Transformer-Based Hierarchical Clustering for Brain Network Analysis (Extended Abstract)

Wei Dai
Dept. of Computer Science
Stanford Univresity
Stanford, United States
dvd.ai@stanford.edu

Hejie Cui
Dept. of Computer Science
Emory University
Atlanta, United States
hejie.cui@emory.edu

Xuan Kan
Dept. of Computer Science
Emory University
Atlanta, United States
xuan.kan@emory.edu

Ying Guo
Dept. of Biostatistics and Bioinformatics
Emory University
Atlanta, United States
yguo2@emory.edu

Sanne van Rooij
Dept. of Psychiatry and Behavioral Sciences
Emory University
Atlanta, United States
sanne.van.rooij@emory.edu

Carl Yang
Dept. of Computer Science
Emory University
Atlanta, United States
j.carlyang@emory.edu

***Abstract*— Brain networks, graphical models such as those constructed from MRI, have been widely used in pathological prediction and analysis of brain functions. Within the complex brain system, differences in neuronal connection strengths parcellate the brain into various functional modules (network communities), which are critical for brain analysis. However, identifying such communities within the brain has been a nontrivial issue due to the complexity of neuronal interactions. In this work, we propose a novel interpretable transformer-based model for joint hierarchical cluster identification and brain network classification with three main contributions. First, we offer an end-to-end transformer-based approach to learning clustering assignments. Through pairwise attention, a clustering layer, BCluster, and a transformer encoder collaboratively learn a globally shared clustering assignment that is continuously tuned to downstream tasks. BCluster enhances the model's performance and reduces run time complexity while also providing clinical insights. Second, we propose a hierarchical structure for the clustering model, enabling the model to learn more abstract, higher-level cluster representations by combining lower-level modules. Each clustering layer is attached to a distinct readout module, which allows the model to utilize the cluster embeddings of every layer effectively. Last but not least, we redesign the attention mechanism of the transformer with stochastic noise, which enhances its cluster learning capability. We compare our model's performance with SOTA models and perform clustering analysis with the ground truth community labels. Extensive experimental results show that with the help of hierarchical clustering, the model achieves increased accuracy and reduced runtime complexity while providing plausible insight into the functional organization of brain regions.**

***Index Terms*—Brain Networks, Neural Imaging Analysis, Graph Neural Networks, Clustering, Machine Learning**

I. Introduction

Graph is a ubiquitous form of data as it captures multiple objects and their interactions simultaneously. It is widely used for representing complex systems of related entities [1]. Brain network is a special kind of graph constructed from MRI images. In brain networks, the anatomical areas named "Region of Interests" (ROIs) are represented as nodes, while connectivities between ROIs are represented as links. Partition atlas defines the set of ROIs in a particular brain network. In recent decades, abundant works have shown strong connections between linking imaging-based brain connectivity and demographic characteristics or mental disorders [2].

Both shallow models and deep models like graph neural networks (GNNs) [3] are researched in the area of brain network analysis. Shallow models exhibits inferior performance as compared to deep models [2], but GNNs also suffer from over-smoothing [4], which limits their ability to model long-distance interactions. Transformers, on the other hand, have recently emerged as a promising approach for various tasks [5], including predictions on graph data [6]. Graph based transformers utilizes pairwise attention across full graphs, unlike GNNs, which only propagate node embeddings locally. BrainTransformer [7] employs transformer on brain networks and demonstrates state-of-the-art performance for brain network analysis.

ROIs in Brain networks are inherently hierarchically clustered [8]. In typical brain network analysis [9], clustered ROIs form communities, with each representing a particular functional module. These functional modules are then further organized into larger functional modules, with each responsible for a more general function. This arrangement creates a hierarchical "module-in-module" structure [10]. Functional modules provide critical information with regard to downstream tasks, and alterations in community patterns sometimes signal pathological lesions [11]. Therefore, learning a globally shared cluster assignment with the awareness of downstream tasks is beneficial for both model optimization and clinical examinations. However, learning such hierarchical cluster representations is difficult. Shallow methods proposed to detect brain communities are mostly based on the Louvain algorithm [12], and Lloyd algorithm [13]. These correlation-based methods fail to capture higher-order connectivity patterns between brain regions [14]. Some GNN models are proposed to detect

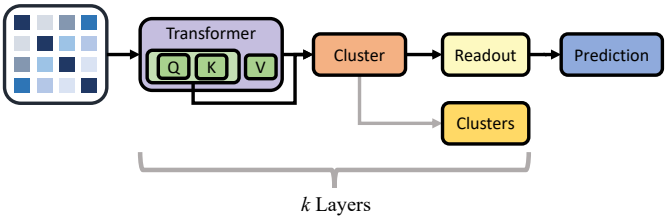

Fig. 1: The overall framework of our proposed model THC.

communities within the brain [3], but these models suffer from the over-smoothing problem of GNN, limiting their ability to aggregate and identify multi-hop connectivity patterns.

To solve the aforementioned challenges, we propose a Transformer-based Hierarchical Clustering model, abbreviated as THC, that is tailored for brain network analysis. An overview of our model is included in Figure 1. We highlight three main contributions of our clustering model.

First, we provide a novel clustering learning approach based on an end-to-end transformer-based clustering model. Particularly, a clustering layer takes an attention matrix from the previous transformer encoder. It multiplies the attention with a learned weighted assignment matrix. A Softmax function is applied so that the model outputs a probabilistic assignment from input embedding to clusters. This soft assignment is differentiable, which allows the model to optimize the assignment directly through gradient descent. The main objective of the clustering layer is maximizing the mutual information of the embedding before and after the cluster operation. This objective encourages the model to cluster similar nodes as the information loss of combining nodes with similar representations is also minimal. The clustering layer enhances the model's performance and reduces run time complexity while also providing clinical insights. During the training process, the assignment is shared batch-wise and optimized jointly with the model. After training is complete, a final globally shared assignment is obtained by averaging across the assignments of all samples.

Second, our model clusters the cluster embedding from the previous layer into new, larger clusters, generating a hierarchical structure. A readout layer is attached to each clustering layer. This allows the model to utilize cluster embeddings from different layers effectively. Moreover, cluster assignments from different layers can be combined to form a tree-like cluster structure, which provides more insights into how different brain regions interact with each other. In addition, the hierarchical assignment is squashable. A hierarchical clustering assignment produced by our model can be flattened into one-layer clustering with a linear number of matrix multiplications.

Last but not least, we redesign the attention mechanism of the transformer with stochastic noise. This allows the clustering layer to effectively learn the clustering assignment without falling into the trivial situation of replicating the results from the attention matrix. We compare our model's performance with SOTA models and perform clustering analysis with the ground truth community labels [15]. Empirical analysis demonstrates the superior prediction power of our model, and the assignment produced by THC aligns well with the ground-truth functional module labels.

To investigate the quality of the model's clustering assignment, we compare the clustering results of our model with ground truth labels. The results reveal that the proposed clustering method is able to produce clustering assignments similar to the existent labels for all functional modules. To quantitatively analyze the result, we further compare our method with other popular clustering methods. We observe that the proposed model produces clusters with the highest quality among both deep and shallow methods. These results further support that the proposed clustering THC can capture the complex structure of brain networks.

References

[1] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec, "Open graph benchmark: Datasets for machine learning on graphs," *NeurIPS*, vol. 33, pp. 22 118–22 133, 2020.

[2] H. Cui, W. Dai, Y. Zhu, X. Kan, A. A. Chen Gu, J. Lukemire, L. Zhan, L. He, Y. Guo, and C. Yang, "BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks," *IEEE TMI*, 2022.

[3] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, vol. 74, p. 102233, 2021.

[4] W. L. Hamilton, "Graph representation learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 14, pp. 1–159, 2020.

[5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT 2019*, 2019, pp. 4171–4186.

[6] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?" *NeurIPS*, 2021.

[7] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," in *NeurIPS*, 2022.

[8] T. J. Akiki and C. G. Abdallah, "Determining the hierarchical architecture of the human brain using subject-level clustering of functional networks," *Scientific Reports*, vol. 9, pp. 1–15, 2019.

[9] T. D. Satterthwaite, D. H. Wolf, D. R. Roalf, K. Ruparel, G. Erus, S. Vandekar, E. D. Gennatas, M. A. Elliott, A. Smith, H. Hakonarson, R. Verma, C. Davatzikos, R. E. Gur, and R. C. Gur, "Linked Sex Differences in Cognition and Functional Connectivity in Youth," *Cerebral Cortex*, vol. 25, pp. 2383–2394, 2015.

[10] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. T. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in neuroinformatics*, vol. 3, p. 37, 2009.

[11] A. Alexander-Bloch, R. Lambiotte, B. Roberts, J. Giedd, N. Gogtay, and E. Bullmore, "The discovery of population differences in network community structure: new methods and applications to brain functional networks in schizophrenia," *Neuroimage*, vol. 59, 2012.

[12] O. Sporns and R. F. Betzel, "Modular brain networks," *Annual review of psychology*, vol. 67, p. 613, 2016.

[13] L. Nanetti, L. Cerliani, V. Gazzola, R. Renken, and C. Keysers, "Group analyses of connectivity-based cortical parcellation using repeated k-means clustering," *Neuroimage*, vol. 47, pp. 1666–1677, 2009.

[14] V. A. Traag, L. Waltman, and N. J. Van Eck, "From louvain to leiden: guaranteeing well-connected communities," *Nature*, vol. 9, 2019.

[15] T. J. Akiki and C. G. Abdallah, "Determining the hierarchical architecture of the human brain using subject-level clustering of functional networks," *Nature*, vol. 9, p. 19290, 2019.