BEYOND LOW-PASS FILTERS: ADAPTIVE FEATURE PROPAGATION ON GRAPHS

Abstract: Graph neural networks (GNNs) have been extensively studied for prediction tasks on graphs. Recent studies show that most GNNs assume local homophily, i.e., strong similarities in local neighborhoods. This assumption however limits the generalizability power of GNNs. To address this limitation, we propose a flexible GNN model, ASGAT, which is capable of handling any graphs without being restricted by their underlying homophily. At its core, ASGAT adopts a node attention mechanism based on multiple learnable spectral filters; therefore, the aggregation scheme is learned adaptively for each graph in the spectral domain. We evaluated ASGAT on node classification tasks over eight benchmark datasets. ASGAT is shown to generalize well to both homophilic and heterophilic graphs.

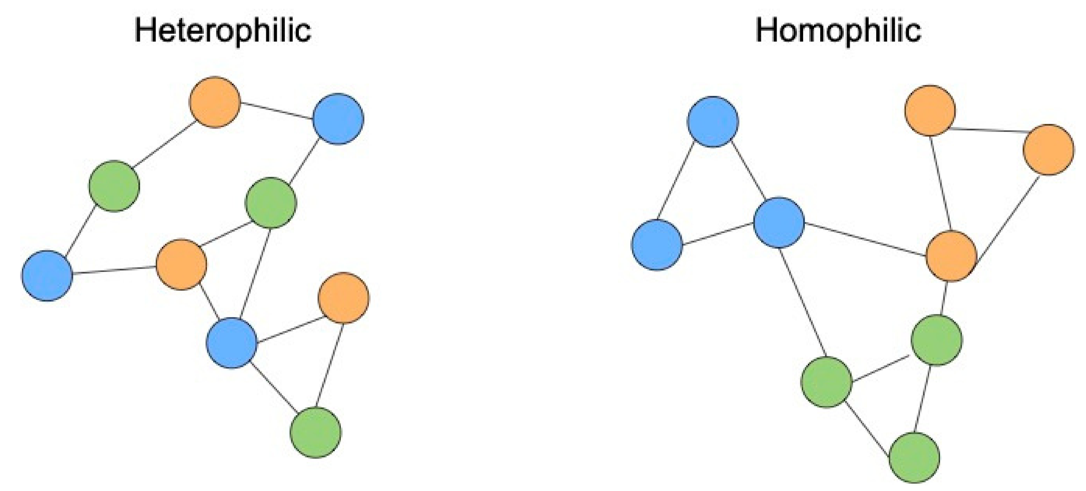
Shouheng Li[1], Dongwoo Kim[2], Qing Wang[1]

[1]School of Computing, Australian National University
[2]Graduate School of Artifical Intelligence, Postech

Australian National University

{shouheng.li; qing.wang}@anu.edu.au; dongwookim@postech.ac.kr

INTRODUCTION

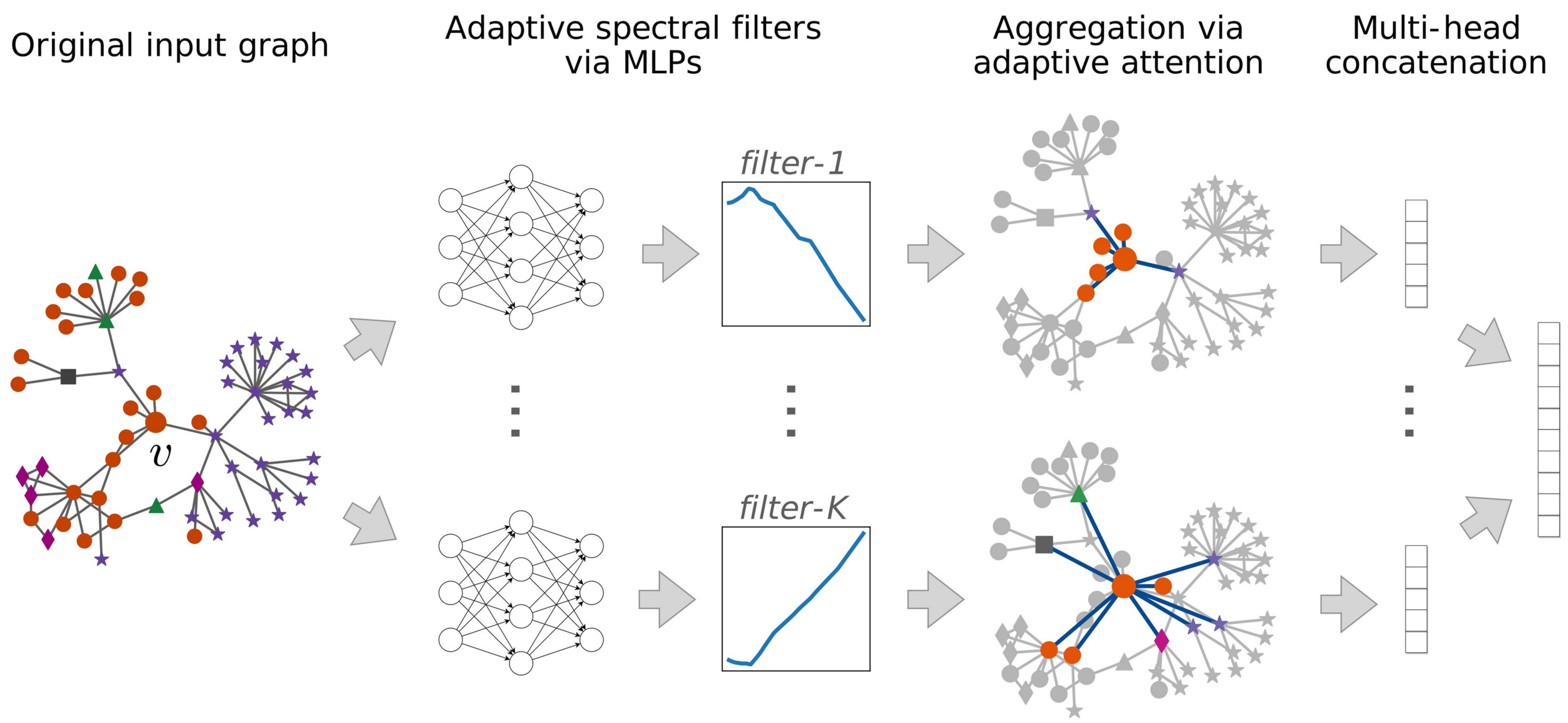


Heterophilic graphs commonly exist in the real world, for instance, people tend to connect to the opposite gender in dating networks, and different amino acid types are more likely to form connections in protein structures.

We study the node classification tasks on such graphs. Our contributions are:

- We show that high-frequency components carry important information on heterophilic graphs.
- We propose a GNN model which performs well on both homophilic and heterophilic graphs.
- We exhibit that multiple spectral filters work better than a single filter, as it is more flexibility.

ASGAT: ILLUSTRATION



EMPIRICAL RESULTS

	Homophily ⟷							Heterophily
	Cora	Pubmed	CiteSeer	Chameleon	Squirrel	Wisconsin	Cornell	Texas
GCN	87.4 ± 0.2	87.8 ± 0.2	78.5 ± 0.5	$59.8 \pm 2.6^{\ddagger}$	$36.9 \pm 1.3^{\ddagger}$	64.1 ± 6.3	59.2 ± 3.2	64.1 ± 4.9
ChevNet	88.2 ± 0.2	89.3 ± 0.3	79.4 ± 0.4	66.0 ± 2.3	39.6 ± 3.0	82.5 ± 2.8	76.5 ± 9.4	79.7 ± 5.0
ARMANet	85.2 ± 2.5	86.3 ± 5.7	76.7 ± 0.5	62.1 ± 3.6	47.8 ± 3.5	78.4 ± 4.6	74.9 ± 2.9	82.2 ± 5.1
GAT	87.6 ± 0.3	83.0 ± 0.1	77.7 ± 0.3	$54.7 \pm 2.0^{\ddagger}$	$30.6 \pm 2.1^{\ddagger}$	62.0 ± 5.2	58.9 ± 3.3	60.0 ± 5.7
SGC	87.2 ± 0.3	81.1 ± 0.3	78.8 ± 0.4	33.7 ± 3.5	46.9 ± 1.7	51.8 ± 5.9	58.1 ± 4.6	58.9 ± 6.1
GraphSAGE	86.3 ± 0.6	89.2 ± 0.5	77.4 ± 0.5	51.1 ± 0.5	$41.6 \pm 0.7^{\ddagger}$	77.6 ± 4.6	67.3 ± 6.9	82.7 ± 4.8
APPNP	**88.4** ± 0.3	86.0 ± 0.3	77.6 ± 0.6	45.3 ± 1.6	31.0 ± 1.6	81.2 ± 2.5	70.3 ± 9.3	79.5 ± 4.6
Geom-GCN	86.3 ± 0.3	89.1 ± 0.1	**81.4** ± 0.3	$60.9^{\dagger}$	$38.1^{\dagger}$	$64.1^{\dagger}$	$60.8^{\dagger}$	$67.6^{\dagger}$
H_2GCN	88.3 ± 0.3	89.1 ± 0.4	78.4 ± 0.5	59.4 ± 2.0	37.9 ± 2.0	86.5 ± 4.4	82.2 ± 6.0	82.7 ± 5.7
MLP	72.1 ± 1.3	88.6 ± 0.2	74.9 ± 1.8	45.7 ± 2.7	28.1 ± 2.0	82.7 ± 4.5	81.4 ± 6.3	79.2 ± 6.1
Vanilla ASGAT	–	–	–	–	–	**86.9** ± 4.2	**84.6** ± 5.8	82.2 ± 3.2
ASGAT-Cheb	87.5 ± 0.5	**89.9** ± 0.9	79.3 ± 0.6	**66.5** ± 2.8	**55.8** ± 3.2	86.3 ± 3.7	82.7 ± 8.3	**85.1** ± 5.7
ASGAT-ARMA	87.4 ± 1.1	88.3 ± 1.0	79.2 ± 1.4	65.8 ± 2.2	51.4 ± 3.2	84.7 ± 4.4	83.2 ± 5.5	79.5 ± 7.7

- ASGAT outperforms all benchmarks on heterophilic graphs.
- ASGAT performs particularly well on classifying nodes of low homophily

PROPOSED METHOD: ASGAT

1. Adaptive Spectral Filtering

Lower-frequency components carry smoothly changing signals, thus often used to "smooth" a homophilic graph. On the contrary, higher-frequency components carry abruptly changing signals, corresponding to the discontinuities and "opposite attraction" on heterophilic graphs. ASGAT utilizes both via trainable spectral filters.

Normalized Laplacian Matrix

$$\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2} = \boldsymbol{U}\Lambda\boldsymbol{U}^H$$

Spectral Convolution/Filtering

$$g(\boldsymbol{L})\vec{x} = \boldsymbol{U}g(\Lambda)\widehat{\vec{x}}$$

Adaptive filtering via MLP

$$\boldsymbol{\psi}_v = \boldsymbol{U}\mathrm{diag}(\mathrm{MLP}(\Lambda))\boldsymbol{U}^H\delta_v$$

2. Attention Sparsification

We use a sparsification technique to keep only the largest k attention entries for each node, before passing them through Softmax normalizing.

Sparsification

$$\bar{\psi}_{vu} = \begin{cases} \psi_{vu} & \text{if } \psi_{vu} \in \mathrm{topK}(\{\psi_{v0}, \ldots, \psi_{vN}\}, k) \\ -\infty & \text{otherwise} \end{cases}$$

Normalization

$$\boldsymbol{a}_v = \mathrm{softmax}(\overline{\boldsymbol{\psi}}_v)$$

3. Aggregate & Combine

An update layer has two steps: **AGGREGATE** and **COMBINE**. Attentions computed from adaptive spectral filtering are not always localized. Hence, ASGAT can adaptively aggregate information from both close and far-distant nodes.

$$\boldsymbol{h}_v^{(l)} = \sigma\left(\sum_{u=1}^{N} a_{vu}\boldsymbol{h}_u^{(l-1)}\boldsymbol{W}^{(l)}\right)$$

RELATED LITERATURE

- Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., Koutra, D.: Beyond homophily in graph neural networks: Current limitations and effective designs, NeurIPS 2020
- Xu, B., Shen, H., Cao, Q., Qiu, Y., Cheng, X.: Graph wavelet neural network, ICLR 2019
- Klicpera, J., Weißenberger, S., Günnemann, S.: Diffusion improves graph learning. NeurIPS 2019
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio,Y.: Graph attention networks, ICLR 2018
- Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P.: The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process.Mag, 2013