# ON THE LIMITS OF APPLYING GRAPH TRANSFORMERS FOR BRAIN CONNECTOME CLASSIFICATION

**Jose Lara-Rangel, Clare Heinbaugh** *
Department of Engineering
University of Cambridge
Cambridge, UK
{jml224,ceh210}@cam.ac.uk

## ABSTRACT

Brain connectomes offer detailed maps of neural connections within the brain. Recent studies have proposed novel connectome graph datasets and attempted to improve connectome classification by using graph deep learning. With recent advances demonstrating transformers' ability to model intricate relationships and outperform in various domains, this work explores their performance on the novel NeuroGraph benchmark datasets and synthetic variants derived from probabilistically removing edges to simulate noisy data. Our findings suggest that graph transformers offer no major advantage over traditional GNNs on this dataset. Furthermore, both traditional and transformer GNN models maintain accuracy even with all edges removed, suggesting that the dataset's graph structures may not significantly impact predictions. We propose further assessing NeuroGraph as a brain connectome benchmark, emphasizing the need for well-curated datasets and improved preprocessing strategies to obtain meaningful edge connections.

## 1 INTRODUCTION

The human brain is a complex network of interconnected regions. Neuroscientists often divide the brain into Regions of Interest (ROIs) and measure signaling between these regions called connectomes (Said et al., 2023). Functional magnetic resonance imaging (fMRI) can be used to observe signals between different regions. Like other areas of biology, the question of how to preprocess the data is often as crucial as the modelling step (Wang et al., 2021). Typically, mapping features signal measurements to graph structures requires domain-specific knowledge of neuroscience. To automate the workflow, Said et al. (2023) recently presented NeuroGraph as a collection of benchmark graph datasets derived from the Human Connectome Project (Essen et al., 2013), giving graph machine learning researchers access to brain connectome graph data.

After testing a variety of deep learning and graph machine learning models on NeuroGraph, Said et al. (2023) identified the ResidualGCN (or GNN*), which uses graph convolutions and residual connections, as the top performer. However, it exhibited overfitting and poor generalization (Said et al., 2023). Following the trend in deep learning to move towards transformers and attention-based models (Vaswani et al., 2017; Islam et al., 2024; Shehzad et al., 2024), we explore the application of Exphormer—a novel sparse Graph Transformer (GT)—along with other attention mechanisms for brain connectome gender, activity, and age prediction tasks, comparing their performance to previous models applied on the NeuroGraph benchmark (Shirzad et al., 2023; Said et al., 2023). Surprisingly, while attention was expected to improve performance and generalization (Hussain et al., 2022), Exphormer's performance matched that of ResidualGCN without surpassing it. Despite extensive regularization efforts, Exphormer also showed signs of overfitting and did not improve performance on the NeuroGraph datasets.

We decided to investigate other potential advantages of GTs and whether the comparable performance with previous models in Said et al. (2023), including graph and non-graph deep learning models, could be attributed to the models themselves or the graph structures in the NeuroGraph

*Equal contribution.

datasets. To this end, we generated synthetic datasets by removing edges, an approach that has been previously explored for other purposes (Han et al., 2025). We hypothesized that given Exphormer's ability to model local and global relationships, combined with its structural encodings, it would be more resilient to edge-dropping compared to ResidualGCN, which relies heavily on local neighborhoods (Shirzad et al., 2023; Hussain et al., 2022). Surprisingly, both models showed resilience, maintaining performance even with complete edge removal. This suggests that node relationships may not be critical for predictions, with models potentially relying solely on node features. Consequently, the graph structure in the NeuroGraph benchmark datasets require further evaluation.

Our contributions are:

- We explore the application of sparse GTs for brain connectome classification tasks, showing the potential of GTs in understanding complex neural connections.
- We generate synthetic datasets with missing edges and show that even with significant alterations to graph structures both ResidualGCN and Exphormer preserve performance compared to the orignal datasets.
- We propose further evaluation of the NeuroGraph static classification datasets as benchmarks, highlighting their potential simplicity and susceptibility to overfitting, and advocate for alternative preprocessing strategies to create more meaningful graph structures.

## BACKGROUND

### NEUROGRAPH DATASET

NeuroGraph was introduced in Said et al. (2023) as a comprehensive collection of static and dynamic brain connectome datasets tailored for graph machine learning classification and regression tasks. We focus on classification tasks on demographics and activity state static graphs, due to data accessibility, computational simplicity, and existing model performance comparisons. The demographic group includes gender classification as male and female, and age classification into 22-25, 26-30, and 31-35 years groups. The activity state task is to identify core functions relevant to brain-cognition-behavior relationships. This consists of seven distinct activities: emotional processing, gambling, language, motor, relational processing, social cognition, and working memory. Hence, the NeuroGraph datasets considered are HCP-Activity, HCP-Gender, and HCP-Age.

In a static graph representation, individual data points are represented by single graphs. In the static connectome graph, $\mathcal{G} = (V, E, \boldsymbol{X})$, the node set $V$ represents ROIs, and the edge set $E \subseteq V \times V$ the positive correlations between pairs of ROIs determined with a defined threshold, and considers a feature matrix $\boldsymbol{X} \in \mathbb{R}^{n \times d}$, where $n$ is the total number of ROIs and $d$ is the dimension of the feature vector. In NeuroGraph, the correlation vectors are used as node features (Said et al., 2023).

### SPARSE GRAPH TRANSFORMERS

GTs extend conventional transformers for graph representation learning (Sáez de Ocáriz Borde, 2024; Vaswani et al., 2017; Min et al., 2022; Dwivedi & Bresson, 2020). Unlike traditional Message-Passing Graph Neural Networks (MPGNNs), GTs operate on a fully connected graph to better model long-range dependencies and interactions among distant nodes. In contrast to local attention in Graph Attention Networks (GATs), GTs allow every node to communicate directly with all others, producing a global attention mechanism (Veličković et al., 2017). By incorporating the input graph's structure as a soft inductive bias and retaining the global attention mechanism of traditional transformers, GTs effectively leverage global graph structural information. Consequently, GTs mitigate common shortcomings observed in MPGNNs, such as over-smoothing, over-squashing, and limited expressivity (Sáez de Ocáriz Borde, 2024; Min et al., 2022).

However, in the dense GT setting, where the graph is fully connected, the computational complexity becomes $O(|V|^2)$ (Min et al., 2022; Dwivedi & Bresson, 2020). Real-world graph datasets, like the NeuroGraph case, often exhibit arbitrary connectivity structures shaped by the application domain, sometimes with an extremely large number of nodes. Although this diversity can offer valuable information, it also makes the use of GTs impractical. Sparse GTs address this issue by restricting attention to specific node pairs, reducing complexity while preserving critical structural information.

Exphormer is a sparse GT introduced in Shirzad et al. (2023) that integrates two sparse attention mechanisms: global nodes and expander graphs. Global nodes are connected to all other nodes, whereas the number of edges in the expander graph is proportional to the number of nodes. For an input graph $\mathcal{G} = (V, E)$, Exphormer builds an undirected interaction graph with three components: **expander graph attention** for long-range connections using $O(|V|)$ edges, **global attention** with virtual nodes linked to all others for universal representation, and **local neighborhood attention** linking immediate neighbors with $O(|E|)$ edges. Expander graphs act as sparse approximations of complete graphs, retaining the spectral properties of a full attention mechanism while requiring $O(|V|)$ edges. Unlike dense attention mechanisms, Exphormer's sparse attention mechanisms do not directly connect all pairs of nodes in the pairwise interaction graph, but by stacking transformer layers most nodes' pairwise interactions can be effectively modeled (Shirzad et al., 2023). Thus, Exphormer has an expander-based sparse attention mechanism with $O(|V| + |E|)$ computational complexity that achieved results comparable to a full transformer (Shirzad et al., 2023).

## METHODOLOGY

To evaluate Exphormer performance, we consider ResidualGCN as a baseline (Said et al., 2023). ResidualGCN's residual nature is derived from the concatenation of outputs from a series of graph convolutional layers in a sequential model, passed into a final multi-layer perceptron. We also explored the introduction of attention to this base architecture. However, it did not produce improvements in performance or other notable benefits, see Appendix 1.

For the Exphormer, we experimented with different numbers of layers, dropout rates for the network and attention mechanism, and numbers of attention heads. The final configuration used dropout probability of $0.1$, attention dropout of $0.3$, $2$ layers, and $4$ attention heads. All experiments used learning rate decay starting at $0.001$, decaying by $1e-5$, over a total of $100$ epochs with $5$ warmup epochs. We used three different seeds for both the Exphormer and ResidualGCN and assessed the alignment with the results in Said et al. (2023), which only included one run for each experiment.

Apart from evaluating performance, we investigated potential advantages of using attention-based models. Our hypothesis was that the attention mechanism could enhance robustness to data noise, particularly in scenarios where certain graph structure components, such as edges, are missing. To verify that the graph structure, nodes and edges taken together, convey meaningful information for prediction, it is important to compare models under noisy or incomplete data settings. We simulate noisy incomplete data by removing edges based on a pre-specified probability of edge removal.

## EXPERIMENTS & DISCUSSION

### PERFORMANCE ON NEUROGRAPH DATASETS

Table 1 shows the performance of both models for each dataset. During training, Exphormer achieves its best performance in the early stages of the training process, whereas ResidualGCN reaches its best performance towards the end of training, see Appendix 1. However, the sparse transformer mechanism of Exphormer does not provide clear performance improvements compared to ResidualGCN. Both models performed similarly across datasets, with ResidualGCN slightly outperforming Exphormer on the HCP-Activity dataset and solidly outperforming on the HCP-Gender dataset. Although the validation and test curves remained similar in most experiments, both models overfit the data. HCP-Age exhibits the most prominent performance drop of approximately 50% between train and test accuracy, see Appendix 1 for additional details.

Regularizing with dropout applied to both the network and attention mechanism, along with reducing layers, did not significantly improve performance. Fewer layers sped up training, higher network dropout slowed learning and reduced accuracy, while attention dropout stabilized learning. No configuration outperformed the baseline, see Appendix 1.

### DROPPED EDGES SYNTHETIC DATASET

Table 1 shows the results of the edge-dropping experiments. As expected, Exphormer maintained performance as the edge-drop probability increased. In contrast, and contrary to what we anticipated,

Table 1: Test accuracy for models trained on NeuroGraph datasets with varying probabilities of dropped edges. Average results over 3 seeds.

| Dataset | Drop edge probability | ResidualGCN | Exphormer |
|---|---|---|---|
| HCP-Activity | 0.0 | $98.10 \pm 0.20$ | $97.54 \pm 0.20$ |
| | 0.5 | $97.85 \pm 0.14$ | $97.40 \pm 0.69$ |
| | 1.0 | $97.76 \pm 0.08$ | $97.63 \pm 0.47$ |
| HCP-Gender | 0.0 | $88.27 \pm 1.16$ | $80.87 \pm 3.25$ |
| | 0.5 | $84.26 \pm 1.61$ | $79.94 \pm 1.41$ |
| | 1.0 | $86.58 \pm 2.02$ | $79.01 \pm 0.54$ |
| HCP-Age | 0.0 | $50.23 \pm 2.35$ | $48.91 \pm 4.21$ |
| | 0.5 | $50.39 \pm 0.71$ | $49.53 \pm 2.81$ |
| | 1.0 | $52.11 \pm 0.00$ | $47.35 \pm 5.15$ |

the performance of ResidualGCN also remained unchanged within error margins as edge-drop probabilities increased, even in the extreme case where all edges were removed.

Since removing edges had no substantial impact on the models' performance, and our prediction tasks focus on graph-level properties rather than node classification, it is possible that the node features in this dataset alone provide sufficient information to make predictions. Alternatively, the edges in the NeuroGraph datasets may carry irrelevant information that adds little to the prediction tasks, or the specific graph structure itself may not be critical for these tasks. This could also explain the lack of improvement with Exphormer, as some of its characteristics rely on edge features. Interestingly, ResidualGCN appears to stabilize as edges are removed while Exphormer fluctuates, especially considering the margins of error for the HCP-Age dataset in 1. While ResidualGCN and Exphormer capture local neighborhood attention, Exphormer also uses global attention, thus, future work could examine how global attention is better leveraged by datasets beyond NeuroGraph.

Having tested models with varying complexities that yielded similar results across the learning process, we can attribute this performance consistency without edges to factors beyond the network architecture's capacity, see Appendix 1. This could also explain why a vanilla neural network achieved competitive performance with the ResidualGCN in Said et al. (2023). This aligns with the recent work of Qiu et al. (2024) suggesting that pair-wise relationships (graph edges) may be insufficient in the application context of brain connectivity and warrants capturing higher-order relationships between ROIs. Consequently, further improvement would be expected from better pre-processing strategies rather than from applying attention-based models to current graph structures.

Unlike recent studies such as Han et al. (2025) that focused on model architectures, our experiments relate the lack of improvement to the graph structures in the NeuroGraph datasets. By experimenting with attention-based mechanisms, we circumvent the claim of Han et al. (2025) that complex graph theory models, like MPGNNs, exhibit poor performance compared to simpler models due to over-squashing. Instead, we hypothesize that the stagnant performance of attention-based graph deep learning models compared to traditional deep learning approaches could be explained by the lack of edge-conveyed information. Thus, neuroscience expertise is still necessary to pre-process brain connectomes to create meaningful graph benchmark datasets.

## CONCLUSION & FUTURE WORK

Contrary to expectations, applying sparse graph transformers like Exphormer, and other attention-based models, to static connectome classification tasks did not improve performance. Although Exphormer showed faster convergence, it performed worse compared to ResidualGCN, originally proposed in Said et al. (2023), and exhibited overfitting despite regularization efforts. In experiments simulating noisy conditions through random edge removal, both models' performance remained unaffected, even without edges. These findings highlight the need to further investigate the relevance of the graph structures in the NeuroGraph benchmark datasets and alternative fMRI processing pipelines to generate meaningful graph representations. This study highlights the importance of well-curated data in complex tasks like brain connectome classification.

## REFERENCES

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2020.

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

Keqi Han, Yao Su, Lifang He, Liang Zhan, Sergey Plis, Vince Calhoun, and Carl Yang. Rethinking functional brain connectome analysis: Do graph deep learning models help?, 2025. URL https://arxiv.org/abs/2501.17207.

Yanqiao Zhu Xuan Kan Antonio Aodong Chen Gu Joshua Lukemire Liang Zhan Lifang He Ying Guo Hejie Cui, Wei Dai and Carl Yang. Braingb: a benchmark for brain network analysis with graph neural networks. *IEEE Transactions on Medical Imaging*, pp. 80:62–79, 2022.

Md Shamim Hussain, Mohammed J. Zaki, and Dharmashankar Subramanian. Global self-attention as a replacement for graph convolution. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 655–665. Association for Computing Machinery, 2022. doi: 10.1145/3534678.3539296.

Saidul Islam, Hanae Elmekki, Ahmed Elsebai, Jamal Bentahar, Nagat Drawel, Gaith Rjoub, and Witold Pedrycz. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, 241, 2024. doi: https://doi.org/10.1016/j.eswa.2023.122666.

Erxue Min, Runfa Chen, Yatao Bian, Tingyang Xu, Kangfei Zhao, Wenbing Huang, Peilin Zhao, Junzhou Huang, Sophia Ananiadou, and Yu Rong. Transformer for graphs: An overview from architecture perspective. *arXiv preprint arXiv:2202.08455*, 2022.

Weikang Qiu, Huangrui Chu, Selena Wang, Haolan Zuo, Xiaoxiao Li, Yize Zhao, and Rex Ying. Learning high-order relationships of brain regions. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=WC14xZIaC2.

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

Anwar Said, Roza G. Bayrak, Tyler Derr, Mudassir Shabbir, Daniel Moyer, Catie Chang, and Xenofon D. Koutsoukos. Neurograph: Benchmarks for graph machine learning in brain connectomics. *ArXiv*, 2023.

Ahsan Shehzad, Feng Xia, Shagufta Abid, Ciyuan Peng, Shuo Yu, Dongyu Zhang, and Karin Verspoor. Graph transformers: A survey. *arXiv preprint arXiv:2407.09777*, 2024.

Hamed Shirzad, Ameya Velingker, Balaji Venkatachalam, Danica J. Sutherl, and Ali Kemal Sinop. Conditional neural processes. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. PMLR, 2023.

Haitz Sáez de Ocáriz Borde. Elucidating graph neural networks, transformers, and graph transformers, 02 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *6th International Conference on Learning Representations*, 2017.

Jiacheng Wang, Quan Zou, and Chen Lin. A comparison of deep learning-based pre-processing and clustering approaches for single-cell RNA sequencing data. *Briefings in Bioinformatics*, 23(1): bbab345, 09 2021. ISSN 1477-4054. doi: 10.1093/bib/bbab345. URL https://doi.org/10.1093/bib/bbab345.

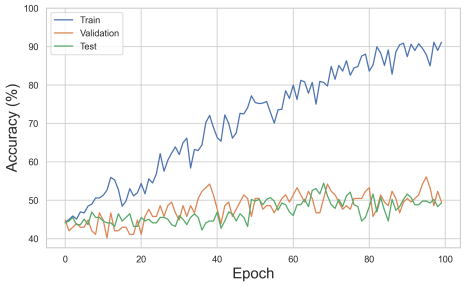
Nicha Dvornek Muhan Zhang Siyuan Gao Juntang Zhuang Dustin Scheinost Lawrence H Staib Pamela Ventola Xiaoxiao Li, Yuan Zhou and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 2021.

# APPENDIX. ADDITIONAL EXPERIMENTAL RESULTS

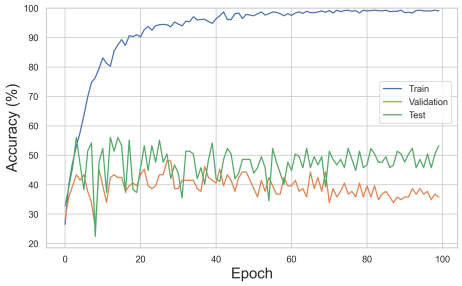
## ADDITIONAL EXPHORMER RESULTS

Figure 1 shows the learning curves for both the ResidualGCN and Exphormer on the HPC-Age datasets, illustrating that both models reach comparable accuracies but exhibit clear overfitting. Among the tasks and datasets considered, HCP-Age displays the largest drop between training and test accuracies, and the sparse GT properties of Exphormer (Shirzad et al., 2023) did not yield significant improvements over the baseline. However, it is worth noting that Exphormer converges to its optimal performance within fewer epochs, whereas ResidualGCN typically requires almost the entire training schedule to reach its best performance.

Similar trends are observed in other HPC datasets. Validation and test curves remain similar, with Exphormer converging faster than ResidualGCN, but even when the gap between training and testing is less pronounced, there remains a distinct sign of overfitting in most experiments, and neither model decisively outperforms the other.




(a) Accuracy vs. epoch number for ResidualGCN trained on the original HCP-Age dataset.




(b) Accuracy vs. epoch number for Exphormer trained on the original HCP-Age dataset.

Figure 1: Accuracy curves for both models on HPC-Age.

Trying to alleviate overfitting, we applied dropout to both the network layers and the Exphormer attention mechanism, varying dropout probabilities from 0.1 to 0.6. The best results were achieved with combinations of dropout probabilities of 0.1 and 0.3. The results for HCP-Gender and HCP-Activity are shown in Tables 2 and 3, respectively. Increasing the network dropout slowed learning and generally reduced accuracy, whereas moderate attention dropout helped stabilize the learning curves. Nevertheless, no configuration outperformed the baseline by a substantial margin on any of the datasets. This suggests that while dropout can help control overfitting, these particular tasks may require additional strategies to improve performance.

Table 2: Validation accuracy for HCP-Gender with different combinations of dropout.

| **Gender** | | **Attention Dropout** | |
|---|---|---|---|
| | | 0.1 | 0.3 |
| **Dropout** | 0.1 | 84.26 | 85.18 |
| | 0.3 | 80.56 | 82.41 |

Table 3: Test accuracy for Activity with different combinations of dropout.

| **Activity** | | **Attention Dropout** | | |
|---|---|---|---|---|
| | | 0.1 | 0.3 | 0.5 |
| **Dropout** | 0.1 | 96.24 | 96.94 | 96.91 |
| | 0.3 | 96.64 | 96.78 | 95.97 |

We also explored using different numbers of layers. As shown in Table 4, 2-layer and 3-layer configurations achieved similar performance, although the 2-layer model trained significantly faster. A 5-layer version was also tested but proved to be much more time-consuming without providing noticeable gains, so we discontinued further experiments.

Table 4: Validation and test accuracy for Exphormer experiments with 2 and 3 layers.

| **Layers** | **Gender** | | **Age** | | **Activity** | |
|---|---|---|---|---|---|---|
| | **Val** | **Test** | **Val** | **Test** | **Val** | **Test** |
| 2 layers | 84.26 | 77.78 | 54.72 | 48.60 | 97.45 | 97.36 |
| 3 layers | 84.26 | 79.63 | 52.83 | 49.53 | 97.71 | 96.78 |

### ADDITIONAL ATTENTION RESULTS

Exphormer builds upon GraphGPS (Rampášek et al., 2022), a modular framework that combines local message passing with a global attention mechanism and leverages various positional and structural encodings. This design provides a flexible platform for incorporating Transformer layers alongside MPNN models. As reported by Shirzad et al. (2023), some experiments showed improved Exphormer performance when additional GraphGPS layers were included. Specifically, by replacing the Transformer component in GraphGPS with an Exphormer layer, Exphormer achieved accuracies comparable to a full Transformer-based GraphGPS. Moreover, when Exphormer was combined with MPNNs in GraphGPS, it reached State-of-the-Art (SOTA) or near-SOTA results (Shirzad et al., 2023; Rampášek et al., 2022). Motivated by these findings, we tested the combination of GraphGPS and Exphormer on the NeuroGraph static datasets. However, as shown in Table 5 for the case of HCP-Gender, this hybrid setup did not yield improvements.

Furthermore, we extended our analysis to assess whether introducing attention layers in ResidualGCN could provide improvements in predictive accuracy. We compared two alternative attention mechanisms different to the one used in Said et al. (2023). The first one consists of an attention layer after each GCN layer, while the second version includes an attention layer after the concatenation of the GCN layers and the residual connections. Moreover, we extended the experiment to apply the attention mechanism according to a specified probability; with a probability of 1, it always applies attention. None of these modifications improved performance. Table 5 exemplifies the validation and test accuracies obtained on HCP-Gender for these alternatives. In some cases, the models with added attention matched or slightly exceeded the baseline accuracy but did not establish a clear advantage overall. Despite the different attention placements and probabilities, none of these modifications consistently improved performance.

Table 5: Validation and test accuracy for alternative models.

| **Model** | **Validation** | **Test** |
|---|---|---|
| GPS+Exphormer | 82.40 | 80.55 |
| ResidualGCN+Attention (1) | 87.25 | 88.25 |
| ResidualGCN+Attention (0.8) | 86.25 | 87.00 |
| ResidualGCN+Attention (0.3) | 84.75 | 88.25 |
| ResidualGCN+Attention2 (1) | 83.50 | 87.00 |
| ResidualGCN+Attention2 (0.6) | 79.50 | 85.00 |

### ADDITIONAL DETAILS ON NEUROGRAPH

There are only a few options for brain connectome datasets (Hejie Cui & Yang, 2022; Xiaoxiao Li & Duncan, 2021; Said et al., 2023; Qiu et al., 2024), which often require specialized processing to be used for graph machine learning. In this work, we explore NeuroGraph, a recent collection of static and dynamic brain connectome benchmark datasets tailored for graph machine learning classification and regression tasks introduced in Said et al. (2023). NeuroGraph was built using publicly

available datasets derived from the Human Connectome Project (HCP), applying minimal preprocessing steps based on standard fMRI procedures to create static and dynamic graph representations (Essen et al., 2013). In a static graph representation, individual data points are represented by single graphs. Conversely, dynamic graph representations encapsulate time-varying interactions and connectivity patterns within the brain. In our study, we focused on classification tasks on static graphs, due to dataset accessibility, computational simplicity, and existing model performance comparisons.

The static connectome graphs in NeuroGraph comprise node sets representing ROIs and edge sets representing positive correlations between pairs of ROIs. These benchmark datasets are categorized into three distinct groups: demographics, activity states, and cognitive traits. We focus on the first two groups, considering HCP-Activity, HCP-Gender, and HCP-Age. Dataset statistics are provided in Table 6.

Table 6: Dataset information for HCP-Activity, HCP-Gender, and HCP-Age (Said et al., 2023).

| **Dataset** | **# of graphs** | **# of node features** | **# of edge features** | **# of classes** |
|---|---|---|---|---|
| HCP-Activity | 7332 | 400 | 1 | 7 |
| HCP-Gender | 1078 | 1000 | 1 | 2 |
| HCP-Age | 1065 | 1000 | 1 | 3 |