BrainATCL: Adaptive Temporal Brain Connectivity Learning for Functional Link Prediction and Age Estimation

Yiran Huang[1] YH87@NJIT.EDU
Amirhossein Nouranizadeh[1] AN857@NJIT.EDU
Christine Ahrends[2] CHRISTINE.AHRENDS@NDCN.OX.AC.UK
Mengjia Xu[*1] MX6@NJIT.EDU
[1] *Department of Data Science, New Jersey Institute of Technology, Newark, NJ, USA*
[2] *Nuffield Department of Clinical Neurosciences, University of Oxford, UK*

Editors: Accepted for publication at MIDL 2026

Abstract

Functional Magnetic Resonance Imaging (fMRI) is an imaging technique widely used to study human brain activity. fMRI signals in areas across the brain transiently synchronise and desynchronise their activity in a highly structured manner, even when an individual is at rest. These functional connectivity dynamics may be related to behaviour and neuropsychiatric disease. To model these dynamics, temporal brain connectivity representations are essential, as they reflect evolving interactions between brain regions and provide insight into transient neural states and network reconfigurations. However, conventional graph neural networks (GNNs) often struggle to capture long-range temporal dependencies in dynamic fMRI data. To address this challenge, we propose *BrainATCL*, an unsupervised, nonparametric framework for adaptive temporal brain connectivity learning, enabling functional link prediction and age estimation. Our method dynamically adjusts the lookback window for each snapshot based on the rate of newly added edges. Graph sequences are subsequently encoded using a `GINE-Mamba2` backbone to learn spatial-temporal representations of dynamic functional connectivity in resting-state fMRI data of 1,000 participants from the Human Connectome Project. To further improve spatial modeling, we incorporate brain structure and function-informed edge attributes, i.e., the left/right hemispheric identity and subnetwork membership of brain regions, enabling the model to capture biologically meaningful topological patterns. We evaluate our BrainATCL on two tasks: functional link prediction and age estimation. The experimental results demonstrate superior performance and strong generalization, including in cross-session prediction scenarios.

Keywords: Dynamic Brain Functional Connectivity, Mamba, Functional MRI, GNNs, Link Prediction, Age Estimation

1. Introduction

Functional magnetic resonance imaging (fMRI) measures brain activity non-invasively via blood-oxygen-level-dependent (BOLD) signals, serving as a key neuroimaging technique for studying neural processes and supporting disease diagnosis (Perovnik et al., 2023; Biswal and Uddin, 2025). In resting-state fMRI, where no explicit task is performed, spontaneous synchrony and desynchrony across brain regions reveal structured, large-scale networks associated with behavior, cognition, aging, and neurological or psychiatric conditions (Biswal

* Corresponding Author

et al., 1995; Shen, 2015; Smith et al., 2013b). Functional brain connectivity can be modeled from fMRI signals as either static functional connectivity (sFC) or dynamic functional connectivity (dFC) (Power et al., 2011). sFC assumes constant interactions between brain regions throughout the fMRI scan, and is usually measured by the Pearson or partial correlation. Significant works utilize sFC to study brain abnormalities linked to various brain diseases, e.g., Alzheimer's disease (Xu et al., 2020, 2021; Baker et al., 2024) or Autism Spectrum Disorder (Zhang et al., 2022). However, as human brain transitions through dynamic functional states, dFC has emerged as a powerful alternative to sFC by capturing time-varying interactions between brain regions, offering insights into transient neural states and network reconfigurations (Ahrends and Vidaurre, 2025).

Capturing temporal dynamics in dFCs remains challenging due to variability from sampling errors, physiological artifacts, and arousal fluctuations, amplified by the low signal-to-noise ratio inherent in fMRI (Laumann et al., 2024). Recent advances in dFC modeling primarily fall into three categories: **1) Mode/Component decomposition-based methods**, which model dFC by decomposing signals or discrete-time connectivity patterns into interpretable bases, using techniques like temporal covariance EVD (eigenvalue decomposition) (Alteriis et al., 2025), principle component analysis (PCA) (Leonardi et al., 2013), Fisher kernel (Ahrends et al., 2025), and the Koopman operator (Turja et al., 2023). These methods often assume linearity and may struggle to generalize across datasets. **2) Clustering-based methods**, which segment time-varying FC matrices into discrete brain states using methods like k-means (van der Horn et al., 2024) or deep clustering (Spencer and Goodfellow, 2022). However, these methods struggle to capture smooth or overlapping transitions between dynamic connectivity states and reply on ad-hoc parameters. **3) Graph-based dynamic modeling**, which integrate static graph neural networks (GNNs) with different temporal modules, such as temporal convolutional networks (TCNs) (Gadgil et al., 2020; Azevedo et al., 2022), recurrent neural networks (RNNs) (Cao et al., 2022; Tang et al., 2025), attention-based transformers (Kim et al., 2021; Yu et al., 2024; Wang et al., 2024b), to capture complex spatiotemporal dynamics of dFCs.

Despite the strong expressive power, existing methods mostly rely on fixed context windows, suffer from high computational overhead, and overlook neuroscience-informed node and edge attributes – limiting generalizability across subjects or sessions. They also struggle to capture long-range temporal dependencies, hindering accurate and scalable modeling of spatio-temporal patterns in dFCs. However, the recent introduction of Mamba (Dao and Gu, 2024), a state-space model (SSM) with linear complexity, presents a promising direction for efficient and scalable sequence modeling. More related works on SSM can be found from section 2.

In this work, to integrate neuroscience-inspired priors, long-range temporal modeling, and adaptive mechanisms to improve the accuracy and robustness of dFC analysis, we propose **BrainATCL**, an unsupervised, nonparametric framework for adaptive temporal brain connectivity representation learning from resting-state fMRI data. It consists of four key innovations: (1) Adaptive determination of the temporal lookback window at each time point based on the rate of newly formed links, enabling the model to capture meaningful transitions in functional connectivity. (2) Incorporation of structure- and function-informed edge attributes, i.e., hemispheric identity and subnetwork membership, to enhance spatial representations and capture biologically meaningful topological patterns. (3) Encoding

of adaptively generated graph sequences via the `GINE-Mamba2` backbone, which integrates edge-aware spatial message passing with efficient long-range temporal modeling to learn expressive temporal graph embeddings. (4) Evaluation of the proposed BrainATCL framework on two different tasks, *functional link prediction (in both within and cross-session settings)* and *age estimation*, using the Human Connectome Project (HCP) dataset including 1,000 subjects. Experimental results demonstrate its effectiveness and strong generalizability in modeling dFCs, outperforming both static and dynamic graph learning baselines.

2. Related Work

Graph neural networks (GNNs) have emerged as a powerful tool for both spatial and temporal graph modeling via incorporating conventional sequence models such as recurrent neural networks (Xu et al., 2022), long-short term memory (LSTM) (Tang et al., 2025), and (Wang et al., 2025) further incorporate a diffusion connection strategy and adaptive self-attention fusion to enhance the GNN-LSTM model stability and spatiotemporal representation. Recently, transformers have shown great potential for temporal graph modeling via the self-attention mechanism (Varghese et al., 2024). (Yu et al., 2024) develop a brain graph transformer to capture long-range dependencies within brain networks. (Wang et al., 2024b) incorporated the node-level attention algorithm for information aggregation on ROI-based brain graphs. (Kan et al., 2022) propose a brain network transformer that models brain graphs with fixed node order and connection profiles and introduce an orthonormal clustering readout. However, attention-based approaches often suffer from high computational cost with quadratic complexity.

State space models (SSMs) (Gu et al., 2021), particularly advanced Mamba model (Dao and Gu, 2024), have gained increasing interest for their linear complexity and efficiency compared to transformers (Vaswani et al., 2017). Several studies have attempt to apply Mamba for dFC modeling, e.g., Brain-GM (Wang et al., 2024a) marks the first integration of the state space model into dynamic brain graph representation learning; FST-Mamba (Wei et al., 2025) is designed for high-dimensional dFC features that can learn dynamic connectivity patterns and predict biological and cognitive outcomes. (Zrimek et al., 2025)'s method learns dynamic spatial connections, adjusting the graph structure based on temporal variations in movement. However, most existing methods rely on fixed temporal windows and overlook neuroscience-informed edge priors, motivating our adaptive framework development with dynamic context selection and biologically grounded edge features

3. Methodology

3.1. Data Preparation

fMRI preprocessing: we used resting-state fMRI recordings of 1,001 healthy subjects (aged 22-35) from the Young Adult S1200 release of the Human Connectome Project (Van Essen et al., 2013; Smith et al., 2013a). Briefly, data were acquired on a 3T MRI scanner. For each participant, resting-state functional scans were acquired in four sessions of 14 mins 33 sec duration each, using multiband echo planar imaging (EPI) at an acceleration factor of 8, a repetition time (TR) of 0.72 sec, and a spatial resolution of $2 \times 2 \times 2$mm. The data were preprocessed as described in (Smith et al., 2013a) using minimal spatial

preprocessing followed by ICA-based temporal preprocessing and high-pass filtering using a 2000 sec cut-off. Since the voxel-level timecourses are too high-dimensional, we parcellated the brain using the Schaefer parcellation (Schaefer et al., 2018) with 100 regions of interest (ROIs). We chose this relatively low-dimensional parcellation, since estimating dFC from more fine-grained parcellations can lead to overfitting (Ahrends et al., 2022). Timecourses were extracted as the first principal component within each ROI. Timecourses from eight ROIs were removed as noise z-score > 6, resulting in 92 ROIs. The input to our experiment is the BOLD time series of $N = 92$ brain regions for $P = 1000$ subjects over $T = 1200$ time steps, and each subject has $Q = 4$ imaging sessions. Hence, we denote the entire time series by $\mathcal{X} \in \mathbb{R}^{N\times P\times Q\times T}$. All detailed symbols are provided in Appendix Table 4.

Sliding window-based dFC construction: we compute the Pearson correlation coefficient between pairs of brain regions with a sliding window of length w and stride s. Each FC matrix is preprocessed by zeroing the diagonal and setting negative correlations to zero, followed by thresholding off-diagonal entries with τ to retain only positive connections and capture temporal dFC patterns. The choice of threshold τ is essential: larger τ yields more dynamic graphs with higher edge turnover, while a smaller τ produces smoother graphs that may retain more noise. The impact of different τ values on model performance is discussed in Section 4.4. To further understand how τ affects the temporal dynamics of graphs, we provide an in-depth analysis in Appendix A.

The resulting data is a set of all discrete-time dynamic graphs for each subject p and session q denoted by $\mathcal{G} = \{G^{p,q} | p \in \{0,\dots,1000\}, q \in \{0,\dots,3\}\}$, where $G^{p,q} = (G_0^{p,q},\dots,G_k^{p,q},\dots G_{K-1}^{p,q})$, k is the sequence index and $K = \lfloor \frac{1200-w}{s} \rfloor + 1$. Note that each functional connectivity graph is represented by its corresponding adjacency, node and edge feature matrices coupled with their target variables $\mathbf{y}^p$, $G_k^{p,q} = (\mathbf{X}_k^{p,q}, \mathbf{A}_k^{p,q}, \mathbf{E}_k^{p,q}, \mathbf{y}^p)$, where $\mathbf{X}_k^{p,q} \in \mathbb{R}^{N\times F_V}$, $\mathbf{E}_k^{p,q} \in \mathbb{R}^{M\times F_E}$ and $\mathbf{A}_k^{p,q} \in \{0,1\}^{N\times N}$, and F_V, F_E are the number of input node and edge features, respectively.

3.2. Problem Definition

The goal is to learn dFC representations that effectively and efficiently encode their spatiotemporal information. Mathematically, the goal is to learn an embedding function f that maps an arbitrary graph sequence $G^{p,q} \in \mathcal{G}$ into an embedding matrix $\mathbf{H}^{p,q} = f(G^{p,q}) \in \mathbb{R}^{N\times D}$, where each row corresponds to a node embedding vector of dimension D. The learned embeddings $\mathbf{H}^{p,q}$ support two downstream tasks: (1) *dynamic link prediction*, where the next adjacency matrix is estimated as $\hat{\mathbf{A}}_K^{p,q} = g(\mathbf{H}^{p,q})$ with $g : \mathbb{R}^{N\times D} \to \{0,1\}^{N\times N}$ as the link prediction function, and $\hat{\mathbf{A}}_K^{p,q}$ is the estimated adjacency matrix of the graph at time step K; (2) *Age prediction*, a dynamic graph regression task, the subject's age is estimated as $\hat{y}^p = r(\mathbf{H}^{p,q})$, where $r : \mathbb{R}^{N\times D} \to \mathbb{R}_+$ is the regression function.

3.3. The Proposed BrainATCL Framework

In our proposed BrainATCL framework, we approach dFC representation learning and downstream machine learning tasks in two phases followed by data preprocessing illustrated in Fig. 1. The workflow of our BrainATCL is shown in Fig. 2.

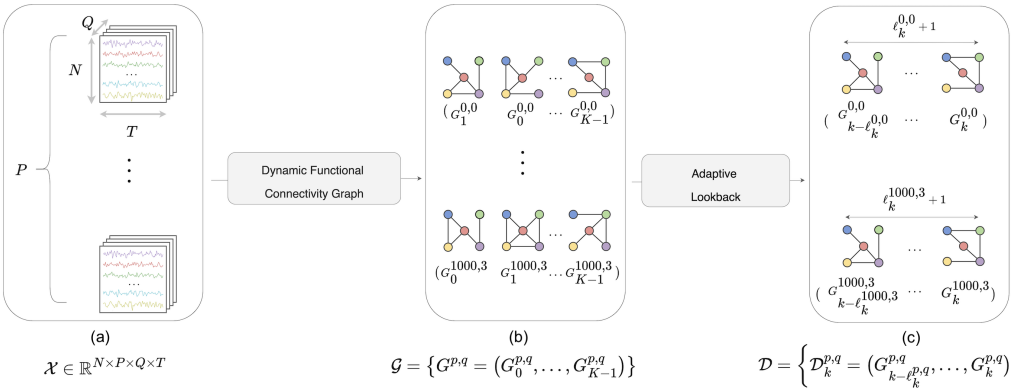

Figure 1: Pipeline for dFC construction and adaptive graph sequence generation. **(a)** The original 4D fMRI BOLD signals $\mathcal{X} \in \mathbb{R}^{N\times P\times Q\times T}$ (with N, P, Q, T denoting ROIs, subjects, imaging sessions, and time points). **(b)** We construct dFC $G^{p,q}$ as a graph sequence of length K for subject p and session q, all subjects' dFC set is denoted as $\mathcal{G}$. **(c)** For training the models, we designed adaptive lookback strategy as a function of the temporal novelty index (Eq. 1) as a heuristic to determine temporal context $\mathcal{D}_k^{p,q}$ dynamically for each graph snapshot at time step k, see detailed computation of adaptive lookback $\ell_k^{p,q}$ in Section 3.3.1.

3.3.1. Novelty-based adaptive lookback computation.

One of the main components of our method is the adaptive computation of the temporal context for each static graph within the dynamic functional connectivity. We define the temporal context as the set of historical graphs prior to each static graph within the graph sequence. We characterize the temporal context of each static graph $G_k^{p,q}$ using the adaptive lookback number $\ell_k^{p,q}$, which indicates the number of past relevant graphs for the static graph corresponding to subject p in session q at time step k. We are motivated to compute the temporal context for each static graph during data preprocessing to introduce an inductive bias. Specifically, our framework explicitly models temporal graph dynamics by aggregating historical context of adaptive length, which alleviates the need for the model to implicitly learn long-range temporal dependencies.

Our intuition is based on the *novelty index* of the dynamic graph, defined as the average ratio of new edges at each time step:

$$n_k = \frac{1}{K}\sum_{k=1}^{K}\frac{|E_k \setminus E_k^{\text{seen}}|}{|E_k|}, \tag{1}$$

where E_k represents the set of edges at time step k, and E_k^{seen} represents the set of edges observed in previous time steps $1, \ldots, k-1$.

We aim to compute the lookback number $\ell_k^{p,q}$ as a function of the novelty index such that the dynamic graph is assigned a shorter lookback at time steps with rapid novel changes and a longer lookback at time steps where the graph evolves more slowly. This intuition is supported by the observation that a dynamic graph undergoing frequent novel changes behaves similarly to a sequence of random graphs, making historical context less relevant,

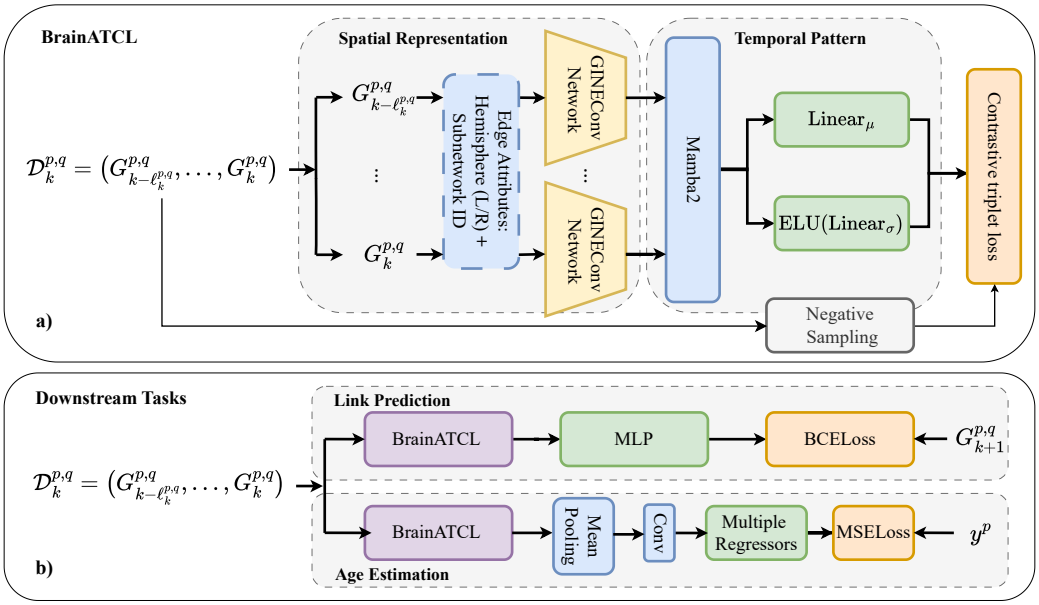

Figure 2: Illustration of our proposed BrainATCL framework for adaptive temporal brain connectivity embedding and two downstream tasks. **(a)** The BrainATCL is trained by minimizing the contrastive triplet loss. It combines a GINEConv-based GNN to encode structural information with edge attributes, followed by a Mamba-2 model for temporal aggregation across the graph sequence. The output of Mamba-2 is fed to linear layers to represent each node as a multivariate Gaussian distribution. **(b)** After pre-training the BrainATCL, the input graph sequence is processed and its node embeddings are used for link prediction classifier and age prediction regressor. These downstream task models are trained separately using binary cross-entropy (BCE) and mean-squared error (MSE) losses.

whereas slow variations in the graph indicate that its structure depends on a longer historical context. We employ the novelty index of the dynamic graph to quantify structural variations and compute the lookback number as a function of this index.

Specifically, let $(n_0^{p,q}, \ldots, n_{K-1}^{p,q})$ be the sequence of novelty indices for the dynamic graph $G^{p,q} = (G_0^{p,q}, \ldots, G_{K-1}^{p,q})$, where each novelty index is computed as defined in Eq. 1. We begin with a desired range of lookback values, $[\ell_{\min}, \ell_{\max}]$, so that $B = \ell_{\max} - \ell_{\min} + 1$ is the number of discrete lookback values. Given the sequence of novelty indices, we compute the $(B+1)$-quantiles as:

$$\alpha_b^{p,q} = \text{Quantile}\big(n_0^{p,q}, \ldots, n_{K-1}^{p,q}; \tfrac{b}{B}\big), \quad b = \{0, \ldots, B\}, \tag{2}$$

so $\alpha_0^{p,q}$ and $\alpha_B^{p,q}$ are equal to the minimum and maximum novelty indices in the sequence, respectively. To compute the adaptive lookback for each time step k of the dynamic graph $G^{p,q}$, we first find the unique bin index $b \in 0, \ldots, B-1$ to which the corresponding novelty index belongs: $\alpha_b^{p,q} \leq n_k^{p,q} < \alpha_{b+1}^{p,q}$, and assign the adaptive lookback value $\ell_k^{p,q} = \ell_{\max} - b$. Thus, graphs with the lowest novelty index ($b = 0$) receive the longest lookback value $\ell_k^{p,q} = \ell_{\max}$, while those with the highest novelty index ($b = B-1$) receive the shortest

lookback value $\ell_k^{p,q} = \ell_{\min}$. We determine an adaptive lookback number for each dynamic graph, resulting in the final dataset. $\mathcal{D}$ structured as follows:

$$\mathcal{D} = \left\{ \mathcal{D}_k^{p,q} = \left(G^{p,q}_{k-\ell_k^{p,q}}, \ldots, G_k^{p,q}\right) \;\middle|\; \begin{array}{l} \ell_{\max} \le k \le K \\ p \in \{0, \ldots, 1000\} \\ q \in \{0, \ldots, 3\} \end{array} \right\} \tag{3}$$

3.3.2. Encoding of dFCs via GINE-Mamba

In the first step of modeling, we employ a contrastive learning approach to encode the spatial and temporal information contained in the dFC graph into the embedding vector space. Inspired by previous work (Bojchevski and Günnemann, 2017; Parmanand Pandey et al., 2024), we model each node in the graph sequence as a multivariate Gaussian distribution by parameterizing its mean vector and covariance matrix. Specifically, for each static graph at time step k in the graph sequence, we encode its structural information using a graph neural network composed of L `GINEConv` (Hu et al., 2019) layers:

$$\mathbf{z}_{k,i}^{\{p,q\},l} = h_{\theta_l}\Big((1+\epsilon)\,\mathbf{z}_{k,i}^{\{p,q\},l-1} + \sum_{j \in \mathcal{N}_k^{p,q}(i)} \mathrm{ReLU}\big(\mathbf{z}_{k,j}^{\{p,q\},l-1} + \mathbf{e}_{k,ji}^{p,q}\big)\Big), \tag{4}$$

where p and q are indicators of the subject and session, respectively. $\mathbf{z}_{k,i}^{\{p,q\},l} \in \mathbb{R}^{D_l}$ is the l^{th} layer's structural embedding of node i at time step k, ϵ is a learnable parameter, $l = 1, \ldots, L$ is the number of `GINEConv` layers, $\mathcal{N}_k^{p,q}(i)$ is the set of node i's neighbors in the graph sequence at time step k, and $\mathbf{e}_{k,ji}^{p,q}$ is the edge feature vector of edge ji at time step k. Note that the input to the first layer is the node feature vector, i.e., $\mathbf{z}_{k,i}^{p,q,0} = \mathbf{x}_{k,i}^{p,q}$. To incorporate biologically meaningful spatial priors into the model, we encode structure- and function-informed edge attributes by leveraging the hemispheric identity and subnetwork membership of the nodes connected by each edge. Specifically, for every edge ji, we obtain the hemisphere and subnetwork labels of both endpoint nodes, embed these categorical labels into learnable vectors, and concatenate them to form the final edge attribute.

After encoding the structural information of static graphs within the graph sequence $\mathcal{D}_k^{p,q} = \left(G^{p,q}_{k-\ell_k^{p,q}}, \ldots, G_k^{p,q}\right)$, we obtain a sequence of embeddings for each node in the graph, i.e., $(\mathbf{z}^{p,q,L}_{k-\ell^{p,q}_k,i}, \ldots, \mathbf{z}^{p,q,L}_{k,i})$ for time steps $k - \ell_k^{p,q}, \ldots, k$. Inspired by recent advances in state-space models for sequence modeling (Gu and Dao, 2023; Dao and Gu, 2024), we employ the `Mamba2` model to encode the temporal evolution of structural embeddings into spatiotemporal embeddings.

$$\mathbf{h}_{k,i}^{p,q} = \texttt{Mamba2}(\mathbf{z}^{\{p,q\},L}_{k-\ell_k^{p,q},i}, \ldots, \mathbf{z}^{\{p,q\},L}_{k,i}). \tag{5}$$

The spatiotemporal embeddings $\mathbf{h}^{p,q}k, i \in \mathbb{R}^{D\text{emb}}$ are then fed into two separate linear layers to compute the mean and covariance of each node:

$$\begin{aligned} \mu_{k,i}^{p,q} &= \texttt{Linear}_{\mu}(\mathbf{h}_{k,i}^{p,q}), \\ \sigma_{k,i}^{p,q} &= \mathrm{ELU}\big(\texttt{Linear}_{\sigma}(\mathbf{h}_{k,i}^{p,q})\big) + 1 + \sigma. \end{aligned} \tag{6}$$

The mean and variance vectors $\boldsymbol{\mu}_{k,i}^{p,q}$ and $\boldsymbol{\sigma}_{k,i}^{p,q}$ are in $\mathbb{R}^{D_{\text{emb}}}$, characterizing the Gaussian distribution of node i's embedding in $\mathcal{D}_k^{p,q}$ as $\mathcal{N}(\boldsymbol{\mu}_{k,i}^{p,q}, \mathrm{diag}(\boldsymbol{\sigma}_{k,i}^{p,q}))$. We add 1 to the ELU output to ensure covariance positivity; σ is a small constant (10^{-14}) for numerical stability.

We train the BrainATCL by optimizing a contrastive triplet loss. For each reference node i of the static graph $G_k^{p,q}$ in the graph sequence $\mathcal{D}_k^{p,q}$, we extract its direct neighbors as positive samples i^{near} and the nodes two hops away from the reference node as negative samples i^{far}. Reference nodes and their corresponding positive and negative samples form a set of triplets $\mathcal{T}_k^{p,q} = (i, i^{\text{near}}, i^{\text{far}}) \mid i \in G_k^{p,q}$ used in the contrastive representation learning.

The contrastive loss function aims to minimize the distance between the embeddings of adjacent nodes while maximizing the distance between the embeddings of non-adjacent nodes. We define the contrastive loss for the graph sequence sample $\mathcal{D}_k^{p,q}$ as

$$\mathcal{L}_{\mathcal{D}_k^{p,q}} = \frac{1}{\ell_k^{p,q}+1} \sum_{G_k^{p,q} \in \mathcal{D}_k^{p,q}} \sum_{\mathcal{T}_k^{p,q}} [E^2(i, i^{\text{near}}) + e^{-E(i,i^{\text{far}})}], \tag{7}$$

where the energy function E computes the Kullback-Leibler divergence between the corresponding Gaussian distribution representations of the input nodes, i.e.,

$$E(i,j) = \text{KL}(\mathcal{N}(\mu_i, \Sigma_i) || \mathcal{N}(\mu_j, \Sigma_j)). \tag{8}$$

The loss objective of the BrainATCL is the sum of embedding losses over all graph sequence samples: $\mathcal{L} = \sum_{\mathcal{D}} \mathcal{L}_{\mathcal{D}_k^{p,q}}$.

3.3.3. Evaluation on downstream prediction tasks.

Once the pre-trained BrainATCL is optimized using contrastive learning, we embed the dFC graph and use the embeddings for downstream tasks:

For the link prediction task, we represent each edge as the concatenation of the mean vectors of its corresponding endpoint node embeddings and feed this representation to a multilayer perceptron (MLP), i.e., $\hat{\mathbf{A}}_{k+1,ij}^{p,q} = \texttt{MLP}([\mu_{k,i}^{p,q} || \mu_{k,j}^{p,q}])$. Here, $\hat{\mathbf{A}}_{k+1,ij}^{p,q}$ represents the probability of observing the edge (i,j) at time step $k+1$ for subject p and session q, given the sequence of graphs $(G_{k-\ell_k^{p,q}}^{p,q}, \ldots, G_k^{p,q})$. The `MLP` is trained using binary cross-entropy between predictions and the ground-truth edges at the next time step.

For age prediction, we use the mean of each node's Gaussian distribution as its embedding and aggregate them into subject-level temporal representations using `MeanPooling` and a `Conv` module, then pass it through a `Regressor` layer to compute the predicted age, i.e., $\hat{y}^p = \texttt{Regressor}(\texttt{Conv}(\texttt{MeanPooling}(\{\mu_{k,i}^{p,q}\}_{i \in G_k^{p,q}})))$. (see Appendix C for details)

4. Experimental Results

4.1. Dataset

We evaluated our approach on the HCP young adult dataset for two tasks: temporal link prediction and age prediction. In the link prediction task, we select the first 300 subjects to train the BrainATCL and conduct the link prediction task. The trained model is then used to generate dynamic functional connectivity (dFC) graph embeddings for a separate set of 300 subjects (not used in training the BrainATCL), which are subsequently used for the age prediction task.

4.2. Implementation Details

We train BrainATCL under two settings: **1) Within-session:** each of the four sessions per subject is split by timestamps: the first 70% of each session is used for training, the next 10% for validation, and the remaining 20% for testing. The training segments from all sessions and all subjects are concatenated to form the final training set; validation and test segments are concatenated similarly. **2) Cross-session:** the first two sessions are used for training, the third for validation, and the fourth for testing. Given the temporal gap between the four sessions, we exclude the first $\ell_{\max}$ graphs at the beginning of each session when computing the adaptive lookback, preventing the model from referencing graphs across session boundaries and ensuring semantic consistency. See Appendix C for more implementation details.

4.3. Evaluation Tasks and Metrics

We assess the performance of BrainATCL on temporal link prediction task using five standard metrics – mean average precision (MAP), mean reciprocal rank (MRR), area under the ROC curve (AUC), area under the precision–recall curve (PR-AUC), and F1-score. See Appendix B for detailed definitions. For age estimation task, the evaluation metrics are mean absolute error (MAE) and mean square error (MSE), both measured in years. We adopted five regression models for age prediction: *ElasticNet, Ridge, Lasso, Random Forest, and Support Vector Regression (SVR)*. For each model, we perform grid search to identify the optimal hyperparameters. The model with the best performance is selected based on validation results.

4.4. Results

We conduct an ablation study under two training setups, systematically varying key hyperparameters such as lookback window, threshold, embedding size, and stride length. As shown in Appendix D Fig. 4, both the mean and variance of MAP and MRR scores are reported for each configuration. Based on this comparison, we select the configuration with threshold 0.5, embedding size 128, adaptive lookback, and stride size 10 for the final model comparison. This setup provides a strong balance between within-session and cross-session performance, yielding high MAP and MRR scores across both evaluation scenarios. For the choice of temporal window size, (Allen et al., 2014; Hutchison et al., 2013) recommend windows of approximately $60s$ to ensure reliable estimation of time-varying correlations. Accordingly, we use a window size of $w = 100$ ($72s$), which provides a principled balance between estimation reliability and sensitivity to transient resting-state fluctuations.

We also evaluate the impact of incorporating edge features into the BrainATCL model. We compare two variants: with and without edge features, and summarize the results for both within-session and cross-session evaluations in Table 1. Including edge features consistently improves performance across all five evaluation metrics, indicating that added relational information helps capture fine-grained variations in dynamic connectivity patterns.

We further conducted an ablation study to evaluate the impact of different training objectives. In particular, we replace the contrastive loss with a VGAE-style reconstruction objective using inner-product decoding and KL regularization (Kipf and Welling, 2016).

Table 1: Ablation study on the effect of edge features (including network and hemisphere labels of the two endpoint nodes) for temporal link prediction under within-session and cross-session settings.

Within-session					
Model Variant	MAP↑	MRR↑	AUC↑	PR-AUC↑	F1↑
w/o Edge Features	0.7313 ± 0.0024	0.6061 ± 0.0015	0.9728 ± 0.0004	0.7143 ± 0.0016	0.6830 ± 0.0011
w/ Edge Features	**0.7681 ± 0.0012**	**0.6265 ± 0.0009**	**0.9746 ± 0.0001**	**0.7468 ± 0.0013**	**0.7050 ± 0.0008**
Cross-session					
Model Variant	MAP↑	MRR↑	AUC↑	PR-AUC↑	F1↑
w/o Edge Features	0.7257 ± 0.0016	0.6051 ± 0.0009	0.9705 ± 0.0004	0.7061 ± 0.0008	0.6815 ± 0.0013
w/ Edge Features	**0.7667 ± 0.0004**	**0.6194 ± 0.0006**	**0.9741 ± 0.0001**	**0.7352 ± 0.0016**	**0.6923 ± 0.0017**

As shown in Table 2, the model trained with the variational loss consistently underperforms the proposed contrastive learning framework across all evaluation metrics. This indicates that purely generative reconstruction objectives are insufficient to capture discriminative temporal patterns in dynamic brain networks.

Table 2: Ablation study on training objectives by replacing the proposed contrastive loss with a VGAE-style reconstruction objective for temporal link prediction under the within-session setting.

Within-session					
Model Variant	MAP↑	MRR↑	AUC↑	PR-AUC↑	F1↑
VGAE Reconstruction	0.7054 ± 0.0019	0.5982 ± 0.0021	0.9701 ± 0.0004	0.6752 ± 0.0012	0.6488 ± 0.0011
Contrastive (BrainATCL)	**0.7681 ± 0.0012**	**0.6265 ± 0.0009**	**0.9746 ± 0.0001**	**0.7468 ± 0.0013**	**0.7050 ± 0.0008**

Under this optimal configuration, we evaluate BrainATCL against a comprehensive set of baselines (classical and recent graph models) on temporal link and age prediction. "Within-session" and "cross-session" refer to the two training setups. As shown in Table 3, BrainATCL achieves the best performance across all evaluation metrics, with MAP of 0.7681, MRR of 0.6265, AUC of 0.9746, PR-AUC of 0.7468, and F1-score of 0.7050, significantly outperforming all other baselines. The baselines include general graph neural networks - GCN (Kipf, 2016) GAT (Hamilton et al., 2017), GraphSAGE (Veličković et al., 2017), GIN (Xu et al., 2018)), GNN-RNNs variants (GCN-GRU, GAT-GRU, GIN-GRU, GraphSAGE-GRU) (Seo et al., 2018; Chen et al., 2022), a dynamic graph variational method (VGRNN) (Hajiramezanali et al., 2019), and two state-of-the-art temporal graph models: TransformerG2G (Varghese et al., 2024) and $\mathcal{GDG}$-Mamba (Pandey et al., 2024). Notably, while $\mathcal{GDG}$-Mamba performs competitively, BrainATCL achieves better performance through the integration of biologically meaningful edge attributes and a novelty-aware adaptive lookback mechanism.

Regarding explainability, since our cohort consists of healthy young adults with a relatively narrow age range (22–37 years), ROI- and edge-level attributions are expected to exhibit a low signal-to-noise ratio and limited translational relevance. To provide a preliminary analysis, we selected the two subjects with the largest age difference (22 and 37 years) and computed the cosine distance between their learned ROI embeddings. The resulting

Table 3: Benchmarking BrainATCL against multiple baselines for temporal link and age prediction under within-session and cross-session settings. MAP, MRR, AUCROC, PR-AUC, and F1 are used for link prediction, while MAE and MSE are used for age prediction. Reported values are means with standard deviations shown in parentheses. All comparisons are conducted under the same experimental configuration (stride=10, window=100, embedding=128, threshold=0.5).

Within-session							
Model	MAP↑	MRR↑	AUC↑	PR-AUC↑	F1↑	MAE↓	MSE↓
GCN	0.2838 (0.0018)	0.2226 (0.0009)	0.9101 (0.0000)	0.2679 (0.0001)	0.3557 (0.0001)	3.09 (0.16)	13.77 (1.89)
GraphSAGE	0.3245 (0.0234)	0.2576 (0.0534)	0.9279 (0.0000)	0.3017 (0.0001)	0.4054 (0.0002)	3.29 (0.58)	16.58 (5.42)
GAT	0.3601 (0.0027)	0.2045 (0.0038)	0.9112 (0.0000)	0.3408 (0.0001)	0.4233 (0.0002)	3.16 (0.49)	15.78 (2.45)
GIN	0.3564 (0.0124)	0.2308 (0.0210)	0.9178 (0.0000)	0.3326 (0.0001)	0.4150 (0.0000)	3.15 (0.07)	14.70 (1.23)
GCN-GRU	0.2991 (0.0055)	0.2351 (0.0037)	0.9172 (0.0000)	0.2785 (0.0000)	0.3622 (0.0000)	3.08 (0.54)	14.68 (0.77)
GraphSAGE-GRU	0.4161 (0.0041)	0.3752 (0.0061)	0.9484 (0.0000)	0.4043 (0.0001)	0.4961 (0.0000)	3.25 (0.41)	16.85 (3.93)
GAT-GRU	0.3765 (0.0023)	0.2256 (0.0167)	0.9183 (0.0000)	0.3489 (0.0001)	0.4176 (0.0000)	3.06 (0.06)	13.49 (1.23)
GIN-GRU	0.3727 (0.0061)	0.2446 (0.0036)	0.9304 (0.0000)	0.3356 (0.0001)	0.4363 (0.0000)	2.92 (0.37)	**10.77** (2.15)
VGRNN	0.3987 (0.0047)	0.2523 (0.0021)	0.9112 (0.0000)	0.3742 (0.0001)	0.4399 (0.0001)	3.32 (0.36)	17.01 (0.59)
TransformerG2G	0.6615 (0.0063)	0.5675 (0.0036)	0.9730 (0.0002)	0.6509 (0.0006)	0.6786 (0.0014)	2.87 (0.41)	12.18 (2.86)
$\mathcal{GDG}$-Mamba	0.7201 (0.0034)	0.5951 (0.0018)	0.9711 (0.0001)	0.6978 (0.0007)	0.6786 (0.0012)	2.84 (0.48)	12.09 (2.65)
BrainATCL	**0.7681** (0.0012)	**0.6265** (0.0009)	**0.9746** (0.0001)	**0.7468** (0.0013)	**0.7050** (0.0008)	**2.81** (0.37)	11.73 (2.57)

Cross-session							
Model	MAP↑	MRR↑	AUC↑	PR-AUC↑	F1↑	MAE↓	MSE↓
GCN	0.2767 (0.0002)	0.2262 (0.0005)	0.9113 (0.0000)	0.2652 (0.0000)	0.3528 (0.0000)	3.46 (0.01)	16.38 (0.79)
GraphSAGE	0.3127 (0.0473)	0.2892 (0.0674)	0.9166 (0.0000)	0.2536 (0.0001)	0.3581 (0.0001)	3.64 (0.59)	17.17 (4.22)
GAT	0.3512 (0.0018)	0.1825 (0.0007)	0.9120 (0.0000)	0.3393 (0.0001)	0.4206 (0.0000)	3.27 (0.001)	15.79 (0.25)
GIN	0.3423 (0.0067)	0.2217 (0.1298)	0.9215 (0.0000)	0.3612 (0.0001)	0.4370 (0.0000)	3.04 (0.74)	14.65 (3.75)
GCN-GRU	0.2801 (0.0035)	0.2285 (0.0016)	0.9186 (0.0000)	0.2778 (0.0000)	0.3613 (0.0000)	3.41 (0.21)	15.53 (0.67)
GraphSAGE-GRU	0.4173 (0.0020)	0.3798 (0.0054)	0.9496 (0.0000)	0.4055 (0.0001)	0.4974 (0.0001)	3.44 (0.31)	17.64 (4.41)
GAT-GRU	0.3689 (0.0025)	0.2058 (0.0314)	0.9203 (0.0000)	0.3507 (0.0001)	0.4190 (0.0000)	3.28 (0.22)	15.98 (2.01)
GIN-GRU	0.3813 (0.0059)	0.2328 (0.0097)	0.9321 (0.0000)	0.3618 (0.0000)	0.4372 (0.0000)	2.97 (0.52)	**11.62** (2.89)
VGRNN	0.3838 (0.0034)	0.2489 (0.0078)	0.9105 (0.0000)	0.3755 (0.0000)	0.4423 (0.0000)	3.58 (0.11)	17.39 (0.76)
TransformerG2G	0.6643 (0.0050)	0.5689 (0.0031)	0.9714 (0.0000)	0.6589 (0.0010)	0.6799 (0.0007)	3.02 (0.45)	13.44 (3.09)
$\mathcal{GDG}$-Mamba	0.7002 (0.0029)	0.5898 (0.0013)	0.9674 (0.0001)	0.6700 (0.0016)	0.6603 (0.0008)	2.83 (0.38)	12.04 (3.04)
BrainATCL	**0.7667** (0.0004)	**0.6194** (0.0003)	**0.9741** (0.0001)	**0.7352** (0.0016)	**0.6923** (0.0017)	**2.80** (0.38)	11.68 (2.97)

cross-subject differences were further grouped according to major functional networks, as shown in Appendix E Fig. 5. As illustrated in the figure, inter-subject variability is not uniformly distributed across networks, but is mainly concentrated in the visual, dorsal attention, and control systems, whereas somatomotor, limbic, and default-mode networks exhibit substantially smaller differences. These observations are consistent with prior resting-state functional connectivity studies reporting network-specific effects of aging, with pronounced changes in task-positive and visual networks (Mowinckel et al., 2012). For more details, see Appendix E. Taken together, our node-level representation analysis suggests that cross-subject differences captured by BrainATCL are structured and network-dependent, reflecting biologically plausible patterns of age-related functional reorganization, even within a relatively narrow age range.

5. Conclusion

Dynamic functional connectivity analysis faces challenges from fixed context windows and the limited integration of neuroscience-informed priors. To address these issues, we propose **BrainATCL**, an unsupervised and adaptive framework that dynamically adjusts temporal context based on temporal novelty index and incorporates biologically meaningful edge attributes to enhance spatial modeling. Experiments on HCP data demonstrate its strong performance and generalizability across both link prediction and age estimation tasks. One limitation of the current framework is the lack of end-to-end training between the embedding model and downstream tasks. In future work, we plan to enhance the adaptivity of the framework by aligning temporal context selection more directly with brain activity dynamics, and to explore the identification of individual-specific brain modes during the resting state.

Acknowledgments

This work is supported in part by DOE SEA-CROGS project (DESC0023191), AFOSR project(FA9550-24-1-0231), and the Grace Hopper AI Research Institute (GHAIRI) Seed Grant at NJIT. We also thank the computing resources provided by the High Performance Computing (HPC) facility at NJIT. Christine Ahrends is supported by the Carlsberg Foundation (CF23-1716). We would like to acknowledge Usama Pervaiz for his contribution to data curation. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

C Ahrends, Angus Stevner, Usama Pervaiz, Morten L Kringelbach, Peter Vuust, Mark W Woolrich, and Diego Vidaurre. Data and model considerations for estimating time-varying functional connectivity in fmri. *NeuroImage*, 252:119026, 2022.

Christine Ahrends and Diego Vidaurre. Dynamic functional connectivity. In *fMRI Techniques and Protocols*, pages 333–375. Springer, 2025.

Christine Ahrends, Mark W Woolrich, and Diego Vidaurre. Predicting individual traits from models of brain dynamics accurately and reliably using the fisher kernel. *eLife*, 13: RP95125, 2025.

Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–676, 2014.

Giuseppe de Alteriis, Oliver Sherwood, Alessandro Ciaramella, Robert Leech, Joana Cabral, Federico E Turkheimer, and Paul Expert. Dysco: A general framework for dynamic functional connectivity. *PLoS Computational Biology*, 21(3):e1012795, 2025.

Tiago Azevedo, Alexander Campbell, Rafael Romero-Garcia, Luca Passamonti, Richard AI Bethlehem, Pietro Lio, and Nicola Toschi. A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional mri data. *Medical Image Analysis*, 79:102471, 2022.

Cole Baker, Isabel Suárez-Méndez, Grace Smith, Elisabeth B Marsh, Michael Funke, John C Mosher, Fernando Maestú, Mengjia Xu, and Dimitrios Pantazis. Hyperbolic graph embedding of meg brain networks to study brain alterations in individuals with subjective cognitive decline. *IEEE journal of biomedical and health informatics*, 28(12):7357–7368, 2024.

Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

Bharat B Biswal and Lucina Q Uddin. The history and future of resting-state functional magnetic resonance imaging. *Nature*, 641(8065):1121–1131, 2025.

Aleksandar Bojchevski and Stephan Günnemann. Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. *arXiv preprint arXiv:1707.03815*, 2017.

Peng Cao, Guangqi Wen, Xiaoli Liu, Jinzhu Yang, and Osmar R Zaiane. Modeling the dynamic brain network representation for autism spectrum disorder diagnosis. *Medical & Biological Engineering & Computing*, 60(7):1897–1913, 2022.

Jinyin Chen, Xueke Wang, and Xuanheng Xu. Gc-lstm: Graph convolution embedded lstm for dynamic network link prediction. *Applied Intelligence*, 52(7):7513–7528, 2022.

Tri Dao and Albert Gu. Transformers are ssms: Generalized models and efficient algorithms through structured state space duality. *arXiv preprint arXiv:2405.21060*, 2024.

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 *International conference on medical image computing and computer-assisted intervention*, pages 528–538. Springer, 2020.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. *Advances in neural information processing systems*, 32, 2019.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.

R Matthew Hutchison, Thilo Womelsdorf, Joseph S Gati, Stefan Everling, and Ravi S Menon. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Human brain mapping*, 34(9):2154–2177, 2013.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer, 2022. URL https://arxiv.org/abs/2210.06681.

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.

Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Timothy O Laumann, Abraham Z Snyder, and Caterina Gratton. Challenges in the measurement and interpretation of dynamic functional connectivity. *Imaging Neuroscience*, 2:1–19, 2024.

Nora Leonardi, Jonas Richiardi, Markus Gschwind, Samanta Simioni, Jean-Marie Annoni, Myriam Schluep, Patrik Vuilleumier, and Dimitri Van De Ville. Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937–950, 2013.

Athanasia M Mowinckel, Thomas Espeseth, and Lars T Westlye. Network-specific effects of age and in-scanner subject motion: a resting-state fmri study of 238 healthy adults. *Neuroimage*, 63(3):1364–1373, 2012.

Ashish Parmanand Pandey, Alan John Varghese, Sarang Patil, and Mengjia Xu. A comparative study on dynamic graph embedding based on mamba and transformers. *arXiv preprint arXiv:2412.11293*, 2024.

Ashish Parmanand Pandey, Alan John Varghese, Sarang Patil, and Mengjia Xu. A comparative study on dynamic graph embedding based on mamba and transformers. *arXiv e-prints*, pages arXiv–2412, 2024.

Matej Perovnik, Tomaž Rus, Katharina A Schindlbeck, and David Eidelberg. Functional brain networks in the evaluation of patients with neurodegenerative disorders. *Nature Reviews Neurology*, 19(2):73–90, 2023.

Farimah Poursafaei, Shenyang Huang, Kellin Pelrine, and Reihaneh Rabbany. Towards better evaluation for dynamic link prediction. *Advances in Neural Information Processing Systems*, 35:32928–32941, 2022.

Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4): 665–678, 2011.

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.

Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *International conference on neural information processing*, pages 362–373. Springer, 2018.

Helen H Shen. Resting-state connectivity. *Proceedings of the National Academy of Sciences*, 112(46):14115–14116, 2015.

Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013a.

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, 2013b.

Arthur PC Spencer and Marc Goodfellow. Using deep clustering to improve fmri dynamic functional connectivity analysis. *NeuroImage*, 257:119288, 2022.

Jun Tang, Jie Chen, Miaojun Hu, Yao Hu, Zixi Zhang, and Liuming Xiao. Diagnosis of autism spectrum disorder (asd) by dynamic functional connectivity using gnn-lstm. *Sensors*, 25(1), 2025. ISSN 1424-8220. doi: 10.3390/s25010156. URL https://www.mdpi.com/1424-8220/25/1/156.

Md Asadullah Turja, Martin Styner, and Guorong Wu. Deepgraphdmd: Interpretable spatio-temporal decomposition of non-linear functional brain network dynamics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 358–368. Springer, 2023.

Harm J van der Horn, Josef M Ling, Tracey V Wick, Andrew B Dodd, Cidney R Robertson-Benta, Jessica R McQuaid, Vadim Zotev, Andrei A Vakhtin, Sephira G Ryman, Joana Cabral, et al. Dynamic functional connectivity in pediatric mild traumatic brain injury. *Neuroimage*, 285:120470, 2024.

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.

Alan John Varghese, Aniruddha Bora, Mengjia Xu, and George Em Karniadakis. Transformerg2g: Adaptive time-stepping for learning temporal graph embeddings using transformers. *Neural Networks*, 172:106086, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Jingjie Wang, Jinwei Lang, Li-Zhuang Yang, and Hai Li. Learning dynamic brain network representation based on graph mamba architecture. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 5143–5150. IEEE, 2024a.

Jiyao Wang, Nicha C Dvornek, Peiyu Duan, Lawrence H Staib, Pamela Ventola, and James S Duncan. Stnagnn: Data-driven spatio-temporal brain connectivity beyond fc. *arXiv preprint arXiv:2406.12065*, 2024b.

Tao Wang, Zenghui Ding, Zheng Chang, Xianjun Yang, Yanyan Chen, Meng Li, Shu Xu, and Yu Wang. A novel graph neural network framework for resting-state functional mri spatiotemporal dynamics analysis. *Physica A: Statistical Mechanics and its Applications*, 669:130582, 2025. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2025.130582. URL https://www.sciencedirect.com/science/article/pii/S0378437125002341.

Yuxiang Wei, Anees Abrol, and Vince D Calhoun. Hierarchical spatio-temporal state-space modeling for fmri analysis. In *International Conference on Research in Computational Molecular Biology*, pages 86–98. Springer, 2025.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Mengjia Xu, Zhijiang Wang, Haifeng Zhang, Dimitrios Pantazis, Huali Wang, and Quanzheng Li. A new graph gaussian embedding method for analyzing the effects of cognitive training. *PLoS computational biology*, 16(9):e1008186, 2020.

Mengjia Xu, David Lopez Sanz, Pilar Garces, Fernando Maestu, Quanzheng Li, and Dimitrios Pantazis. A graph gaussian embedding method for predicting alzheimer's disease progression with meg brain networks. *IEEE Transactions on Biomedical Engineering*, 68 (5):1579–1588, 2021.

Mengjia Xu, Apoorva Vikram Singh, and George Em Karniadakis. Dyng2g: An efficient stochastic graph embedding method for temporal graphs. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):985–998, 2022.

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

Hao Zhang, Ran Song, Liping Wang, Lin Zhang, Dawei Wang, Cong Wang, and Wei Zhang. Classification of brain disorders in rs-fmri via local-to-global graph neural networks. *IEEE transactions on medical imaging*, 42(2):444–455, 2022.

Zakariae Zrimek, Youssef Mourchid, and Mohammed El Hassouni. Dynstg-mamba: Dynamic spatio-temporal graph mamba with cross-graph knowledge distillation for gait disorders recognition, 2025. URL https://arxiv.org/abs/2503.13156.

Appendix A. Impact of Threshold τ on Temporal Graph Dynamics

To better understand how the choice of threshold τ influences the dynamic properties of the constructed temporal graphs, we analyze the edge evolution across time using the Temporal Edge Appearance (TEA) plot (Poursafaei et al., 2022). The TEA plot quantifies, at each time point, the proportion of newly appearing edges compared to the previous graph snapshot, thereby offering a visual measure of temporal variability. Fig. 3 shows how the choice of threshold τ affects the temporal dynamics of functional connectivity graphs. When τ is too small, the resulting graphs are overly dense and smooth, capturing many weak or noisy connections. This leads to a large number of repeated edges across time and consequently a lower average novelty score. In contrast, when τ is too large, each graph contains very few edges, making the connectivity patterns highly sparse and more dynamic, but potentially less stable. Thresholds around 0.4 and 0.5 provide a good balance, capturing meaningful changes while avoiding excessive noise. We further validate this observation in our ablation study, which shows that performance is sensitive to τ, and based on these results, we select $\tau = 0.5$ for our final model.

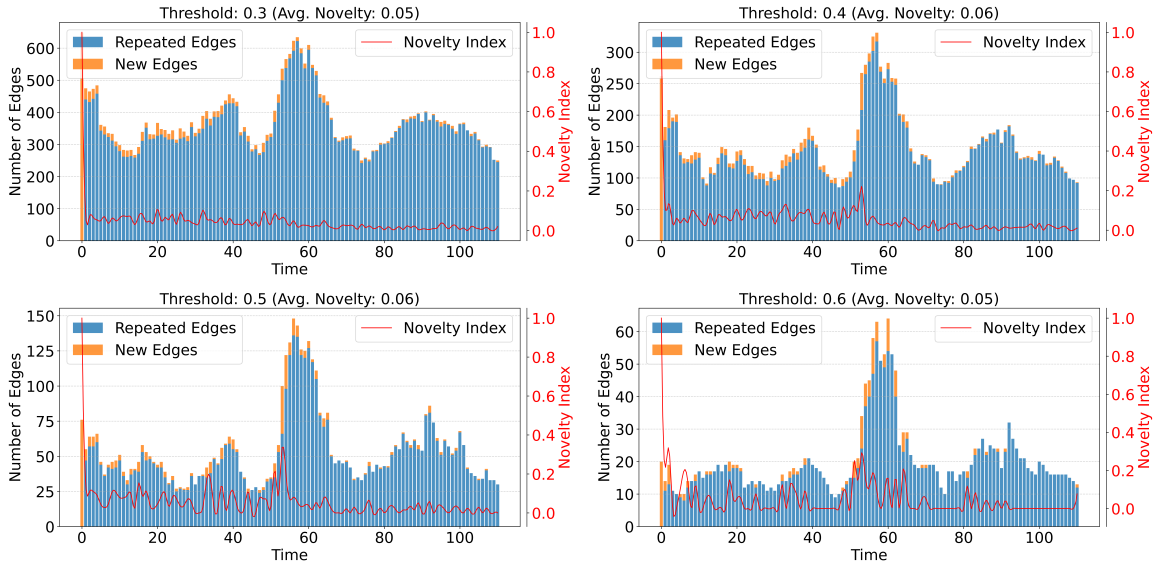

Figure 3: Temporal Edge Appearance (TEA) plots for subject 10, session 1 under different threshold values. For each time point, the blue bars indicate the number of repeated edges (i.e., edges that have appeared in any previous time point), while the orange bars represent newly appeared edges. The red line shows the novelty index at each time point, quantifying the proportion of new edges relative to total edges. The average novelty across the time series is reported in the title of each subplot.

Appendix B. Evaluation Metrics

(1) Mean Average Precision (MAP): For each query node q, predictions are ranked in descending order by their predicted probabilities. The average precision (AP) is calculated using Eq. 1, where $P(k)$ is the precision at rank k, and $rel(k)$ is 1 if the k-th item is relevant. The final MAP is the mean of AP values across all nodes in Q.

$$\text{MAP} = \frac{1}{|Q|}\sum_{q\in Q} AP(q); \quad AP(q) = \frac{1}{m}\sum_{k=1}^{n}[P(k)\times rel(k)] \tag{1}$$

(2) Mean Reciprocal Rank (MRR): MRR is the average of reciprocal ranks of the first relevant prediction for each query node q, as shown in Eq. 2. Unlike MAP, which considers all relevant items, MRR focuses on the rank of the first correct prediction.

$$\text{MRR} = \frac{1}{|Q|}\sum_{q\in Q}\frac{1}{k_q} \tag{2}$$

For F1-score, predicted link probabilities are binarized using an optimal threshold selected on the validation set. Specifically, for each experimental setting, the threshold that maximizes the validation F1-score is chosen and then applied to the corresponding test set. AUC and PR-AUC are computed in a threshold-free manner based on ranked prediction scores.

Appendix C. Implementation Details

We use 5 NVIDIA A100 GPUs with 8GB of memory each. Our model is implemented using the mamba-ssm library (Dao and Gu, 2024). The Mamba2 temporal module is configured with hidden dimension $d_{\text{model}} = 96$, state dimension $d_{\text{state}} = 64$, convolution kernel size $d_{\text{conv}} = 4$, and head dimension set to 16. Our models were trained for 50 epochs using the AdamW optimizer with a weight decay of 8×10^{-5}. The learning rate was adjusted based on the generated graph sequences and varied between 10^{-6} and 10^{-7} across different experimental settings. To regularize the model during training, a dropout rate of 0.38 was applied after the Mamba block. To prevent overfitting, we also employed early stopping with a patience of 5 epochs.

For the baseline model TransformerG2G, the input sequence of graph embeddings with adaptive lookback is first projected to a d-dimensional space using a linear layer, where we set $d = 256$. The model consists of 8 encoder layers, each with a single attention head. A tanh activation is applied in the nonlinear layer following the encoder, and the resulting representation is further projected to the D_{emb}-dimensional space, consistent with our BrainATCL architecture. In the nonlinear projection head, we apply an ELU activation function. The TransformerG2G model is optimized using the Adam optimizer with a fixed learning rate of 10^{-6} across all datasets. For the $\mathcal{GDG}$-Mamba baseline, we adopt the same training setup as used for our proposed BrainATCL model. For all adaptive lookback settings, we set $\ell_{\min} = 1$ and $\ell_{\max} = 4$.

We also compared BrainATCL against several static and dynamic GNNs for link prediction and age estimation tasks. Static GNNs include GCN (Kipf, 2016), GraphSAGE (Hamilton et al., 2017), GAT (Veličković et al., 2017), and GIN (Xu et al., 2018). Dynamic GNNs comprise compositions of static GNNs with gated recurrent units, GCN-GRU,

GraphSAGE-GRU, GAT-GRU, and GIN-GRU (Chen et al., 2022; Seo et al., 2018), as well as the variational graph recurrent neural network (VGRNN) (Hajiramezanali et al., 2019).

Both static and dynamic GNNs were trained end-to-end for both tasks. For static GNNs, the input is a static graph $G_k^{p,q}$, and the target is either the next graph $G_{k+1}^{p,q}$ or the subject's age y^p. For dynamic GNNs, the input is a graph sequence of fixed lookback size four, $(G_{k-4}^{p,q}, \ldots, G_k^{p,q})$, with the same targets.

For the *temporal link prediction task*, we trained an MLP classifier consists of one hidden layer with ReLU activation, as defined in Section 3.3, to predict whether a pair of nodes is connected based on their embeddings. The classifier was trained to minimize the weighted binary cross-entropy loss using an Adam optimizer with a learning rate of 10^{-3} and evaluated using MAP and MRR. For this task, each subject's data was split into train/validation/test sets along the temporal dimension, following the same split setting used for training the BrainATCL. The reported MAP and MRR scores are averaged over all test graphs to evaluate embedding quality. For F1-score, the decision threshold is selected based on validation performance.

In the *age estimation task*, we ensured that subjects with familial relationships (e.g., twins) were assigned to the same dataset partition (training or testing) to avoid data leakage. This guarantees that related individuals do not appear in both sets simultaneously during model training and evaluation. We use a separate set of 300 subjects that are not involved in the link estimation experiments to avoid information leakage. The data is split at the subject level, with 80% of the subjects used for training and 20% for testing. All available sessions are included for each subject. Results are averaged over five runs, each with a different random 80/20 train-test split. At first, we generate a temporal embedding for each subject by concatenating the graph embeddings from all sessions. Before feeding these subject-level temporal embeddings into the age regression model, we apply a dimensionality reduction module to extract compact subject-level representations. Given an input tensor of shape $(P, K, N, D_{\text{emb}})$, where P is the number of subjects, K is the number of temporal graphs per subject, N is the number of brain regions, and D_{emb} is the embedding dimension, we first perform mean pooling over the region dimension to obtain a tensor of shape (P, K, D_{emb}). The resulting sequence is then passed through a 1D convolutional network consisting of two convolutional layers with ReLU activations and an adaptive average pooling layer to aggregate temporal information into a fixed-size vector. For the K-dimensional representation, since it combines four sessions, we also experimented with an alternative approach in which we first averaged the four sessions before applying the convolutional module. In addition, we tested removing the first 20% of graphs from each session, as the initial segment often contains substantial noise and fluctuations. We selected the best-performing configuration as the final result. Finally, a fully connected layer maps the output to a 256-dimensional representation, which is then used for downstream age prediction.

Appendix D. Ablation study on the effect of hyperparameters

See Fig. 4.

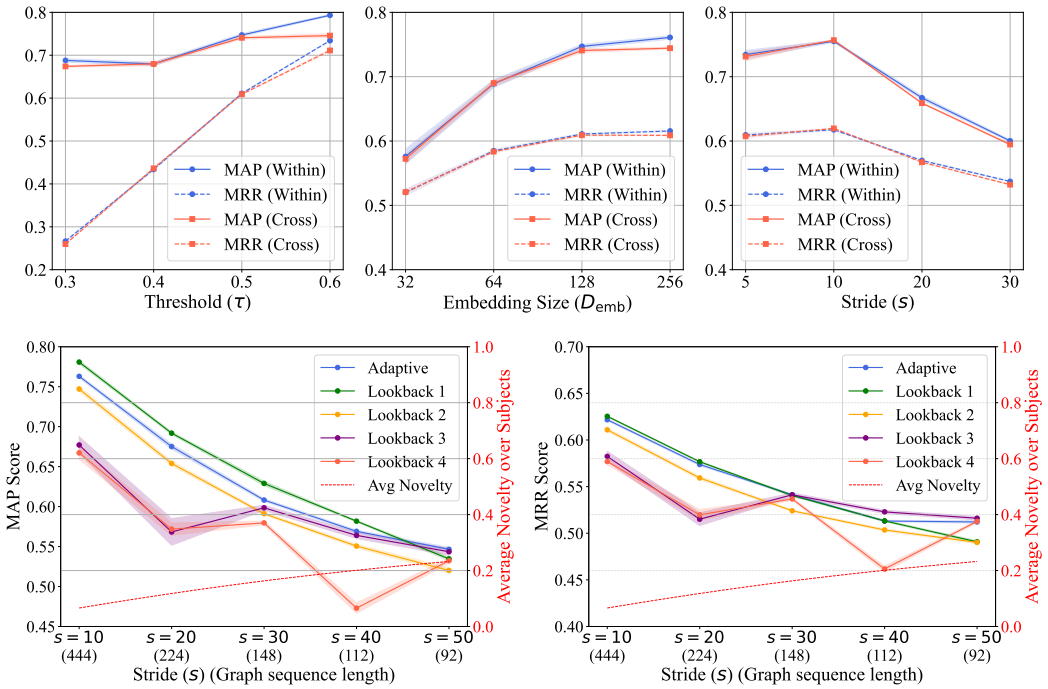

Figure 4: **The Top panel** shows ablation study on the effect of hyperparameters (threshold, embedding size and stride) on *within-session* (blue) and *cross-session* (red) temporal link prediction performance in terms of MAP (solid lines) and MRR (dashed lines). Solid/dashed lines denote the mean performance, while the shaded areas correspond to the standard deviation across five runs. **The bottom panel** shows MAP and MRR scores across different lookback strategies and stride sizes for the link prediction task. The red dashed line shows the average novelty score corresponding to each stride setting. The x-axis denotes the stride size (s), with the number of temporal graphs per subject indicated in parentheses.

Appendix E. Explainability

See Fig. 5. Each box summarizes the distribution of ROI-level cosine distances within a functional network. This visualization reflects the heterogeneity of cross-subject embedding differences across ROIs within each network.

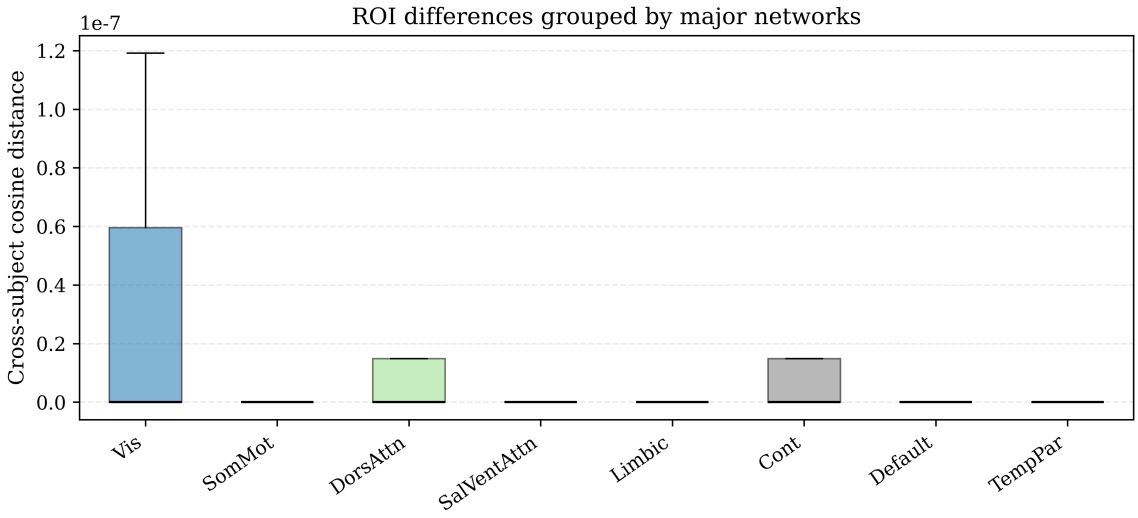

Figure 5: Distribution of cross-subject cosine distances between time-averaged ROI embeddings of the two subjects with the largest age difference (15 years), grouped by major functional networks. All subnetworks are defined based on the Schaefer parcellation (Schaefer et al., 2018)

Table 4: List of Symbols and Their Descriptions

Symbol	Description
N	Number of nodes
M	Number of edges
P	Number of subjects
Q	Number of sessions
T	Number of time steps in BOLD time series
$\mathcal{X}$	BOLD time series
w	Window size for computing the Pearson correlation coefficient
s	Stride size for computing the Pearson correlation coefficient
$\mathcal{G}$	Set of all dynamic functional connectivity graphs
$G^{p,q}$	Dynamic functional connectivity graph for subject p and session q
K	Number of static graphs in the dynamic functional connectivity
$G_k^{p,q}$	Static graph for subject p and session q at time step k
$\mathbf{X}_k^{p,q}$	Node feature matrix for subject p and session q at time step k
$\mathbf{E}_k^{p,q}$	Edge feature matrix for subject p and session q at time step k
$\mathbf{A}_k^{p,q}$	Adjacency matrix for subject p and session q at time step k
$\mathbf{y}^p$, y^p	Target variable vector and scalar for subject p
F_V	Number of node features
F_E	Number of edge features
f	Embedding function
$\mathbf{H}^{p,q}$	Node embedding matrix for subject p and session q
$\hat{\mathbf{A}}_K^{p,q}$	Predicted adjacency matrix for subject p and session q at time step K
g	Link prediction function
r	Age prediction function
$\mathbf{z}_{k,i}^{\{p,q\},l}$	l^{th} layer's spatial embedding vector of node i for subject p, session q and time index k
$\mathcal{N}_k^{p,q}(i)$	Set of neighbors of node i for subject p and session q at time step k
$\mathbf{e}_{k,ji}^{p,q}$	Feature vector of edge (j,i) for subject p and session q at time step k
L	Number of `GINEConv` layers
$\mathbf{h}_{k,i}^{p,q}$	Hidden state of Mamba2 for node i at time step k corresponding to subject p and session q
$\mu_{k,i}^{p,q}, \sigma_{k,i}^{p,q}$	Mean and standard deviation of Gaussian distribution corresponding to the embeddings of node i at time step k for subject p and session q
$D_{\mathbf{emb}}$	Embedding dimension size
$E(i,j)$	Energy function that computes KL divergence between embeddings of nodes i and j
$\mathcal{L}$	Loss function for BrainATCL's training
$\mathcal{L}_{\mathcal{D}_k^{p,q}}$	Embedding loss corresponding to the sample $\mathcal{D}_k^{p,q}$
$n_k^{p,q}$	Novelty index of the graph $G_k^{p,q}$
B	Number of discrete lookback values and number of quantile bins
$\alpha_b^{p,q}$	The b^{th} quantile of novelty index for subject p and session q
$\ell_k^{p,q}$	Adaptive lookback number for the graph $G_k^{p,q}$
$\mathcal{D}_k^{p,q}$	Graph sequence sample with adaptive lookback corresponding to subject p, session q and time step k