligence: A survey. *ACM Computing Surveys*, 57(5):1–42, 2025.

Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An end-to-end deep learning architecture for graph classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Qi Zhu, Shengrong Li, Xiangshui Meng, Qiang Xu, Zhiqiang Zhang, Wei Shao, and Daoqiang Zhang. Spatio-temporal graph hubness propagation model for dynamic brain network classification. IEEE, 2024.

## Appendix A. Correlation-based Edge Construction Details

Here we formally define Pearson Correlation Coefficient, Spearman Correlation, and Kendall's Tau, in the context of fMRI neuroimaging data, i.e., signal relations between ROIs.

*Pearson Correlation Coefficient* between ROI $r_i$ and $r_j$ is defined as:

$$\rho_{r_i,r_j} = \frac{\text{cov}(\mathbf{T}_{r_i}, \mathbf{T}_{r_j})}{\sigma_{r_i}\sigma_{r_j}} \tag{6}$$

where $\sigma_x$ and $\sigma_y$ correspond respectively to the standard deviation of signal $\mathbf{T}_{r_i}$ and $\mathbf{T}_{r_j}$, respectively, and defined in Equation 2. We note for simplicity here we use $\mathbf{T}_{r_i}$ to denote the temporal BOLD signal of ROI $r_i$, but is similarly defined if combining with the high-amplitude signal retention process defined in Section 4.

*Spearman Correlation* is a non-parametric rank correlation that measures the strength and direction of a monotonic relationship between two variables (Ahmadi et al., 2023). It first ranks the data and is then calculated thereafter following the same method in Equation 6.

*Kendall's Tau* is another widely used non-parametric method for computing the correlation between two variables (Ahmadi et al., 2023). It evaluates the probability a randomly chosen pair of observations will have the same order in both variables. When applied to two signals, $\mathbf{T}_{r_i}$ and $\mathbf{T}_{r_j}$ (for regions $r_i$ and $r_j$), Kendall's Tau quantifies the degree of ordinal association between them. Specifically, for each pair of time points $(t_a, t_b)$, where $a < b$, the pair is considered concordant if the relative ordering of the signals $T_{r_i}(t_a)$, $T_{r_j}(t_a)$ is the same as that of $T_{r_i}(t_b)$, $T_{r_j}(t_b)$. Conversely, the pair is discordant if the order is reversed. This ensures that it captures the consistency of rank ordering between the signals $\mathbf{T}_{r_i}$ and $\mathbf{T}_{r_j}$ over time, making it a robust and interpretable alternative to Pearson's and Spearman's correlation coefficients. Kendall's Tau is calculated as follows:

$$\tau_{r_i,r_j} = \frac{n_c(r_i, r_j) - n_d(r_i, r_j)}{\frac{1}{2}n(n-1)} \tag{7}$$

where $n_c(r_i, r_j)$ and $n_d(r_i, r_j)$ correspond to the number of concordant and discordant pairs between $r_i$ and $r_j$, having $n$ pairs.

## Appendix B. Additional Details on fMRI to Graph Construction

The most common approach for constructing a brain graph from fMRI data is visualized in Figure 1 and involves the following steps: First, the fMRI data must be carefully pre-processed to remove known artifacts and noise. After preprocessing, the fMRI signals are extracted and averaged within the same ROIs, which are pre-defined based on a brain atlas template. Next, pairwise correlation coefficients are calculated for each pair of ROIs to construct a functional connectivity (FC) matrix based on their representive z-normalized time series. The Pearson correlation coefficient is the most commonly used measure in previous research. This FC matrix is often used as the node feature in the graph. Finally, edges are defined by applying a threshold to the FC matrix, representing strong FC in the brain graph while avoiding an undesired fully connected graph (Said et al., 2023); we take top 5% positive correlation value in the functional connectivity matrix as the edges in the graph.

## Appendix C. Dataset Details

**Human Connectome Project (HCP) S1200** (Van Essen et al., 2013): This publicly available neuroimaging dataset was collected from healthy young adult participants and includes both resting-state and task-based fMRI data. For the resting-state dataset, participants remained at rest during the scan, and their demographics served as labels for graph classification. We used resting-state fMRI data from 1,078 subjects, excluding those with fewer than 1,200 frames. Our methods were applied to the task of gender classification, where each subject's gender served as the label (2 classes). For the task dataset, participants performed specific cognitive or motor tasks during the scan. The dataset includes over 7,000 scans, covering seven task types: working memory, social cognition, relational processing, motor function, language, gambling, and emotion. The type of task performed served as the label for graph classification (7 classes). For preprocessing, we used the Minimally Preprocessed HCP Pipeline for all scans (Glasser et al., 2013), followed by additional steps to enhance data quality. We applied brain parcellation using the Schaefer atlas (Schaefer et al., 2018) to extract the mean fMRI timeseries for each region of interest (ROI). To reduce instrumental noise, we removed scanner drifts by regressing out linear and quadratic trends. Motion artifacts were mitigated using rigid-body motion regression, removing six motion parameters and their derivatives before graph construction.

The HCP-Task experiments in the main tables use split seed 123 and model seeds 123–127. The ten participant-disjoint robustness splits use split seeds 123, 124, 125, 126, 127, 128, 129, 221, 223, and 224.

**Autism Brain Imaging Data Exchange (ABIDE)** (Di Martino et al., 2014): A collaborative initiative involving 17 international research sites, aimed at enhancing our understanding of Autism Spectrum Disorder (ASD). ABIDE provides resting-state fMRI data from individuals with and without ASD. We used data from 1,009 subjects, with binary labels indicating ASD diagnosis. For preprocessing, we applied the C-PAC pipeline (Craddock et al., 2012), including brain parcellation, slice timing correction, motion correction, and spatial normalization to the MNI152 template before graph construction.

## Appendix D. Detailed Result Tables

In Tables 7 and 8 we report the more detailed results of Tables 4 and 6, respectively. We also provide the full detailed results tables for selected lagged correlation experiments, specifically Tables 9 and 10. Finally, the detailed results for the edge featurization are provided in Tables 11 and 12.

Table 7: Classification performance of GNNs on ABIDE dataset under different feature selection strategies and ROI settings.

| | ROI 200 | | | | ROI 400 | | | | Avg. across ROIs & GNNs |
|---|---|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | Avg. across GNNs | GCN | GAT | SAGE | Avg. across GNNs | |
| **Baseline** | 60.00 ± 2.94 | 62.13 ± 2.69 | 62.32 ± 1.67 | 61.48 | 62.32 ± 3.25 | 64.06 ± 1.61 | 63.86 ± 2.15 | 63.41 | 62.45 |
| $\Theta_{\theta_{p30}\gamma_0}$ | 61.74 ± 3.14 | 62.13 ± 1.58 | 62.22 ± 2.19 | 62.03 | 61.64 ± 2.84 | 63.09 ± 2.49 | 57.87 ± 4.30 | 60.87 | 61.45 |
| $\Theta_{\theta_{p30}\gamma_1}$ | 58.94 ± 6.11 | 61.45 ± 4.19 | 61.35 ± 2.63 | 60.58 | 62.13 ± 1.82 | 63.96 ± 1.13 | 60.00 ± 5.29 | 62.03 | 61.31 |
| $\Theta_{\theta_{sd1}\gamma_0}$ | 62.02 ± 2.22 | 62.32 ± 6.70 | 61.72 ± 3.28 | 62.02 | 60.77 ± 4.24 | 62.73 ± 1.17 | 61.53 ± 4.05 | 61.68 | 61.85 |
| $\Theta_{\theta_{sd1}\gamma_1}$ | 65.86 ± 2.11 | 63.03 ± 4.86 | 63.84 ± 2.36 | 64.24 | 61.64 ± 1.27 | 61.75 ± 0.91 | 62.08 ± 2.71 | 61.82 | 63.03 |

Table 8: Classification performance of GNNs on ABIDE dataset under different correlation strategies and ROI settings.

| | ROI 200 | | | | ROI 400 | | | | Avg. across ROIs & GNNs |
|---|---|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | Avg. across GNNs | GCN | GAT | SAGE | Avg. across GNNs | |
| **Baseline** | 60.00 ± 2.94 | 62.13 ± 2.69 | 62.32 ± 1.67 | 61.48 | 62.32 ± 3.25 | 64.06 ± 1.61 | 63.86 ± 2.15 | 63.41 | 62.45 |
| **Spearman** | 61.93 ± 1.77 | 65.22 ± 2.66 | 60.87 ± 2.03 | 62.67 | 63.00 ± 3.37 | 63.29 ± 1.53 | 63.48 ± 2.83 | 63.26 | 62.97 |
| **Kendall** | 62.13 ± 3.03 | 63.09 ± 1.48 | 61.06 ± 1.87 | 62.09 | 63.77 ± 1.76 | 63.77 ± 1.65 | 63.38 ± 2.13 | 63.64 | 62.87 |

Table 9: Classification performance of GNNs on HCP-Rest dataset under different lagged correlation values and ROI settings.

| | ROI 100 | | | | ROI 400 | | | | ROI 1000 | | | | Avg. ROIs & GNNs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | |
| **Baseline** | 74.07 ± 1.13 | 74.35 ± 2.56 | 71.11 ± 3.76 | 73.18 | 74.63 ± 0.68 | 76.02 ± 0.80 | 75.37 ± 2.08 | 75.34 | 81.39 ± 2.81 | 78.33 ± 3.43 | 81.48 ± 1.80 | 80.4 | 76.31 |
| **Lag = 1** | 72.13 ± 3.43 | 71.30 ± 2.21 | 72.69 ± 1.17 | 72.04 | 75.56 ± 2.22 | 74.91 ± 3.67 | 75.19 ± 1.12 | 75.22 | 78.33 ± 8.95 | 80.09 ± 2.41 | OOM | 79.21 | 75.03 |
| **Lag = 2** | 76.39 ± 2.57 | 73.06 ± 2.86 | 73.15 ± 2.01 | 74.2 | 76.94 ± 1.26 | 76.48 ± 2.58 | 76.30 ± 0.99 | 76.57 | 78.15 ± 10.90 | 80.56 ± 1.94 | OOM | 79.36 | 76.38 |
| **Lag = 3** | 74.17 ± 3.42 | 73.89 ± 1.96 | 74.07 ± 1.80 | 74.04 | 77.96 ± 1.57 | 75.56 ± 1.26 | 78.43 ± 2.10 | 77.32 | 84.91 ± 1.65 | 82.69 ± 0.86 | OOM | 83.8 | 77.71 |
| **Lag = 4** | 73.06 ± 1.81 | 73.06 ± 2.49 | 73.52 ± 1.42 | 73.21 | 79.81 ± 2.32 | 78.89 ± 1.54 | 78.70 ± 0.97 | 79.13 | 84.72 ± 0.65 | 83.24 ± 3.23 | OOM | 83.98 | 78.13 |
| **Lag = 5** | 75.37 ± 1.29 | 73.33 ± 1.33 | 74.63 ± 1.79 | 74.44 | 81.85 ± 1.74 | 79.26 ± 0.90 | 80.46 ± 1.35 | 80.52 | 84.07 ± 1.48 | 85.28 ± 2.04 | OOM | 84.68 | 79.28 |
| **Lag = 6** | 73.80 ± 3.77 | 74.91 ± 1.42 | 74.44 ± 1.72 | 74.38 | 80.00 ± 2.24 | 80.46 ± 2.51 | 79.07 ± 0.80 | 79.84 | 83.24 ± 3.25 | 82.78 ± 2.06 | OOM | 83.01 | 78.59 |
| **Lag = 7** | 73.52 ± 1.79 | 75.46 ± 1.52 | 73.33 ± 2.51 | 74.1 | 78.80 ± 2.18 | 81.11 ± 1.64 | 79.07 ± 1.50 | 79.66 | 85.09 ± 1.45 | 84.26 ± 1.13 | OOM | 84.68 | 78.83 |
| **Lag = 8** | 73.80 ± 1.62 | 75.46 ± 1.99 | 73.89 ± 0.81 | 74.38 | 79.91 ± 1.33 | 78.24 ± 1.28 | 77.96 ± 1.45 | 78.7 | 84.26 ± 1.37 | 84.44 ± 1.51 | OOM | 84.35 | 78.5 |
| **Lag = 9** | 74.26 ± 2.39 | 72.69 ± 1.68 | 73.80 ± 2.04 | 73.58 | 77.41 ± 1.84 | 77.78 ± 1.40 | 76.02 ± 0.61 | 77.07 | 81.85 ± 2.24 | 84.54 ± 1.36 | OOM | 83.2 | 77.29 |
| **Lag = 10** | 72.96 ± 1.67 | 75.19 ± 1.93 | 71.85 ± 2.24 | 73.33 | 75.83 ± 0.54 | 77.41 ± 1.07 | 76.30 ± 3.59 | 76.51 | 82.41 ± 1.40 | 83.43 ± 1.07 | OOM | 82.92 | 76.92 |

Table 10: Classification performance of GNNs on ABIDE dataset under different lagged correlation values and ROI settings.

| | ROI 200 | | | | ROI 400 | | | | Avg. ROIs & GNNs |
|---|---|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | |
| **Baseline** | 60.00 ± 2.94 | 62.13 ± 2.69 | 62.32 ± 1.67 | 61.48 | 62.32 ± 3.25 | 64.06 ± 1.61 | 63.86 ± 2.15 | 63.41 | 62.45 |
| **Lag = 1** | 63.96 ± 1.58 | 64.64 ± 1.20 | 62.13 ± 2.04 | 63.58 | 65.51 ± 2.97 | 61.16 ± 2.74 | 65.89 ± 4.09 | 64.19 | 63.88 |
| **Lag = 2** | 61.35 ± 1.88 | 57.39 ± 2.49 | 58.45 ± 2.57 | 59.06 | 61.55 ± 3.17 | 60.77 ± 5.12 | 66.67 ± 5.37 | 63.00 | 61.03 |
| **Lag = 3** | 59.52 ± 3.58 | 57.29 ± 3.66 | 58.94 ± 4.12 | 58.58 | 61.74 ± 4.43 | 63.19 ± 7.18 | 65.12 ± 4.69 | 63.35 | 60.97 |
| **Lag = 4** | 58.26 ± 5.01 | 57.78 ± 2.78 | 60.68 ± 2.71 | 58.91 | 63.96 ± 3.45 | 63.29 ± 4.39 | 64.15 ± 4.44 | 63.80 | 61.35 |
| **Lag = 5** | 60.68 ± 2.47 | 61.84 ± 2.57 | 62.13 ± 1.95 | 61.55 | 64.64 ± 2.67 | 67.15 ± 4.81 | 62.51 ± 3.85 | 64.77 | 63.16 |

Table 11: Classification results on HCP-Rest and HCP-Task with and without edge features.

| Dataset | | ROI 100 | | | | ROI 400 | | | | ROI 1000 | | | | Avg. across |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | ROIs & GNNs |
| **Rest** | **Baseline** | 74.07 ± 1.13 | 74.35 ± 2.56 | 71.11 ± 3.76 | 73.18 | 74.63 ± 0.68 | 76.02 ± 0.80 | 75.37 ± 2.08 | 75.34 | 81.39 ± 2.81 | 78.33 ± 3.43 | 81.48 ± 1.80 | 80.40 | 76.31 |
| | **w/ Edge Feat.** | 70.22 ± 1.33 | 69.29 ± 1.43 | 71.30 ± 0.38 | 70.27 | 75.46 ± 2.27 | 77.62 ± 2.28 | 75.00 ± 2.73 | 76.03 | 80.56 ± 1.31 | 83.64 ± 3.86 | 81.64 ± 1.22 | 81.95 | 76.08 |
| **Task** | **Baseline** | 91.05 ± 0.29 | 91.93 ± 0.50 | 91.71 ± 0.30 | 91.56 | 94.57 ± 0.60 | 95.08 ± 0.51 | 95.51 ± 0.19 | 95.05 | 96.06 ± 0.36 | 96.26 ± 0.62 | 96.49 ± 0.28 | 96.27 | 94.30 |
| | **w/ Edge Feat.** | 91.22 ± 0.98 | 92.05 ± 1.63 | 92.93 ± 0.43 | 92.07 | 94.09 ± 0.58 | 95.46 ± 0.28 | 95.90 ± 0.33 | 95.15 | OOM | OOM | OOM | / | 93.61 |

Table 12: Classification results on ABIDE with and without edge features.

| | ROI 200 | | | | ROI 400 | | | | Avg. ROIs & GNNs |
|---|---|---|---|---|---|---|---|---|---|
| | GCN | GAT | SAGE | Avg. GNNs | GCN | GAT | SAGE | Avg. GNNs | |
| **Baseline** | 60.00 ± 2.94 | 62.13 ± 2.69 | 62.32 ± 1.67 | 61.48 | 62.32 ± 3.25 | 64.06 ± 1.61 | 63.86 ± 2.15 | 63.41 | 62.45 |
| **w/ Edge Feat** | 61.84 ± 0.79 | 61.03 ± 2.28 | 60.39 ± 2.20 | 61.09 | 63.29 ± 3.88 | 64.09 ± 0.46 | 61.67 ± 2.41 | 63.02 | 62.05 |

Table 13: GCN paired accuracy differences relative to Pearson over ten participant-disjoint HCP-Task splits. Differences and 95% confidence intervals are in percentage points. BH correction covers all nine comparisons. At 100 ROIs, split-level accuracies average replicates from two compute environments.

| ROI | Method | $\Delta$ (pp) | 95% CI | $q$ |
|---|---|---|---|---|
| 100 | Spearman | +0.69 | [0.24, 1.14] | 0.016 |
| | Kendall | +0.79 | [0.45, 1.13] | 0.0024 |
| | Lag=5 | +0.86 | [0.62, 1.10] | 0.0002 |
| 400 | Spearman | +0.49 | [0.04, 0.94] | 0.0555 |
| | Kendall | +0.17 | [−0.16, 0.51] | 0.31 |
| | Lag=5 | +1.21 | [0.62, 1.80] | 0.0036 |
| 1000 | Spearman | +0.16 | [−0.25, 0.58] | 0.40 |
| | Kendall | +0.33 | [−0.20, 0.86] | 0.25 |
| | Lag=5 | +0.64 | [0.13, 1.14] | 0.034 |

Table 14: Additional evaluation metrics on ABIDE at the default 5% graph density. Accuracy is reported as mean ± standard deviation over five model seeds. AUROC provides an additional assessment of discriminative performance.

| **ROI** | **Method** | **Accuracy (%)** | **AUROC** |
|---|---|---|---|
| 200 | Pearson | 60.00 ± 2.94 | 0.648 ± 0.015 |
| | Spearman | 61.93 ± 1.77 | 0.633 ± 0.014 |
| | Kendall | 62.13 ± 3.03 | 0.647 ± 0.013 |
| | Lag-5 | 60.68 ± 2.47 | 0.609 ± 0.049 |
| 400 | Pearson | 62.32 ± 3.25 | 0.668 ± 0.007 |
| | Spearman | 63.00 ± 3.37 | 0.654 ± 0.024 |
| | Kendall | 63.77 ± 1.76 | 0.668 ± 0.016 |
| | Lag-5 | 64.64 ± 2.67 | 0.604 ± 0.023 |