

When Do Graph Neural Networks Help with Node Classification?

-Investigating the Impact of Homophily Principle on Node Distinguishability

Sitao Luan, Chenqing Hua, Minkai Xu, Qincheng Lu, Jiaqi Zhu, Xiao-Wen Chang, Jie Fu, Jure Leskovec, Doina Precup

Why Do We Choose GNNs over NNs?

Graph-aware model: a network that includes the feature aggregation step according to graph structure, e.g., GCN, SGC-1.

Graph-agnostic model: a network that does not use graph structure information, such as Multi-Layer Perceptron (MLP)

GCN: $\text{Softmax}(\hat{A}\text{ReLU}(\hat{A}XW_0)W_1)$, **MLP-2:** $\text{Softmax}(\text{ReLU}(XW_0)W_1)$

SGC-1: $\text{Softmax}(\hat{A}XW_0)$, **MLP-1:** $\text{Softmax}(XW_0)$

The difference between graph-aware model and graph-agnostic model only lies in whether \hat{A} is multiplied in each layer, e.g., GCN vs. MLP-2 and SGC-1 vs. MLP-1. Therefore, whether we choose GNNs or NNs depends on the “quality” of \hat{A} (the aggregation step).

Question: How to know if \hat{A} is “good” or “bad”? Can homophily tells?

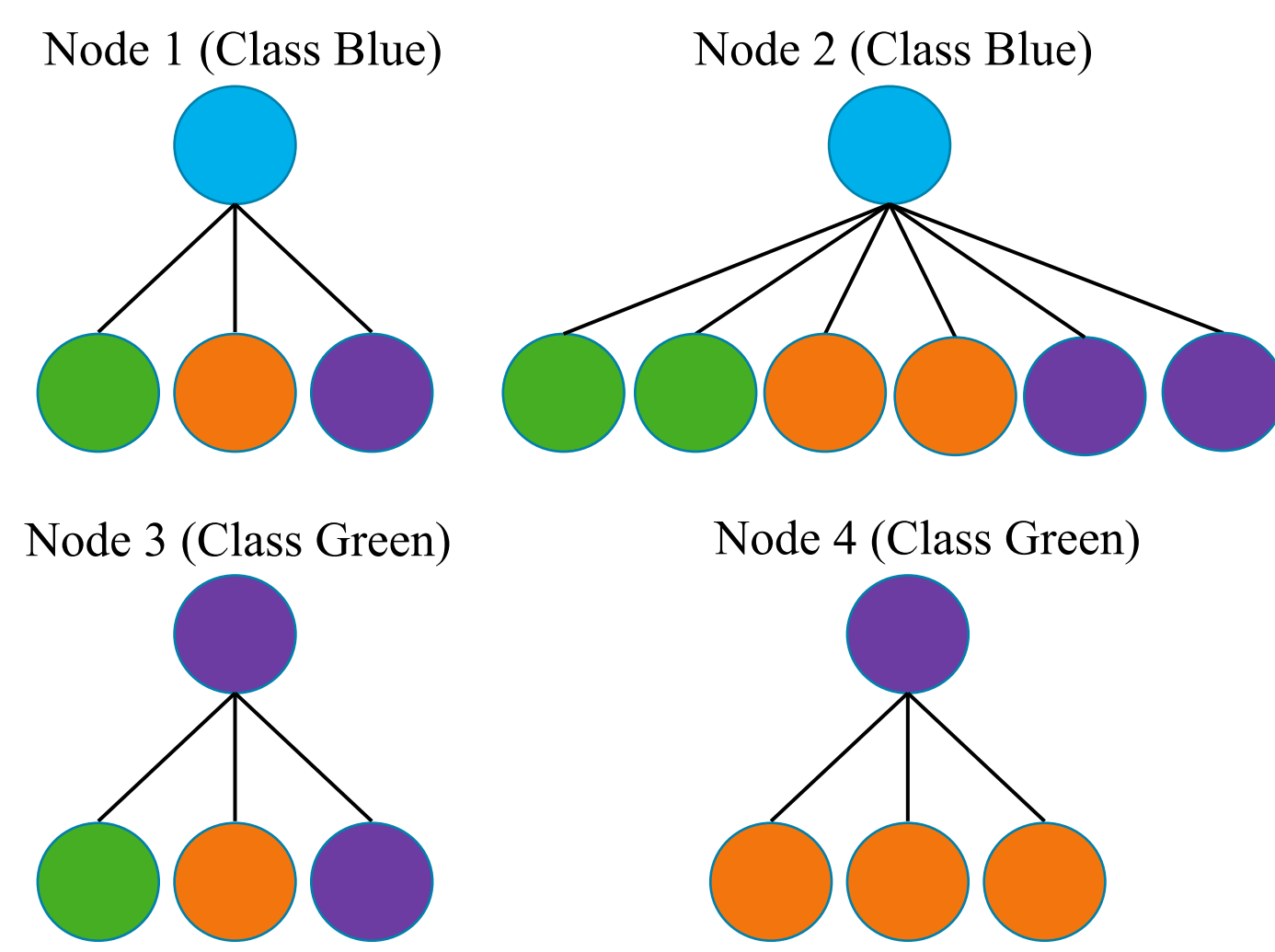
Homophily and Heterophily

Homophily: nodes from the same class are more likely to be connected. The most important reason for the success of GNNs on node-level tasks.

Heterophily: opposite of homophily, nodes from different classes are more likely to be connected. Node features from different classes will be mixed and become indistinguishable.

Homophily wins, Heterophily loses? NO [1,2] ❌

Homophily is not necessary [1]: as long as nodes within the same class share similar neighborhood patterns, their embeddings will be similar after aggregation. For example, although nodes {1,2} have heterophilic neighborhood patterns, their neighborhood patterns are the same. As a result, their aggregated features are still be similar and will be classified into the same class.



Problem: The above understanding is incomplete, because it only discusses intra-class ND, but forgets inter-class ND. For example, node 3 is from class green and has the same neighborhood pattern as {1,2}, which means node 3 and nodes {1,2} will become indistinguishable after aggregation.

Takeaway: This highlights the necessity for careful consideration of both intra- and inter-class ND when evaluating the impact of homophily on the performance of GNNs. An ideal case would be node {1,2,4}, where we have smaller intra-class "distance" than inter-class "distance".

Question: How to quantify node distinguishability and its correlation with homophily?

Quantify Node Distinguishability

CSBM-H The features of 2 classes of nodes, $i \in \mathcal{C}_0$ and $j \in \mathcal{C}_1$, are generated independently, with $\mathbf{x}_i \sim N(\boldsymbol{\mu}_0, \sigma_0^2 \mathbf{I})$ and $\mathbf{x}_j \sim N(\boldsymbol{\mu}_1, \sigma_1^2 \mathbf{I})$; d_0, d_1 are node degree for i, j . For $i \in \mathcal{C}_0$, its neighbors are generated by independently sampling from $h \cdot d_0$ intra-class nodes and $(1 - h) \cdot d_0$ inter-class nodes. The neighbors of $j \in \mathcal{C}_1$ are generated in the same way. The parameters of low-pass and high-pass filtered features can be computed accordingly.

Theorem 1 (Optimal Bayes Classifier CL_{Bayes} for CSBM-H) Suppose the prior distribution for \mathbf{x} is $\mathbb{P}(\mathbf{x} \in \mathcal{C}_0) = \mathbb{P}(\mathbf{x} \in \mathcal{C}_1) = 1/2$, then we have

$$\text{CL}_{\text{Bayes}}(\mathbf{x}) = \begin{cases} 1, \eta(\mathbf{x}) \geq 0.5 \\ 0, \eta(\mathbf{x}) < 0.5 \end{cases}, \eta(\mathbf{x}) = \mathbb{P}(z = 1|\mathbf{x}) = \frac{1}{1 + \exp(Q(\mathbf{x}))}$$

where $Q(\mathbf{x}) = \mathbf{a}\mathbf{x}^T\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, z_i is the label of node i . See the calculation of $\mathbf{a}, \mathbf{b}, c$ in the paper.

Probabilistic Bayes Error (PBE) can be calculated for the optimal Bayes classifier of CSBM-H. PBE can be used to measure ND as follows,

$$\frac{\text{CDF}_{\tilde{\chi}(\omega_0, F_h, \lambda_0)}(-\xi) + (1 - \text{CDF}_{\tilde{\chi}(\omega_1, F_h, \lambda_1)}(-\xi))}{2}$$

See the parameters and the calculation in our paper. PBE can be numerically calculated to show the relation between h and ND precisely. However, it doesn't have an analytic expression, which makes it less explainable.

Generalized Jeffreys Divergence D_{NGJ} A KL-divergence based statistical measure which can provide an explainable metric of ND as follows

$$D_{NGJ}(\text{CSBM-H}) = -d_X^2 \left(\frac{1}{4\sigma_1^2} + \frac{1}{4\sigma_0^2} \right) - \frac{F_h}{4} [\rho^2 + \frac{1}{\rho^2} - 2]$$

Where $d_X^2 = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$ describes inter-class ND; $\rho = \frac{\sigma_0}{\sigma_1}$.

- The normalized distance term verifies that ND depends on both intra- and inter-class ND; The larger the difference between the two intra-class ND is, the nodes are more distinguishable and vice versa.
- Plot the relation between homophily and metrics of ND.

Visualization & Analysis

Mid-homophily pitfall: For low-pass (LP) filtered features, the most indistinguishable area is the medium homophily level, instead of the extreme low homophily levels.

Three regimes: low-pass (LP), high-pass (HP) and full-pass (FP) filters dominate different homophily areas. LP: very low and very high homophily intervals (two ends); HP: low to medium homophily levels; FP: medium to high homophily areas.

HP filter: high-pass filter works better in heterophily areas than in homophily areas

Ablation Study: From the comparison of Fig 1 and Fig 2, it is observed that as σ_1^2 increases, the PBE and D_{NGJ} of the three curves all go up, which means the node embeddings become less distinguishable under HP, LP and FP filters. The significant shrinkage of the HP regimes and the expansion of the FP regime indicates that the original features are more robust to imbalanced variances, especially in the low homophily area. (See more detailed and comprehensive analysis in our paper)

More General Theoretical Analysis we investigate how significant the intra-class embedding distance is smaller than inter-class embedding distance under more general conditions. The theorem verifies that ND is related to the combined effect of inter- and intra-class ND. Besides, we found that the impact of HP filter depends on the relative center distance (See the above Fig), which is novel.

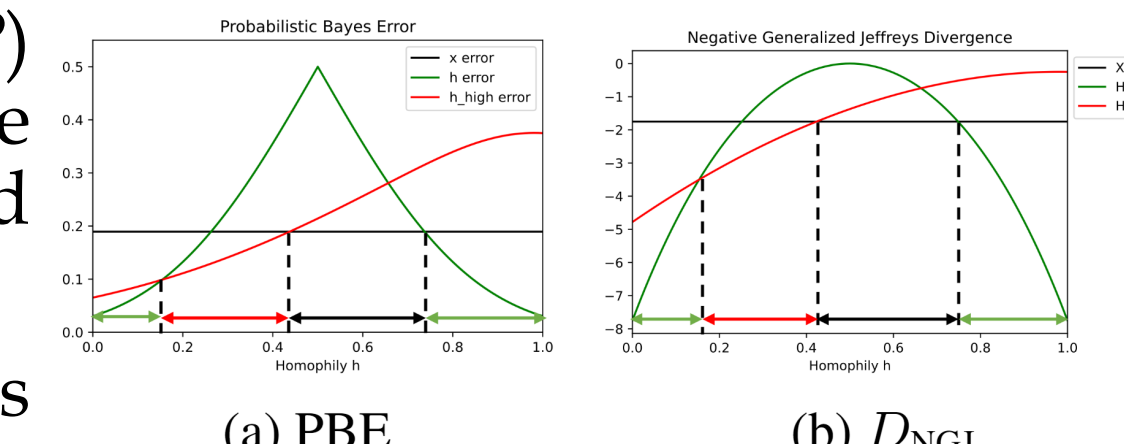


Fig 1: $\mu_0 = [-1, 0], \mu_1 = [1, 0], \sigma_0^2 = 1, \sigma_1^2 = 2, d_0 = 5, d_1 = 5$

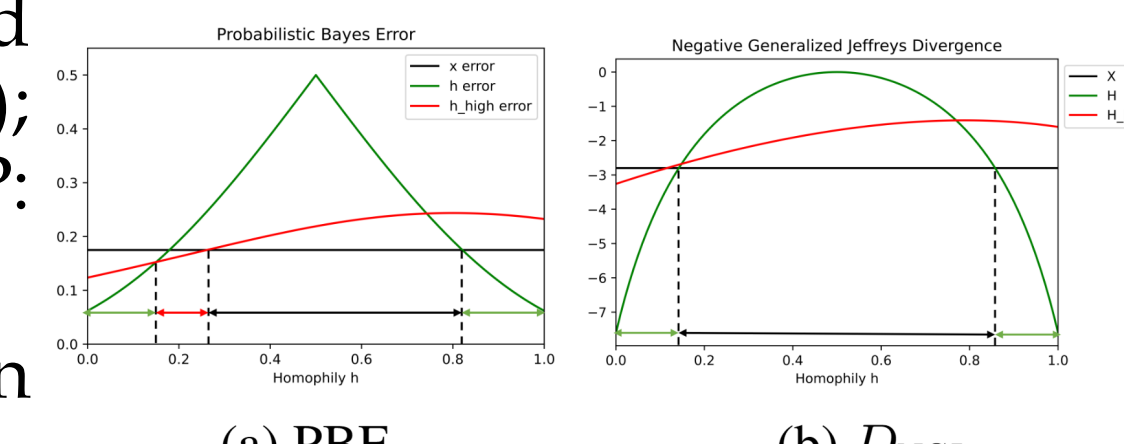
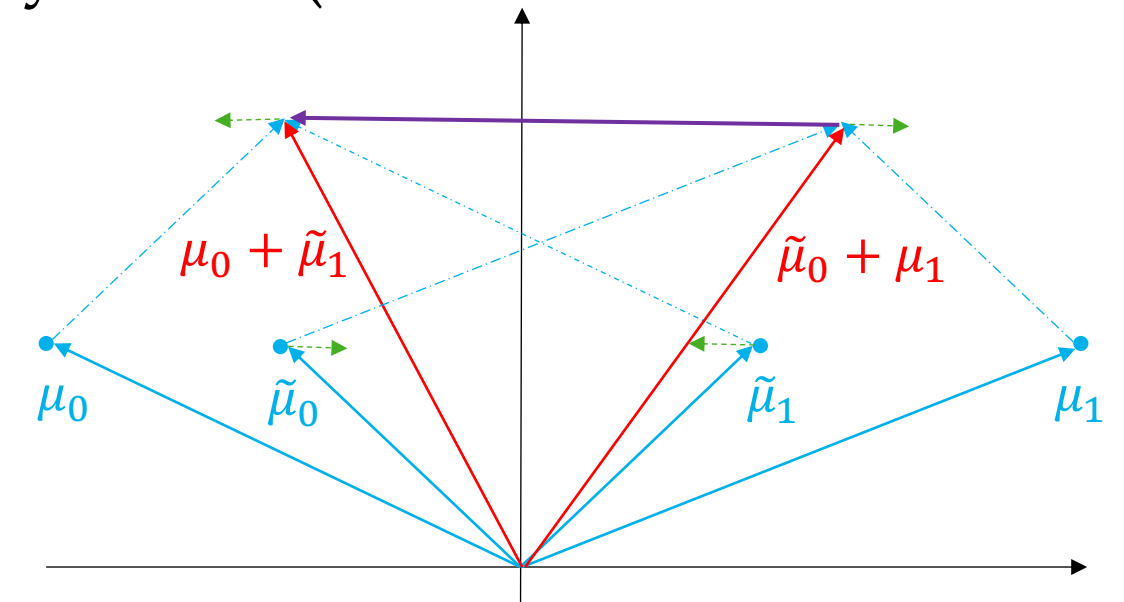
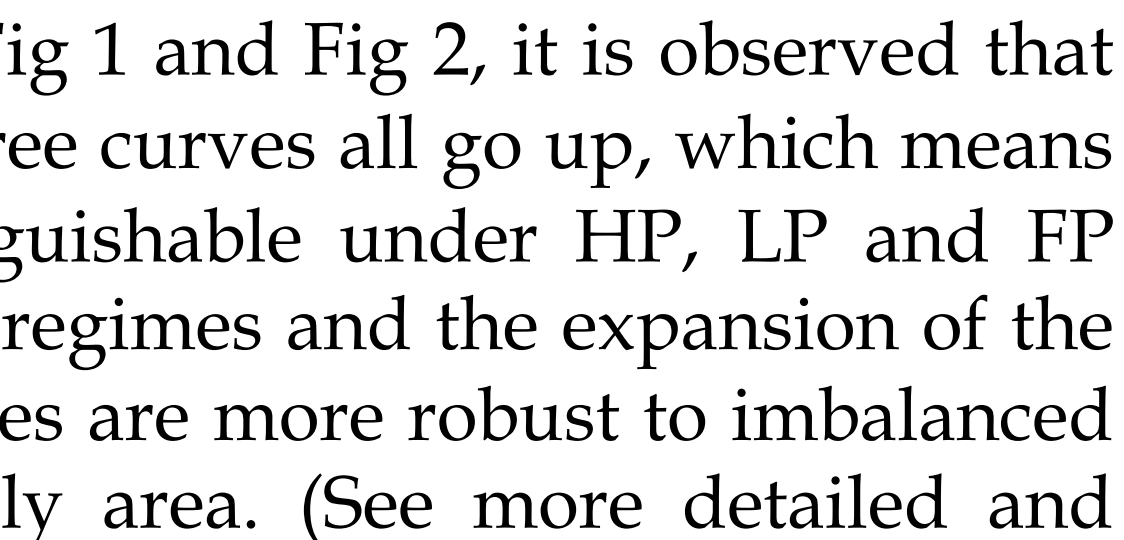


Fig 2: $\mu_0 = [-1, 0], \mu_1 = [1, 0], \sigma_0^2 = 1, \sigma_1^2 = 5, d_0 = 5, d_1 = 5$



Empirical Study of ND

We conduct hypothesis tests with GCN vs. MLP-2 and SGC-1 vs. MLP-1 to verify **whether "intra-class embedding distance is smaller than the inter-class embedding distance" strongly relates to the superiority of graph-aware models**. We compute the proportion of nodes whose intra-class node distance is significantly smaller than inter-class node distance (Prop). Then, $H_0: \text{Prop}(\text{G-aware}) \geq \text{Prop}(\text{G-agnostic})$; $H_1: \text{Prop}(\text{G-aware}) < \text{Prop}(\text{G-agnostic})$

We compare the Prop results 100 times on 9 benchmark datasets. With the 100 samples, we conduct T-test to obtain the p-values. It is observed that, in most cases, **when H_1 significantly holds, graph-aware models will underperform the coupled graph-agnostic models and vice versa, no matter the homophily levels**. This confirms the strong correlation.

		Cornell	Wisconsin	Texas	Film	Chameleon	Squirrel	Cora	CiteSeer	PubMed
Baseline Homophily Metrics	H_{edge}	0.5669	0.4480	0.4106	0.3750	0.2795	0.2416	0.8100	0.7362	0.8024
	H_{node}	0.3855	0.1498	0.0968	0.2210	0.2470	0.2156	0.8252	0.7175	0.7924
	H_{class}	0.0468	0.0941	0.0013	0.0110	0.0620	0.0254	0.7657	0.6270	0.6641
	H_{agg}	0.8032	0.7768	0.694	0.6822	0.61	0.3566	0.9904	0.9826	0.9432
	H_{GE}	0.31	0.34	0.35	0.16	0.0152	0.0157	0.17	0.19	0.27
	H_{adj}	0.1889	0.0826	0.0258	0.1272	0.0663	0.0196	0.7388	0.7431	0.7431
	LI	0.0169	0.1311	0.1923	0.0002	0.048	0.0015	0.5904	0.4508	0.4093
Classifier-based Performance Metrics	KR w/acc	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
	GNN	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
	p-value	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00
	ACC SGC	70.98 \pm 8.39	70.38 \pm 2.85	83.28 \pm 5.43	25.26 \pm 1.18	64.86 \pm 1.81	47.62 \pm 1.27	85.12 \pm 1.64	79.66 \pm 0.75	85.5 \pm 0.76
SGC vs. MLP-1	ACC MLP-1	93.77 \pm 3.34	93.87 \pm 3.33	93.77 \pm 3.34	34.53 \pm 1.48	45.01 \pm 1.58	29.17 \pm 1.46	74.3 \pm 1.27	75.51 \pm 1.35	86.23 \pm 0.54
	Diff Acc	-22.79	-23.49	-10.49	-9.27	19.85	18.45	10.82	4.15	-0.73
	p-value	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00
GCN vs. MLP-2	ACC GCN	82.46 \pm 3.11	75.5 \pm 2.92	83.11 \pm 3.2	35.51 \pm 0.99	64.18 \pm 2.62	44.76 \pm 1.39	87.78 \pm 0.96	81.39 \pm 1.23	88.9 \pm 0.32
	ACC MLP-2	91.30 \pm 0.70	93.87 \pm 3.33	92.26 \pm 0.71	38.58 \pm 0.25	46.72 \pm 0.46	31.28 \pm 0.27	76.44 \pm 0.30	76.25 \pm 0.28	86.43 \pm 0.13
	Diff Acc	-8.84	-18.37	-9.15	-3.07	17.46	13.48	11.34	5.14	2.47

Beyond homophily metrics p-value is a better performance metric, because it can provide a statistical threshold value for the superiority of graph-aware models. This property is not present in the current homophily metrics. However, it is required to train and fine-tune the models to obtain the p-values, which make it less practical because of computational costs.

Classifier-based Performance Metrics

We use the simple classifiers (Gaussian Naïve Bayes (GNB) and Kernel Regression (KR) with Neural Network Gaussian Process (NNGP)) that don't require iterative training for hypothesis testing. From the above table and the figures below on synthetic graphs, we can see that the new classifier-based performance metrics are much more effective than the existing homophily metrics on revealing the advantage and disadvantage of GNNs.

Performance Metrics	Linear or Non-linear	Feature Dependency	Sparse Labels	Statistical Threshold
H_{edge}	linear	✗	✗	✗
H_{node}	linear	✗	✗	✗
H_{class}	linear	✗	✗	✗
H_{agg}	linear	✗	✓	✗
H_{GE}	linear	✓	✓	✗
H_{adj}	linear	✗	✗	✗
LI	linear	✗	✗	✗
Classifier	both	✓	✓	✓

References

- [1] Is Homophily a Necessity for Graph Neural Networks?. In *International Conference on Learning Representations*, 2022.
- [2] Revisiting heterophily for graph neural networks. *Advances in neural information processing systems*, 35, 1362-1375.



Code



Paper

