Comparing Shallow and Deep Graph Models for Brain Network Analysis

Erica Choi[1] Sally Smith[2] Ethan Young[3]

[1]Columbia University

[2]Georgia Institute of Technology

[3]University of California - Los Angeles

IEEE Big Data 2022

Introduction

- Analyze different approaches for classifying brain networks
  - kernelized SVM[1]
  - message passing GNNs[2]
  - graph kernel GNNs[3]
- Suggest several methods to motivate further research in brain network analysis

[1] Hofmann, Schölkopf, and Smola, “Kernel methods in machine learning”, 2008

[2] Cui et al., *BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks*, 2022

[3] Feng et al., “KerGNNs: Interpretable Graph Neural Networks with Graph Kernels”, 2022

Classification Task

- The standard graph classification task considers the problem of classifying graphs into two or more categories
- In this project, we perform binary classification on neuroimaging data to distinguish between negative and positive diagnoses

Datasets

- We are working with 2 datasets, one classifying HIV and the other classifying bipolar disorder
- Each dataset consists of:
  - diffusion tensor imaging (DTI) scans
  - functional magnetic resonance imaging (fMRI) scans
  - classification labels: positive diagnosis, negative diagnosis

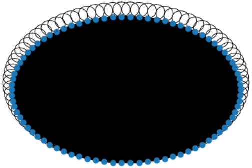
Datasets

- The DTI and fMRI brain scans of each patient i are represented as weighted adjacency matrices $\mathbf{W}_i \in \mathbb{R}^{M \times M}$
  - The fMRI scans are considered to be more robust than DTI scans, so our experiments prioritize working with them
  - The fMRI datasets have been cleaned for us and consist of 70 (HIV) and 97 (bipolar disorder) patients
- Nodes in the brain network represent regions of interest (ROIs), and edge links between nodes indicate the strength of the connection between ROIs

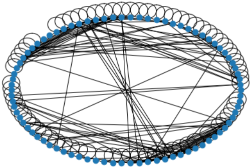
Threshold Rounding

- We implement a rounding scheme to remove edge weights and sparsify the adjacency matrices
- We have: $A_{ij}= \begin{cases} 1 & \text{if } \mathrm{A}_{ij} \geq \alpha \\ 0 & \text{otherwise} \end{cases}$, where A_{ij} is the ij-th entry of the adjacency matrix A and $\alpha \in [0, 1]$ is our rounding threshold

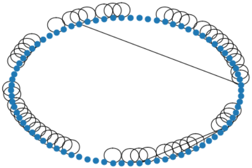
Threshold Rounding



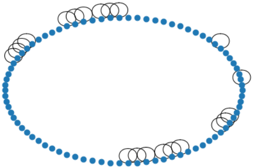
(a) Threshold: 0



(b) Threshold: 0.01



(c) Threshold: 0.1



(d) Threshold: 0.25

Figure 1: Effect of threshold rounding on network density.

Graph Kernels

- Popular in graph-based learning because they can be computed implicitly (inner product)
- We compute graph kernel matrices using the *GraKel* Python package and plug them into SVM to perform classification
- Consider WL, WLOA, shortest path, and graphlet sampling kernels in experiments

Graph Kernels

- **Weisfeiler-Lehman subtree** kernel is built on the Weisfeiler-Lehman graph isomorphism test[4] and is essentially a relabeling procedure
  - Computationally inexpensive, taking $O(hm)$ time, where h is the number of iterations and m is the number of edges.
- **WL optimal assignment** kernel uses valid assignment theory to improve the performance of the WL subtree kernel[5]
  - Computed in linear time, taking $O(|X| + |Y|)$ time, where X and Y are elements of $[\mathcal{X}]^n$. $[\mathcal{X}]^n$ denotes the set of all n-element subsets of the set $\mathcal{X}$.

[4] Weisfeiler and Lehman, "The reduction of a graph to canonical form and the algebra which appears therein", 1968

[5] Kriege, Giscard, and Wilson, "On Valid Optimal Assignment Kernels and Applications to Graph Classification", 2016

Graph Kernels

- **Shortest path** kernel decomposes graphs into shortest paths and compares pairs of them[6]
  - Computationally expensive when number of n nodes is large, taking $O(n^4)$ time
- **Graphlet sampling** kernel decomposes graphs into graphlets of k nodes and compares the number of matching graphlets between two graphs[7]
  - Computationally intractable for large k, taking $O(n^k)$ time
  - Experiments show k=5 generally performs the best

[6] Borgwardt and Kriegel, "Shortest-path kernels on graphs", 2005
[7] Przulj, "Biological network comparison using graphlet degree distribution", 2007

Support Vector Machines

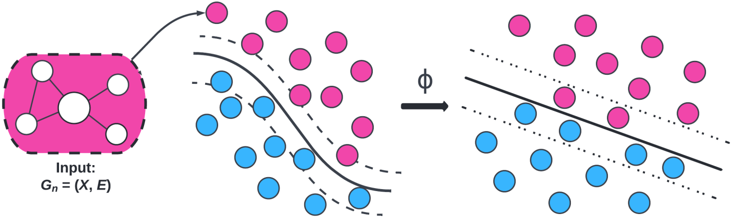


Figure 2: Overview of kernel SVM.

Graph Neural Networks

- GNNs combine node features and graph structures to perform prediction tasks
- General framework:
  - computing the representation of each node
  - applying a pooling strategy to obtain the graph representation
  - multilayer perceptron (MLP) can be applied to make predictions

BrainGB

- We implement MPGNNs using the *BrainGB* Python package and focus on two types of MPGNNs:
  - **Graph attention network** (GAT) is a type of convolutional neural network that operates on graphs
  - **Graph convolutional network** (GCN) is a special case of GATs with attention fully determined by graph structure alone, without node features
- Conduct experiments using settings based on extensive studies from Cui et al (2022)

BrainGB

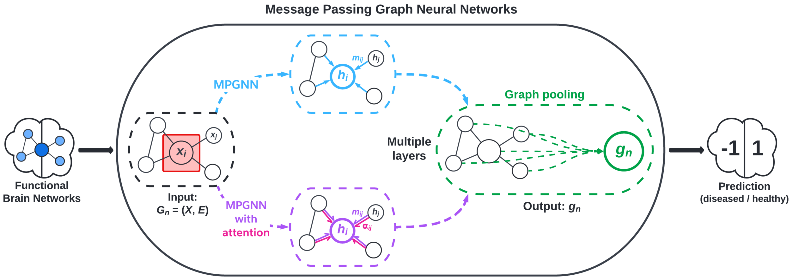
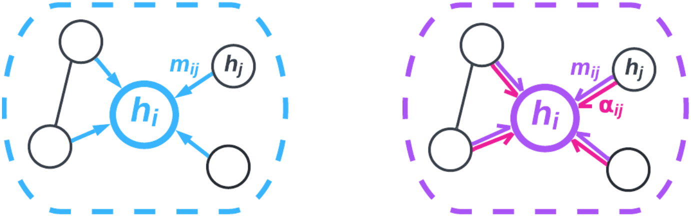


Figure 3: BrainGB framework. Adapted from Fig. 1 in Cui et al (2022). The node representation of node x_i is h_i, the message from node x_j to x_i is m_{ij}, and the attention weight from node x_j to x_i is a_{ij}.

BrainGB



(a) Standard message passing (b) MP with attention

Figure 4: The message passing schemes in the BrainGB framework.

BrainGB

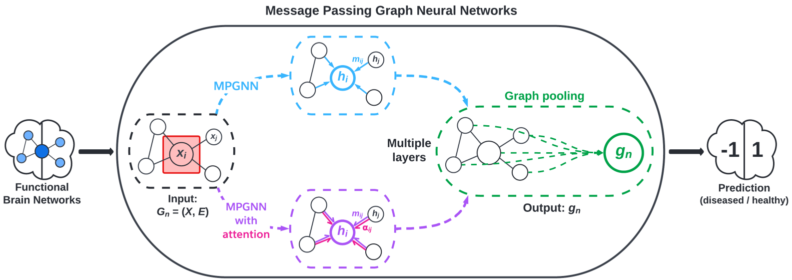


Figure 5: BrainGB framework. Adapted from Fig. 1 in Cui et al (2022). The output g_n is the pooled information that will be passed through a MLP to make the prediction.

Kernel GNNs

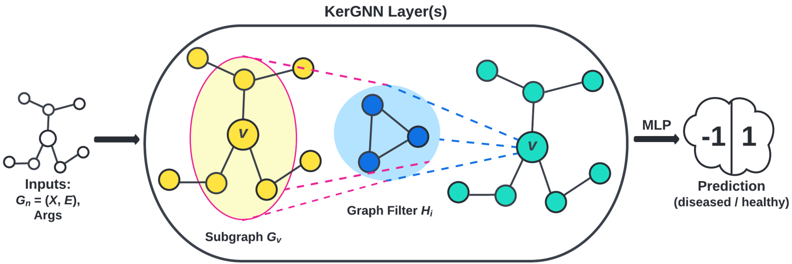


Figure 6: KerGNN framework. Adapted from Fig. 3 in Feng et al (2022).

Kernel GNNs

Number of epochs	100; 150; 200; 250; 300; 350; 400; 450; 500
Learning rate	10^{-2}; 10^{-3}; 10^{-4}; 10^{-5}; 10^{-6}
Dropout rate	0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9
Nodes in graph filter	2; 4; 6; 8; 10; 12; 14; 16; 18; 20
Subgraph size	5; 10; 15; 20
k-hop neighborhood	1; 2; 3
Max step of RW	1; 2; 3; 4; 5

Table 1: Hyperparameter Search Range

Data	Method	Accuracy	F1	AUC
HIV	WL-0.21	$0.67_{\pm 0.17}$	—	—
	WLOA-0.21	$0.65_{\pm 0.17}$	—	—
	SP-0.01	$0.66_{\pm 0.20}$	—	—
	GS-0.03	$0.66_{\pm 0.18}$	—	—
	GCN-concat	$0.64_{\pm 0.15}$	$0.59_{\pm 0.20}$	$0.77_{\pm 0.20}$
	GAT-concat	$0.73_{\pm 0.16}$	$0.71_{\pm 0.17}$	$0.81_{\pm 0.19}$
	GCN-edge concat	$0.71_{\pm 0.11}$	$0.69_{\pm 0.12}$	$0.77_{\pm 0.17}$
	GAT-edge concat	$0.69_{\pm 0.18}$	$0.67_{\pm 0.19}$	$0.73_{\pm 0.24}$
	KerGNN	$0.64_{\pm 0.19}$	—	—
BP	WL-0.4	$0.63_{\pm 0.19}$	—	—
	WLOA-0.42	$0.66_{\pm 0.12}$	—	—
	SP-0.02	$0.64_{\pm 0.12}$	—	—
	GS-0.04	$0.62_{\pm 0.15}$	—	—
	GCN-concat	$0.53_{\pm 0.13}$	$0.51_{\pm 0.14}$	$0.54_{\pm 0.16}$
	GAT-concat	$0.53_{\pm 0.13}$	$0.50_{\pm 0.13}$	$0.57_{\pm 0.19}$
	GCN-edge concat	$0.63_{\pm 0.12}$	$0.61_{\pm 0.13}$	$0.61_{\pm 0.17}$
	GAT-edge concat	$0.52_{\pm 0.17}$	$0.51_{\pm 0.16}$	$0.59_{\pm 0.19}$
	KerGNN	$0.68_{\pm 0.16}$	—	—

Discussion

- Limited data (70 and 97 patients in each dataset)
- GNNs are usually shallow; deep GNNs are still an active area of research
- For brain networks, what kinds of graph structures are effective beyond the pairwise connections are still unknown

Discussion

- Cui et al (2021)[8] notes HIV affects 2 sub-networks, while bipolar disorder only affects 1 sub-network
  - This may make accurate classification difficult
- Li et al (2020)[9] found utilizing multimodal neuroimaging (fMRI and MRI) improves SVM classification performance

[8] Cui et al., "BrainNNExplainer: an interpretable graph neural network framework for brain network based disease analysis", 2021

[9] Li et al., "Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques", 2020

Future Work

- There are many graph kernels and GNNs that we hope are useful in the area of brain network analysis
- Some of these include: graph kernel neural networks[10] (GKNN), graph stochastic attention[11] (GSAT), k-dimensional GNNs[12] (k-GNN), message passing graph kernels[13] (MPGK), and motif convolutional networks[14] (MCN)

---

[10] Cosmo et al., *Graph Kernel Neural Networks*, 2021

[11] Miao, Liu, and Li, *Interpretable and Generalizable Graph Learning via Stochastic Attention Mechanism*, 2022

[12] Morris et al., *Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks*, 2018

[13] Nikolentzos and Vazirgiannis, *Message Passing Graph Kernels*, 2018

[14] Lee et al., *Higher-order Graph Convolutional Networks*, 2018

Acknowledgements & Contact Info

- The authors thank Dr. Carl Yang, Ms. Hejie Cui and Mr. Dai Wei for their mentorship and technical assistance
- This work was financially supported in part by the US NSF awards DMS-2051019 and DMS-1751636; this work was also partially supported by the internal funds and GPU servers provided by the Department of Computer Science of Emory University

- If you have any questions or would like to discuss further, feel free to contact us

  - Erica Choi - erica.c@columbia.edu
  - Sally Smith - sallysmith@gatech.edu
  - Ethan Young - young.j.ethan@gmail.com

References I

Borgwardt, Karsten M. and Hans-Peter Kriegel. "Shortest-path kernels on graphs". In: *Fifth IEEE International Conference on Data Mining*. 2005, pp. 74–81.

Cosmo, Luca et al. *Graph Kernel Neural Networks*. 2021. DOI: 10.48550/ARXIV.2112.07436. URL: https://arxiv.org/abs/2112.07436.

Cui, Hejie et al. *BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks*. 2022. DOI: 10.48550/ARXIV.2204.07054. URL: https://arxiv.org/abs/2204.07054.

Cui, Hejie et al. "BrainNNExplainer: an interpretable graph neural network framework for brain network based disease analysis". In: *ICML-IMLH*. 2021.

Feng, Aosong et al. "KerGNNs: Interpretable Graph Neural Networks with Graph Kernels". In: *AAAI*. 2022.

References II

Hofmann, Thomas, Bernhard Schölkopf, and Alexander J Smola. "Kernel methods in machine learning". In: *The annals of statistics* 36.3 (2008), pp. 1171–1220.

Kriege, Nils, P.-L Giscard, and Richard Wilson. "On Valid Optimal Assignment Kernels and Applications to Graph Classification". In: *NeurIPS*. 2016, pp. 1615–1623.

Lee, John Boaz et al. *Higher-order Graph Convolutional Networks*. 2018. DOI: 10.48550/ARXIV.1809.07697. URL: https://arxiv.org/abs/1809.07697.

Li, Hao et al. "Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques". In: *BMC Psychiatry*. 2020.

References III

- Miao, Siqi, Miaoyuan Liu, and Pan Li. *Interpretable and Generalizable Graph Learning via Stochastic Attention Mechanism*. 2022. DOI: 10.48550/ARXIV.2201.12987. URL: https://arxiv.org/abs/2201.12987.
- Morris, Christopher et al. *Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks*. 2018. DOI: 10.48550/ARXIV.1810.02244. URL: https://arxiv.org/abs/1810.02244.
- Nikolentzos, Giannis and Michalis Vazirgiannis. *Message Passing Graph Kernels*. 2018. DOI: 10.48550/ARXIV.1808.02510. URL: https://arxiv.org/abs/1808.02510.
- Przulj, Natasa. "Biological network comparison using graphlet degree distribution". In: *Bioinformatics* 23.2 (2007), pp. 177–183.

References IV

Weisfeiler, Boris and A.A. Lehman. "The reduction of a graph to canonical form and the algebra which appears therein". In: *NTI* 9 (1968), pp. 12–16.