p-LAPLACIAN BASED GRAPH NEURAL NETWORKS

Tencent AI Lab

GUOJI FU, PEILIN ZHAO, YATAO BIAN

ICML International Conference On Machine Learning

INTRODUCTION

Homophily Assumption of GNNs: The labels of linked nodes in a graph are the same or consistent.

Problem: Most low-pass filter GNNs (e.g., GCN, SGC, APPNP) fail on heterophilic graphs where the labels of linked nodes are more likely to differ.

Contributions:

1. **New methodologies.** We propose the p-Laplacian message passing and ^{p}GNNs to adapt GNNs to heterophilic graphs.
2. **Theoretical justification.** We theoretically demonstrate that p-Laplacian message passing works as low-high-pass filters.
3. **New paradigm of designing GNN architectures.** We bridge the gap between discrete regularization framework and GNNs, which could be used to develop new message passing schemes and GNN architectures.

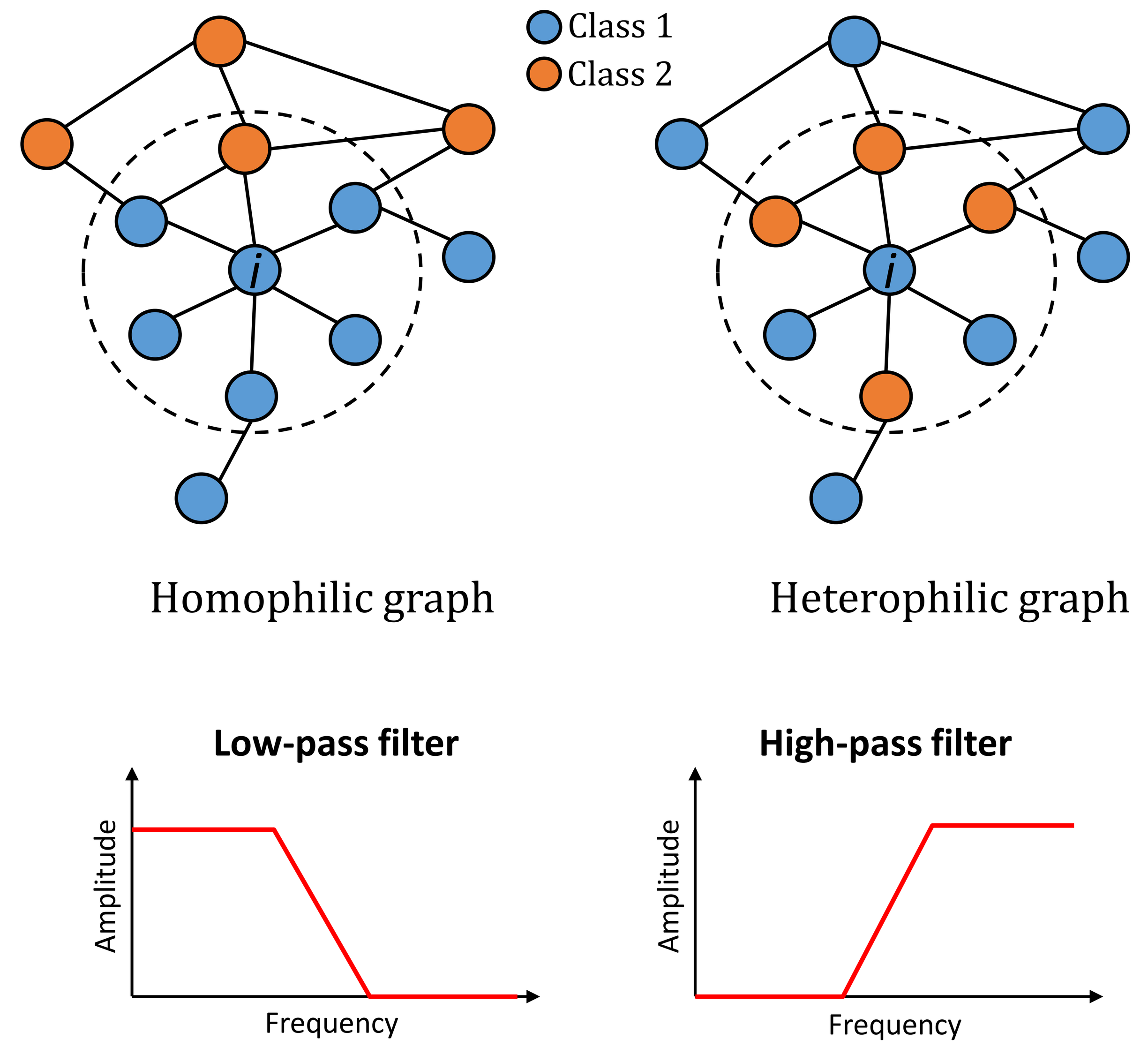


Figure 1: Homophilic & heterophilic graphs.

p-LAPLACIAN MESSAGE PASSING & ^{p}GNNS ARCHITECTURE

1. p-Laplacian Regularization Framework:

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node features $\mathbf{X}$,

$$\mathcal{L}_p(\mathbf{F}) := \min_{\mathbf{F}} \mathcal{S}_p(\mathbf{F}) + \mu \sum_{i=1}^{N} \|\mathbf{F}_{i,:} - \mathbf{X}_{i,:}\|^2,$$

where $\mu \in (0, \infty)$ and

$$\mathcal{S}_p(f) := \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \sqrt{\frac{W_{i,j}}{D_{j,j}}} f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}} f(i) \right\|^p.$$

Regularization with $p > 1$: we could use the following iteration algorithm to get a solution: for all $i, j \in [N]$,

$$\mathbf{F}_{i,:}^{(k+1)} = \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{\sqrt{D_{i,i} D_{j,j}}} \mathbf{F}_{j,:}^{(k)} + \beta_{i,i}^{(k)} \mathbf{X}_{i,:},$$

$$M_{i,j}^{(k)} = W_{i,j} \left\| \sqrt{\frac{W_{i,j}}{D_{i,i}}} \mathbf{F}_{i,:}^{(k)} - \sqrt{\frac{W_{i,j}}{D_{j,j}}} \mathbf{F}_{j,:}^{(k)} \right\|^{p-2},$$

$$\alpha_{i,i}^{(k)} = 1 \Big/ \left(\sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{D_{i,i}} + \frac{2\mu}{p} \right), \quad \beta_{i,i}^{(k)} = \frac{2\mu}{p} \alpha_{i,i}^{(k)}.$$

2. p-Laplacian Message Passing:

$$\mathbf{F}^{(k+1)} = \boldsymbol{\alpha}^{(k)} \mathbf{D}^{-1/2} \mathbf{M}^{(k)} \mathbf{D}^{-1/2} \mathbf{F}^{(k)} + \boldsymbol{\beta}^{(k)} \mathbf{X}.$$

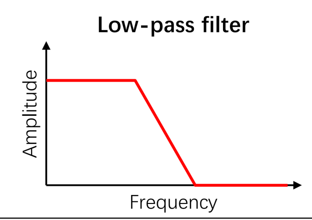
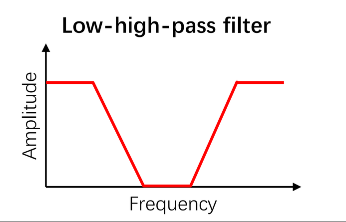
It is guaranteed to converge under merit settings.

3. ^{p}GNN Architecture:

$$\mathbf{F}^{(0)} = \mathrm{ReLU}(\mathbf{X}\Theta^{(1)}),$$
$$\mathbf{F}^{(k+1)} = \boldsymbol{\alpha}^{(k)} \mathbf{D}^{-1/2} \mathbf{M}^{(k)} \mathbf{D}^{-1/2} \mathbf{F}^{(k)} + \boldsymbol{\beta}^{(k)} \mathbf{F}^{(0)},$$
$$\mathbf{Z} = \mathrm{softmax}(\mathbf{F}^{(K)}\Theta^{(2)}),$$

4. Spectral Analysis.

It adaptively acts as **low-pass** & **low-high-pass** filters on each node in terms of its graph gradients.

		
$1 \le p < 2$	$\|\nabla f(i)\| < 2(2\sqrt{N_k})^{1/(p-2)}$	$\|\nabla f(i)\| \ge 2(2\sqrt{N_k})^{1/(p-2)}$
$p = 2$	—	Always
$p > 2$	$\|\nabla f(i)\| > 2^{(p-1)/(p-2)}$	$\|\nabla f(i)\| \le 2^{(p-1)/(p-2)}$

RESULTS

Table 1: Results on heterophilic benchmark datasets.

Method	Chameleon	Squirrel	Actor	Wisconsin	Texas	Cornell
MLP	48.02±1.72	**33.80**±1.05	39.68±1.43	93.56±3.14	79.50±10.62	80.30±11.38
GCN	34.54±2.78	25.28±1.55	31.28±2.04	61.93±3.00	56.54±17.02	51.36±4.59
SGC	34.76±4.55	25.49±1.63	30.98±3.80	66.94±2.58	59.99±9.95	44.39±5.88
GAT	45.16±2.10	31.41±0.98	34.11±1.28	65.64±6.29	56.41±13.01	43.94±7.33
JKNet	33.28±3.59	25.82±1.58	29.77±2.61	61.08±3.71	59.65±12.62	55.34±4.43
APPNP	36.18±2.81	26.85±1.48	31.26±2.52	64.59±3.49	82.90±5.08	66.47±9.34
GPRGNN	43.67±2.27	31.27±1.76	36.63±1.22	88.54±4.94	80.74±6.76	78.95±8.52
$^{1.0}$GNN	**48.86**±1.95	**33.75**±1.50	**40.62**±1.25	**95.37**±2.06	84.06±7.41	**82.16**±8.62
$^{1.5}$GNN	**48.74**±1.62	33.33±1.45	**40.35**±1.35	**95.24**±2.01	84.46±7.79	78.47±6.87
$^{2.0}$GNN	**48.77**±1.87	33.60±1.47	40.07±1.17	91.15±2.76	**87.96**±6.27	72.04±8.22
$^{2.5}$GNN	**48.80**±1.77	**33.79**±1.45	39.80±1.31	87.08±2.69	83.01±6.80	70.31±8.84

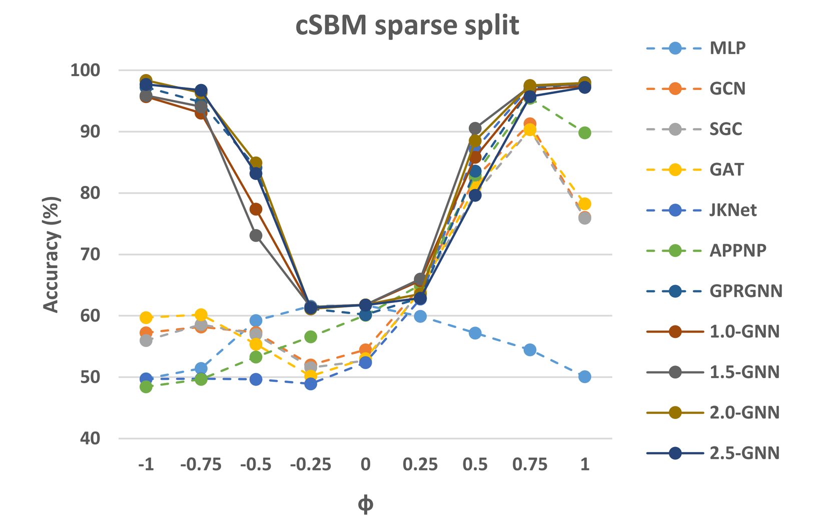


Figure 3: Results on cSBM (sparse split).

Results: Table 1 and Figure 3 show that ^{p}GNNs significantly outperform several SODA GNN architectures on heterophilic benchmarks and cSBM datasets. Figure 2 shows that ^{p}GNNs can adaptively learn aggregation weights and Figure 4 shows that ^{p}GNNs are more robust to noisy edges.

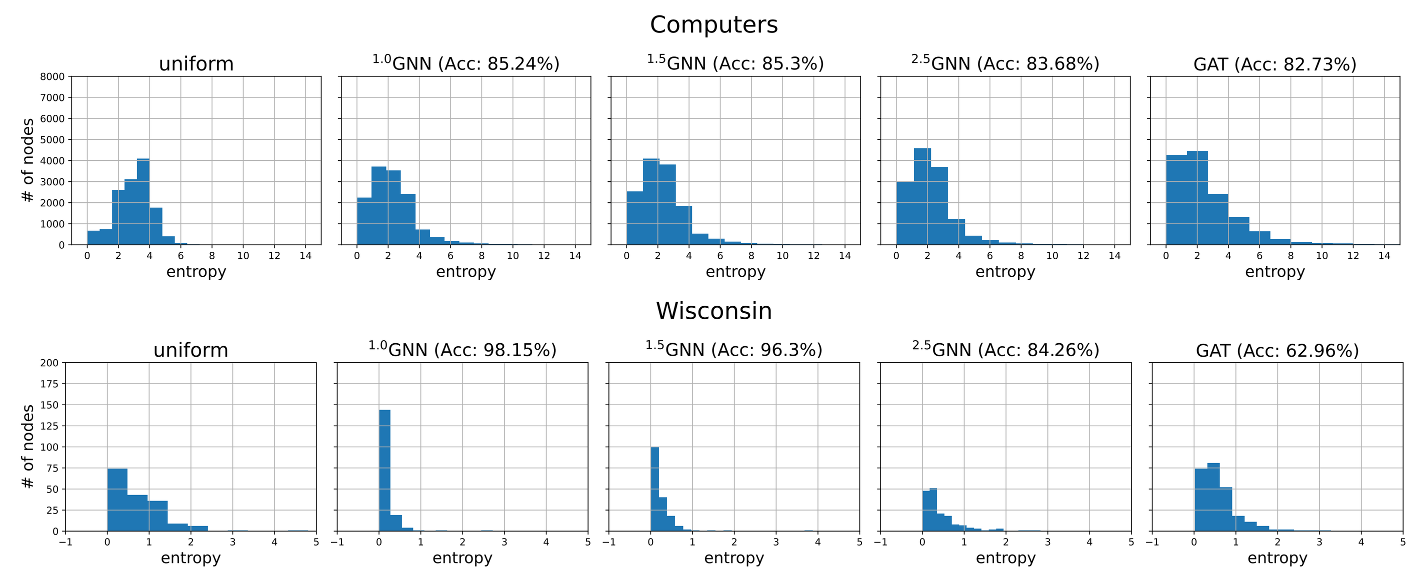


Figure 2: Aggregation weight entropy distribution.

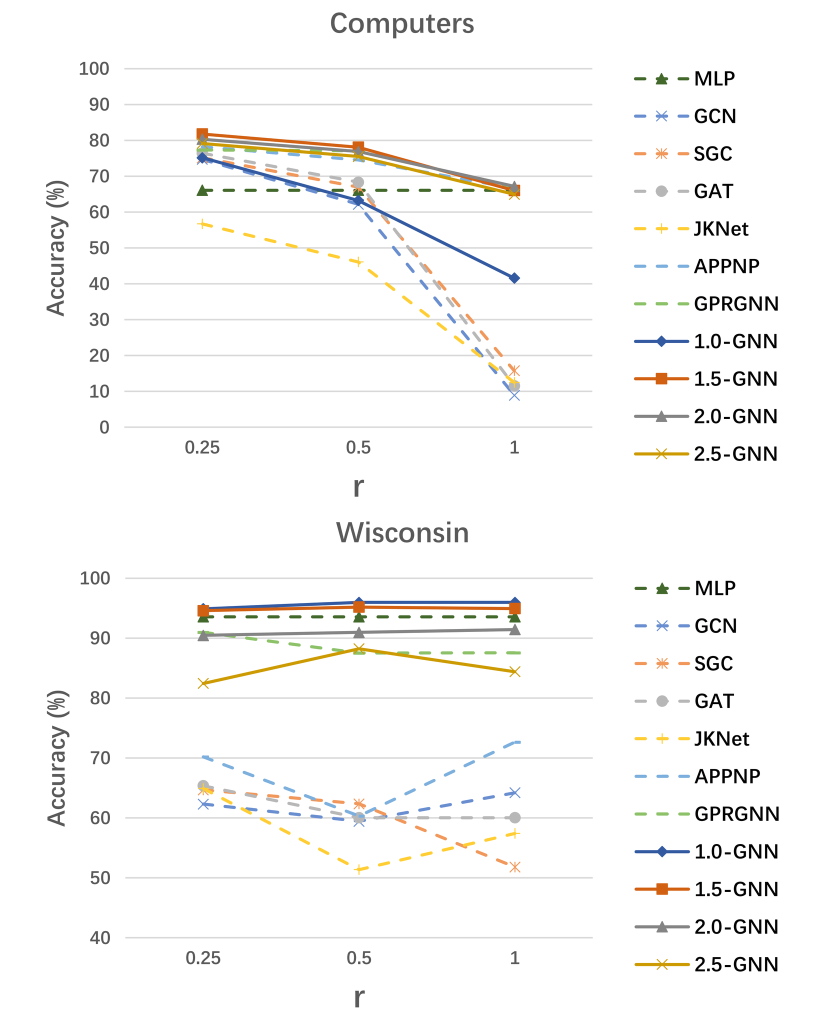


Figure 4: Results on graphs with noisy edges.

CONCLUSION

We addressed the problem of generalizing GNNs to heterophilic graphs. To this end, we derived a novel p-Laplacian message passing scheme from a discrete regularization framework and proposed a new ^{p}GNN architecture. We theoretically demonstrate our method works as low-high-pass filters and thereby applicable to both homophilic and heterophilic graphs. We empirically validate our theoretical results and show the advantages of our methods on heterophilic graphs and graphs with non-informative topologies. We also bridge the gap between discrete regularization framework, graph-based semi-supervised learning, and GNNs, which provides a new paradigm of designing new GNN architectures.

REFERENCES

[1] Dengyong Zhou and Bernhard Schölkopf. Regularization on discrete spaces. In *DAGM*, volume 3663, pages 361–368, 2005.

[2] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[3] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

[4] Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In *ICML*, pages 6861–6871, 2019.

[5] Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. In *ICLR*. OpenReview.net, 2019.