Neurodegenerative Brain Network Classification via Adaptive Diffusion with Temporal Regularization

Hyuna Cho[1] Jaeyoon Sim[1] Guorong Wu[2] Won Hwa Kim[1]

[1]Pohang University of Science and Technology (POSTECH) [2]University of North Carolina at Chapel Hill

ICML International Conference On Machine Learning

Motivation

Task. Classification of degenerative (i.e., progressive) brain networks (i.e., graphs).

- Given C **chronological** labels y_c $(c = 1, ..., C)$, neurodegeneration progresses from a healthy group (y_1) to the most deteriorated state (y_C).

Brain network as a graph.

- Brain regions are nodes (i.e., regions of interest).
- (Structural/functional) Connections between ROIs become edges.

Motivation.

- How to adaptively analyze graphs with homophily and heterophily characteristics?
- How to capture sequential variations in progressive neurodegeneration?
- How to identify neuroscientifically significant regions to classify degenerative brain disease stages?

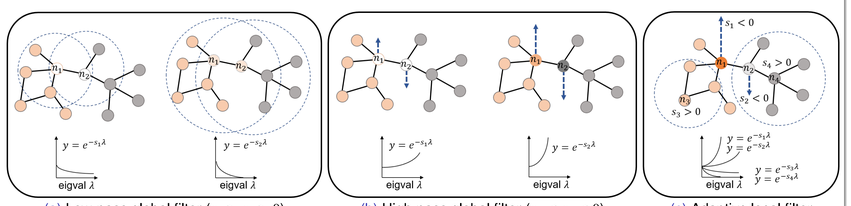
Spectral Graph Wavelet Transform

(Definition 1) Let $\mathbf{X} \in \mathbb{R}^{N\times F}$ be a node feature matrix with F features for N nodes, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ be a connectivity matrix.

(Definition 2) A graph Laplacian $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^T$ is decomposed with non-negative eigenvalues $\Lambda = diag(\lambda_1, ..., \lambda_N)$ and orthonormal eigenvectors $\mathbf{U} = [u_1, ..., u_N]$.

(Definition 3) A wavelet basis $\psi_s = \mathbf{U}k(s\Lambda)\mathbf{U}^T$ is defined with a kernel function $k(\cdot)$ and a scaling parameter s.

- Typically, a low-pass kernel is used for graph data, and the s is a hyperparameter.
- (Key Idea) In this work, we used a diffusion kernel $k(s\lambda) = e^{-s\lambda}$ with **trainable scaling parameters** $\mathbf{s} = \{s_i\}_{i=1}^N$ **paired with nodes** $\{n_i\}_{i=1}^N$.



(a) Low-pass global filter ($s_2 > s_1 > 0$) (b) High-pass global filter ($s_2 < s_1 < 0$) (c) Adaptive local filter

Figure: (a) A low-pass filter smooths out features among neighboring nodes, (b) while a high-pass filter accentuates the difference between nodes. (c) Unlike these filters, the adaptive local filter captures optimal bandwidths for each node with trainable node-wise scaling parameters.

- SGWT projects graph signals $\mathbf{X}$ onto the spectral domain as

$$W_{\mathbf{X}}(s) =< \psi_s, \mathbf{X} >= \psi_s\mathbf{X}, \tag{1}$$

which yields a wavelet coefficient $W_{\mathbf{X}}(s)$.

- Inverse transform (IGWT) completely reconstructs the signal $\mathbf{X}$ by projecting $W_{\mathbf{X}}(s)$ back to the graph domain as follows:

$$\mathbf{X} = \frac{1}{C_k}\int_0^\infty \psi_s \cdot W_{\mathbf{X}}(s)\frac{ds}{s} \tag{2}$$

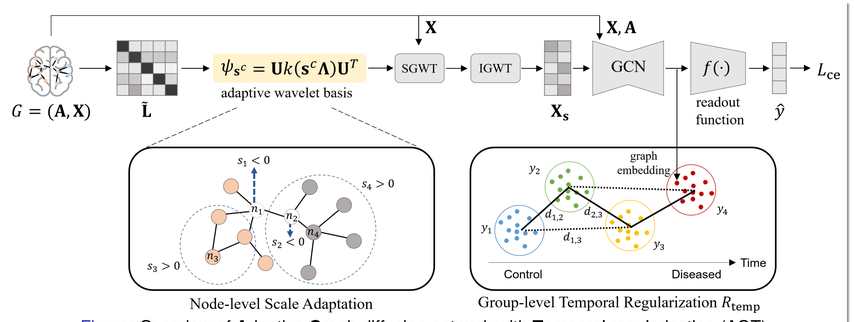
with an admissibility constant $C_k = \int_0^\infty \frac{k(\tau)^2}{\tau}d\tau < \infty$.

- The Eq. (2) is a superposition of *multi-resolution* representation of $\mathbf{X}$ over scales $s \in [0, \infty)$. Therefore, a signal $\mathbf{X}_s$ in the graph space filtered at the scale s is defined as

$$\mathbf{X}_s =< \psi_s, W_{\mathbf{X}}(s) >= \mathbf{U}k^2(s\Lambda)\mathbf{U}^T\mathbf{X}. \tag{3}$$

While the raw graph data $\mathbf{X}$ may contain unnecessary resolutions for solving a given task, Eq. (3) allows us **to extract task-relevant graph information at specific resolution(s)** in the spatial domain.

Model Architecture



Node-level Scale Adaptation Group-level Temporal Regularization R_{temp}

Figure: Overview of **A**daptive **G**raph diffusion network with **T**emporal regularization (AGT).

(1) Node-level Scale Adaptation

- $\{\mathbf{s}^c\}_{c=1}^C$: C number of trainable scale sets, where each set is a set of ROI-wise scales $\mathbf{s}^c = \{s_i^c\}_{i=1}^N$.
- The class- and node-wise adaptive wavelet basis $\psi_{\mathbf{s}^{c_i}}$ localized at n_i on n_j is defined as

$$\psi_{\mathbf{s}_{i,i}^c}(j) = \sum_{l=1}^{N} e^{-\mathbf{s}_i^c\lambda_l}u_l^*(i)u_l(j). \tag{4}$$

(2) Group-level Temporal Regularization

- Given $\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{X}_\mathbf{s}] \in \mathbb{R}^{N\times 2F}$, a graph embedding is $e = \sigma(\mathbf{A}\,\sigma(\mathbf{A}\tilde{\mathbf{X}}W^{(1)})\,W^{(2)})$.
- $\bar{e}_c$: The average graph representation for class c.
- $d_{c,c+1} = ||\bar{e}_c - \bar{e}_{c+1}||_{l2}$: Distance between adjacent classes (i.e., c-th and $(c+1)$-th class).
- To enforce the sequential relations between classes, distances between three adjacent classes are aligned in the latent space as $d_{c,c+1} + d_{c+1,c+2} = d_{c,c+2}$.

$$R_{\text{temp}} = \frac{1}{C-2}\sum_{c=1}^{C-2}\Big(d_{c,c+1} + d_{c+1,c+2} - d_{c,c+2}\Big) \tag{5}$$

(3) Training Objective: $\mathcal{L} = L_{\text{ce}} + \alpha R_{\text{temp}}$, where L_{ce} is a standard cross-entropy loss.

Dataset

(a) Alzheimer's Disease Neuroimaging Initiative (ADNI)

Node feature	Edge feature	CN	SMC	EMCI	LMCI	AD
Cortical Thickness	# of white matter tracts	359	181	437	180	166
FDG		345	186	461	231	162

(b) Parkinson's Progression Markers Initiative (PPMI)

Node feature	Edge feature	CN	Prodromal	PD
BOLD	Correlation of BOLD	15	67	113

Table: Number of subjects for ADNI and PPMI datasets. The diseases progress from CN to AD/PD class.

Quantitative Result

Table: Classification performance on the ADNI and PPMI datasets with 5-fold cross-validation.

Model	ADNI-CT			ADNI-FDG			PPMI			
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Specificity
SVM (Linear)	82.4%	82.2%	85.2%	85.3%	85.7%	86.9%	60.5%	30.2%	28.0%	67.5%
MLP (2-layers)	78.8%	79.2%	79.9%	87.5%	88.2%	88.2%	68.9%	36.3%	39.0%	70.0%
GCN	61.4%	59.8%	62.6%	68.8%	67.7%	69.7%	78.8%	48.1%	70.3%	82.5%
GAT	64.2%	62.7%	66.8%	69.2%	67.0%	73.6%	81.2%	51.4%	77.2%	87.0%
GDC	77.1%	76.9%	78.5%	86.2%	86.7%	87.0%	73.0%	36.5%	61.8%	79.1%
GraphHeat	70.9%	70.3%	71.8%	77.0%	77.5%	77.3%	79.1%	48.4%	84.5%	68.8%
ADC	82.1%	77.6%	72.8%	88.6%	70.8%	75.3%	78.8%	50.7%	66.9%	80.5%
Exact	86.2%	86.6%	86.7%	90.2%	90.7%	90.7%	79.5%	48.1%	76.6%	87.3%
LSAP	87.0%	86.8%	88.5%	90.4%	90.9%	91.4%	79.1%	47.5%	83.2%	72.9%
BrainGNN	69.3%	20.1%	23.4%	68.9%	20.3%	31.9%	69.6%	38.5%	70.5%	39.9%
BrainNetTF	87.8%	65.9%	70.6%	87.1%	66.5%	66.3%	71.3%	42.6%	76.1%	53.3%
AGT (Ours)	**90.3%** (+2.5)	**91.3%** (+4.5)	**89.9%** (+1.4)	**94.8%** (+4.4)	**94.3%** (+3.4)	**95.3%** (+3.9)	**83.6%** (+2.4)	**62.4%** (+11.0)	**87.6%** (+3.1)	**92.3%** (+5.0)

Qualitative Results

(1) Analysis on Trained Scales. High-frequency components focus on more node-specific details and thus potentially highlight the importance of individual nodes.

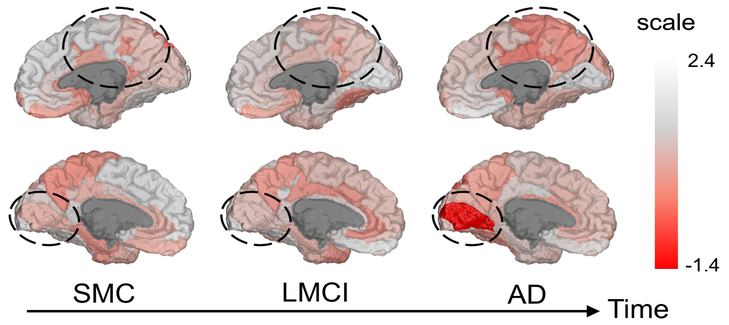
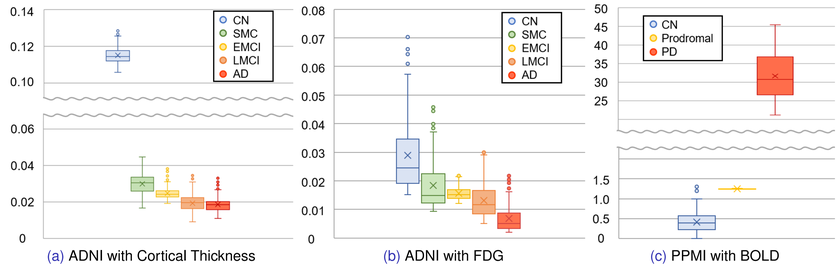


Figure: On the ADNI-FDG experiment, scales of *the precuneus gyrus* and *the medial occipitotemporal gyrus* decrease at later stages, accentuating high-frequency features on these ROIs. They are known to be regions where early-onset AD symptoms appear (Karas et al., 2007; Convit et al., 2000).

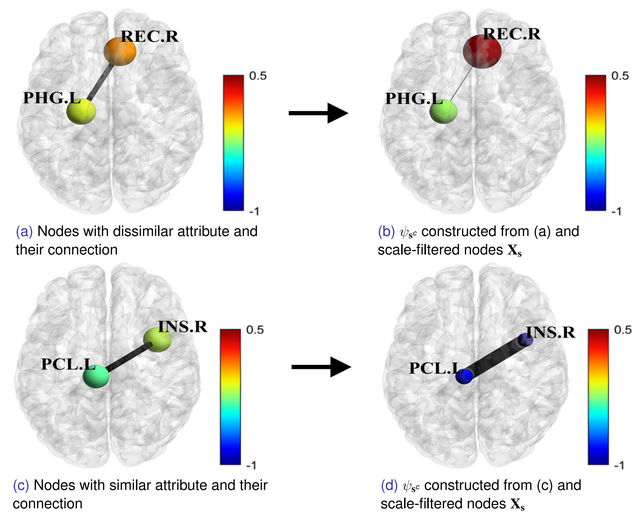
(2) Temporal Analysis.



(a) ADNI with Cortical Thickness (b) ADNI with FDG (c) PPMI with BOLD

Figure: Comparison of distributions of the graph embedding $\bar{e}_c$ between diagnostic classes. Temporal variations between the groups appear along with the disease progression.

(3) Effect of Node-wise Wavelet Basis.



(a) Nodes with dissimilar attribute and their connection (b) $\psi_{\mathbf{s}^c}$ constructed from (a) and scale-filtered nodes $\mathbf{X}_\mathbf{s}$

(c) Nodes with similar attribute and their connection (d) $\psi_{\mathbf{s}^c}$ constructed from (c) and scale-filtered nodes $\mathbf{X}_\mathbf{s}$

Figure: (a) and (c) are from a brain network (i.e., $\mathbf{X}$ and $\mathbf{A}$) of a subject with PD. Edges in (b) and (d) are the adaptive wavelet basis $\psi_{\mathbf{s}^c}$ derived from a trained model. By using $\psi_{\mathbf{s}^c}$, the node features in (b) become more discriminative from (a), while those in (d) become more similar compared to the nodes in (c).