Appendix for "**All Roads Lead to Rome: Exploring Edge Distribution Shifts for Heterophilic Graph Learning**"

Authors: Yi Wang, Changqin Huang, Ming Li, Tingyi Cai, Zhonglong Zheng, Xiaodi Huang

IJCAI 2025

Appendix A: Training Details

Datasets

All experiments are conducted on both homophilic and heterophilic graphs, with the relevant data described as follows:
Homophilic Datasets: The homophily ratio of a dataset ranges from 0.5 to 1, where $\mathcal{H} \approx 1$ indicates strong homophily. Three popular benchmarks (Welling and Kipf 2017; Xu et al. 2019), such as Cora, Citeseer, and PubMed, are representative homophilic graphs. In these citation networks, edges denote citation relationships between papers, with node features represented by bag-of-words vectors for each paper. The labels associated with the nodes indicate the respective fields of study.
Heterophilc Datasets: In contrast to homophily datasets, heterophily datasets display homophily ratios ranging from strong heterophily ($\mathcal{H} \approx 0$) up to 0.5. Heterophily is a relatively common trait in real-world datasets, including webpage networks and actor co-occurrence networks.

- **Webpage Networks:** Texas, Wisconsin, Cornell, Chameleon, and Squirrel are five web datasets where nodes represent web pages and edges denote hyperlinks between them. Texas, Wisconsin, and Cornell are subsets of the WebKB[1] dataset, collected by Carnegie Mellon University from the computer science departments of various universities. These web pages are manually categorized into five classes: student, project, course, staff, and faculty. Chameleon and Squirrel, on the other hand, are page-to-page networks on specific topics from Wikipedia, where web pages are classified into five categories based on their average monthly traffic.
- **Actor Co-occurrence Networks[2]:** The actor co-occurrence network is an actor-induced subgraph of a movie-director-writer network. Each node represents an actor, and an edge between two nodes indicates co-occurrence on the same Wikipedia page. Node features correspond to certain keywords on the Wikipedia pages. Five categories are classified based on words from the actors' Wikipedia pages.

Competitors

We evaluate the performance of H_2OGNN by comparing it with graph-agnostic method MLP and popluar GNNs.

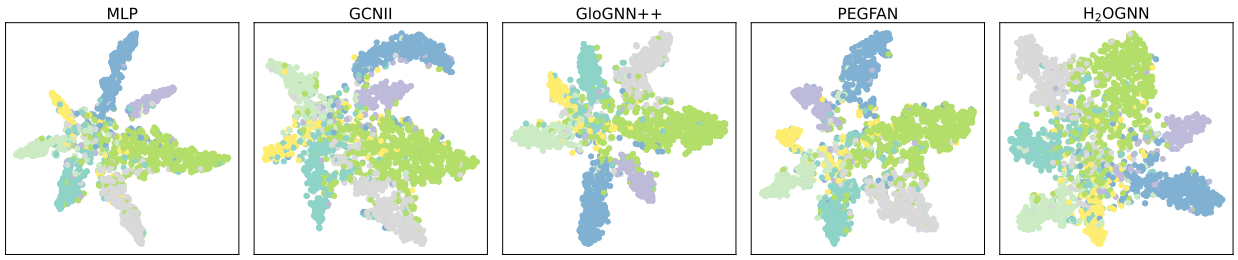
- MLP (Popescu et al. 2009) is a graph-agnostic approach, and its performance is representive as a benchmark to evaluate the efficacy of Graph Neural Networks (GNNs), which incorporate relational learning.
- GCN (Welling and Kipf 2017) and GAT (Veličković et al. 2018) are typical classical standard GNNs designed for homophily graphs, where message aggregation operations tend to feature smoothing, and make them challenging to learn distinguishable representation on heterophily nodes.
- GCNII (Chen et al. 2020) is a particular classical GNN that has shown improved performance on several heterophily graphs. It retains the initial feature information in the deepened graph neural network, thus alleviating the problem of feature oversmoothing in general GNNs, and performing well on heterophily graphs.
- MixHop (Abu-El-Haija et al. 2019) and H_2GCN (Zhu et al. 2020) are two high-order neighbor mixing models that represent early studies in heterophily-aware GNNs. They employ the straightforward approach of leveraging homophily information from higher-order neighbors to offset the distribution of heterophilic information in lower-order neighborhoods.
- GPR-GNN (Chien et al. 2021), LINKX(Lim et al. 2021), ACM-GCN (Luan et al. 2022), GloGNN++ (Li et al. 2022b), CAGNN (Chen et al. 2024), HopGNN+ (Chen et al. 2023) and PEGFAN (Li et al. 2024) are recent research efforts focused on heterophilic graphs. They aim to design guidance mechanisms that consider the degree of homophily in the message-passing process to mitigate the impact of heterophilic information on updated representations.

Hyperparameters Selection

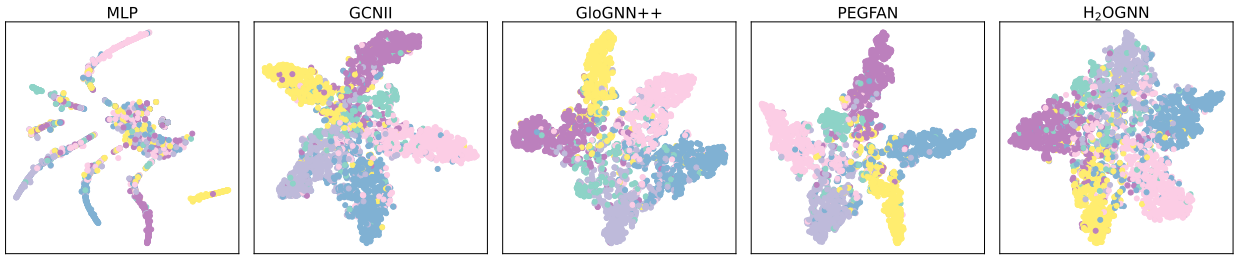
We implement the H_2OGNN using Pytorch (Paszke et al. 2017) and geometric deep learning library Torch-Geometric (Fey and Lenssen 2019). The detailed experimental settings on each benchmarks are as follows:

- **Cora**: layers number is 16, hidden dimensions is 128, learning rate is 1e-2, weight decay is 5e-3, dropout is 0.7.
- **Citeseer**: layers number is 1, hidden dimensions is 64, learning rate is 1e-3, weight decay is 5e-4, dropout is 0.7.
- **Pubmed**: layers number is 64, hidden dimensions is 512, learning rate is 1e-2, weight decay is 5e-4, dropout is 0.2.
- **Actor**: layers number is 1, hidden dimensions is 32, learning rate is 5e-3, weight decay is 5e-3, dropout is 0.2.

[1] https://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wwkb
[2] https://www.aminer.org/lab-datasets/soinf/



(a) Cora



(b) Citeseer

Fig. S-1: Visualization of the deep aggregation feature in several baselines and H_2OGNN on homophilic graphs.

- **Cornell**: layers number is 4, hidden dimensions is 128, learning rate is 1e-2, weight decay is 5e-3, dropout is 0.2.
- **Texas**: layers number is 16, hidden dimensions is 64, learning rate is 1e-2, weight decay is 5e-4, dropout is 0.7.
- **Wisconsin**: layers number is 4, hidden dimensions is 256, learning rate is 1e-2, weight decay is 5e-4, dropout is 0.7.
- **Chameleon**: layers number is 1, hidden dimensions is 512, learning rate is 1e-2, weight decay is 5e-4, dropout is 0.1.
- **Squirrel**: layers number is 1, hidden dimensions is 512, learning rate is 1e-2, weight decay is 5e-4, dropout is 0.0.

Evaluation Metric

Accuracy is employed to evaluate the performance of models. It measures the proportion of the number of classified samples to the total number of samples. It can be expressed as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{S-1}$$

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively.

Appendix B: Supplementary Experiments

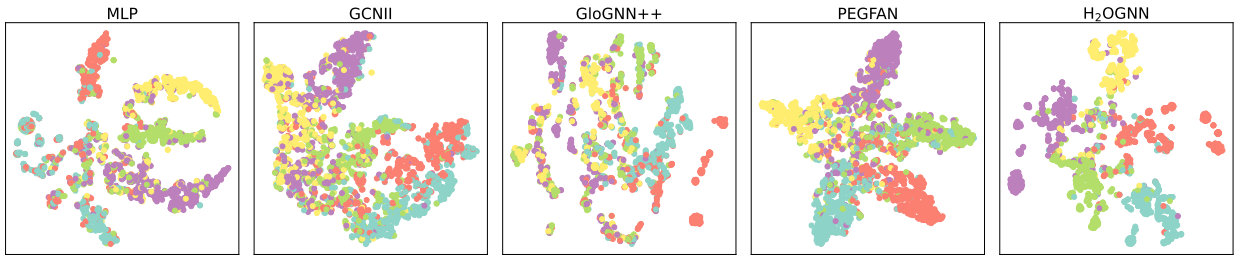
Feature visualization serves as compelling evidence for evaluating the effectiveness of representation learning. Here, we provide more comprehensive visualizations of deep aggregated features in Figs. S-1 and 2, covering both homophilic and heterophilic graphs. The results demonstrate that on heterophilic graphs, the aggregation performance of the classic GCNII algorithm is inferior to that of MLP, while other heterophilic GNNs generally outperform MLP.

Appendix C: Supplementary Explanation

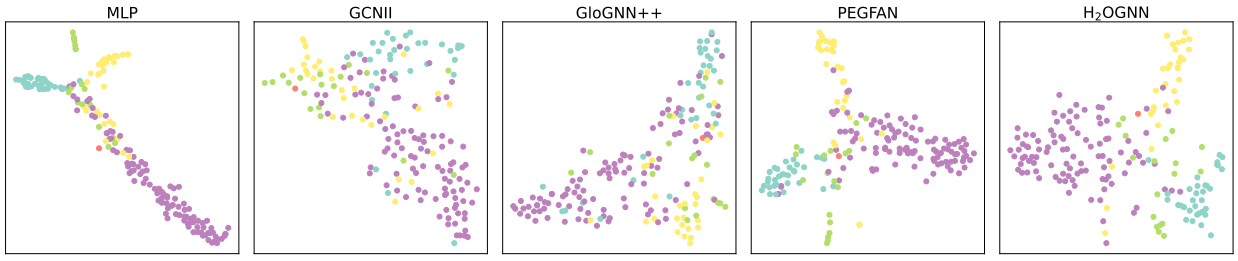
Design Motivation

This section aims to reveal the benefits of heterophilic graph representation learning under the perspective of out-of-distribution (OOD) detection, using both intuitive and statistical analysis. Additionally, based on the analysis results, we introduce a new problem formulation to guide the design of the subsequent framework.

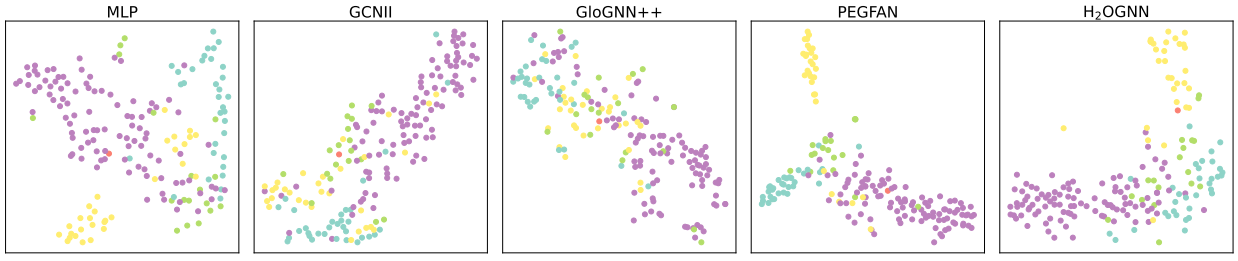
Intuitive Insights. Propagation is a fundamental operation that transmits information along the edges of graphs in GNNs (Welling and Kipf 2017; Garcia Duran and Niepert 2017). Numerous studies (Bo et al. 2021; Li et al. 2022b; Luan et al. 2023; Huang et al. 2024) have demonstrated that heterophilic edges will hinder the effective feature representation learning due to the differentiated message passing along these edges. Therefore, it is essential to design an appropriate propagation method that enhances the relevant information transmission. Typically, this relevant knowledge is believed to propagate solely along homophilic edges. However, many real-world examples support the view that heterophilic edges can also derived from common knowledge shared between nodes. For instance, in citation networks, a classic paper (such as GCN) is often cited by other papers from different fields due to its strong foundational technology, with the shared knowledge between these papers



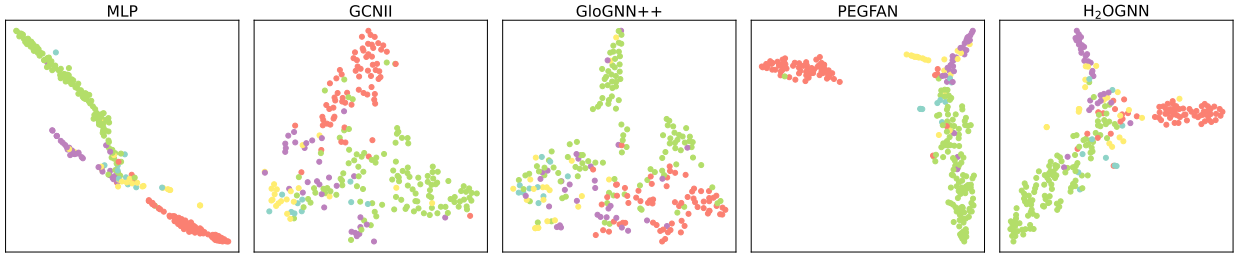
(a) Chameleon



(b) Cornell



(c) Texas



(d) Wisconsin

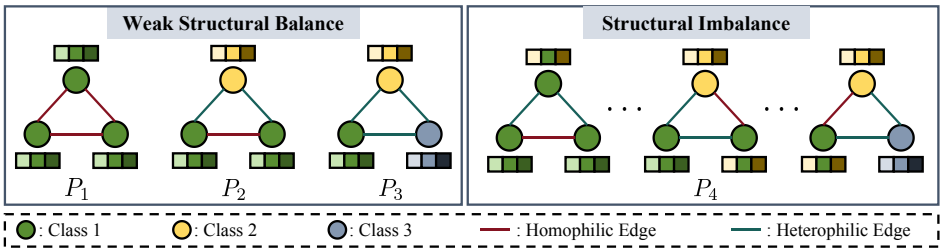
Fig. S-2: Visualization of the deep aggregation feature in several baselines and H_2OGNN on heterophilic graphs.

often implicitly encoded in their features. Based on this observation, we formalize this phenomenon as Hypothesis 1, which supports our subsequent problem formulation and theoretical analysis. Additionally, building on this hypothesis, we analyze several possible connection patterns that may coexist within real-world data, as illustrated in Fig. S-3 (a).

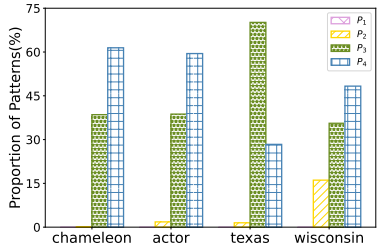
Hypothesis 1 *Two dissimilar nodes in a graph are often connected due to similarities in certain knowledges, which may also be relevant to the linking patterns of some similar nodes.*

However, the diverse edge patterns present in the graph structure pose challenges to current edge attribute identification methods from multiple perspectives, including generalization, imbalance, etc. This challenge arises because these methods primarily rely on node feature similarity calculations, failing to capture the intrinsic knowledge of edge attributes. Our intuitive insight is on that this issue can be addressed through modeling identifiable attribute distributions. There remains a significant gap in existing research on heterophilic graph representation learning regarding how to learn edge attribute distributions, and this paper aims to fill this gap.

Emperical Findings. Guided by these intuitive insights, we further analyzed different structure patterns from the perspective of structural balance theory (Cartwright and Harary 1956; Liang et al. 2023), revealing the challenges face in handling diverse structures over heterophilic graphs. In general, a meaningful graph data structure should adhere to the structural weak



(a) Undirected triadic structure patterns in a graph.



(b) Statistical results of structural patterns.

Fig. S-3: Guided by appropriate class features, symbolic P_1, P_2, and P_3 align with weak structural balance theory, allowing their label relationships to be reconstructed using existing methods. However, we find that triadic structures in most heterophilic graphs are unbalanced, aligning more closely with real-world assumptions (see Hypothesis 1) but also posing greater challenges for edge identification.

balance theory [3]. We defined and categorized potential structural patterns in the graph according to weak balance theory, as shown in Fig. S-3 (a). Based on the fundamental assumptions of most existing research, the three patterns that conform to weak structural balance should rely on consistent features for similar nodes and inconsistent features for dissimilar nodes. However, when the features between dissimilar nodes exhibit similarity, their structural patterns become unbalanced. Under current basic assumptions, we find that most structural patterns in heterophilic graphs are unbalanced, as illustrated in Fig. S-3 (b). Existing assumptions are incomplete, posing challenges for edge attribute identification. In the real world, many dissimilar nodes connect due to common knowledge, which is often implicitly reflected in their node features, as described in Hypothesis 1. The introduction of Hypothesis 1 challenges existing methods, highlighting the urgent need to design a new edge identification framework to model the intrinsic attributes of edges.

Related Work

Heterophilic Graph Neural Networks. Heterophilic graph neural networks (GNNs) (Abu-El-Haija et al. 2019; Li et al. 2022b, 2024) have garnered widespread attention due to their stronger generalization capabilities. Compared to standard GNNs research (Welling and Kipf 2017; Veličković et al. 2018; Chen et al. 2020), heterophilic GNNs are required to perform well on both homophilic and heterophilic graphs. To achieve this, most existing studies (Zhu et al. 2020; Chien et al. 2021; Chen et al. 2023) have converged on a common strategy: implementing homophilic/heterophilic edge attribute-aware message passing. As a result, researchers (Abu-El-Haija et al. 2019; Li et al. 2022b; Bo et al. 2021; Li et al. 2024) have made extensive efforts to design mechanisms for recognizing homophilic and heterophilic edge attribute over local neighborhoods, which can be categorized into two types: structural-based and feature-based methods. In general, structure-based methods (Abu-El-Haija et al. 2019; Zhu et al. 2020; Li et al. 2022b) aim to enhance local message passing under edge attribute-aware by exploring homophilic attribute neighbors. These methods extend local neighborhoods by capturing potential homophilic neighbors from higher-order neighborhoods. For example, MixHop (Abu-El-Haija et al. 2019) and H_2GCN (Zhu et al. 2020) allow ego nodes to receive latent representations from both their local 1-hop neighbors and distant k-hop neighbors. GPR-GNN (Chien et al. 2021), GloGNN++ (Li et al. 2022b), and HopGNN+ (Chen et al. 2023) re-establish local neighborhoods by exploring local homophily over the entire graph topology. Feature-based methods (Bo et al. 2021; Luan et al. 2022; Li et al. 2024) focus on GNNs design in both the spectral and spatial domains. The core intuition is that low-pass filters retain commonality (often obtained from neighbors connected with homophilic edges) in node features, while high-pass filters capture differences (often obtained from neighbors connected with heterophilic edges). These approaches employ an adaptive combination of low-pass and high-pass filters, either from the graph signal perspective (e.g., FAGCN (Bo et al. 2021), AutoGCN (Wu et al. 2023), PEGFAN (Li et al. 2024)) or through feature decomposition (e.g., ACM-GCN (Luan et al. 2022)), aiming to enhance homophilic connections and weaken heterophilic connections in the message passing process.

Although existing methods perform well on heterophilic graphs, we find that many of them experience a significant performance decline on graphs with high heterophily. This paper investigates the reasons for this performance decline, suggesting that it may stem from semantic biases in edge connections between training and testing phases. These biases are caused by the complex and diverse edge connectivity patterns, which in turn challenge the stability of message passing.

Out-of-Distribution Learning on Graphs. Out-of-distribution (OOD) learning addresses situations where training and testing data follow semantic shift. Recently, OOD learning on graphs has emerged as a new research field (Li et al. 2022a; Fan et al. 2024; Liu et al. 2023). Generally, it can be divided into two types: OOD generalization and OOD detection. The former aims to train more robust GNNs that can generalize better to new class samples (i.e., OOD data), while the latter focuses on enabling

[3]The weak balance theory states fundamental principles for specific patterns: "*an enemy of my friend is my enemy*" (e-f-e), "*a friend of my friend is my friend*" (f-f-f), and "*a friend of my enemy is my enemy*" (f-e-e).

GNNs to effectively identify OOD data. Due to different tasks, these two approaches impose distinct training constraints on GNNs. OOD generalization (Li et al. 2022a; Yu, Liang, and He 2023; Li et al. 2023; Fan et al. 2024) seeks to make GNNs more inclusive by learning a unified distribution that can cover all classes, often through techniques like data augmentation (Yu, Liang, and He 2023), domain adaptation (Li et al. 2023), or other learning paradigms (Fan et al. 2024) to improve model generalization. In contrast, OOD detection (Song and Wang 2022; Li et al. 2022c; Liu et al. 2023) aims to enhance GNNs' ability to distinguish between different distributions, focusing on modeling distribution differences. This is often achieved by using probability estimation (Li et al. 2022c), distance metrics (Song and Wang 2022), or feature space distribution properties (Liu et al. 2023) to widen the gap between OOD and in-distribution (IND) data.

The reliability of message passing along edge attributes can be ensured by incorporating OOD detection strategies. This paper aims to achieve intrinsic edge attribute identification by treating heterophilic edges as OOD samples and homophilic edges as IND samples, with the support of OOD detection. However, it is important to note that when integrating this strategy into the overall objective of graph representation learning, semantic shifts between training and testing still exist within homophilic edges. Therefore, in addition to ensuring the effective detection of OOD samples, it is also necessary to preserve the generalization ability of IND samples. Existing OOD detection methods typically focus on maximizing the distinction between IND and OOD samples, which leads to certain limitations. Based on this, this paper proposes a novel and effective OOD detection model to ensure an appropriate classification boundary between the attribute distributions of homophilic and heterophilic edges, thereby aligning the OOD detection strategy more effectively with the task of heterophilic graph representation learning.

Appendix D: Theoretical Properties of H_2OGNN

Theoretical Soundness for Theorem 1

Let $\mathbf{H}$ be the result of mapping $\mathbf{X}$ as given in Eq. (8) in manuscript. Hypothesis 1 inspires us to define $\mathbf{A}_{+,(u,v)}$ and $\mathbf{A}_{-,(u,v)}$ as the dual facets for linking nodes with shared knowledge across homophilic and heterophilic edges. Incorporating these constraints, we derive a linear relation $\mathbf{A}_{+,(u,v)} - \mathbf{A}_{-,(u,v)} = 1$ in the scenario where $\mathbf{A}_{(u,v)} = 1$. Solving this relation, we can express both $\mathbf{A}_{+,(u,v)}$ and $\mathbf{A}_{-,(u,v)}$ in attribute distribution $\mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}|\mathcal{G})$ and $\mathbb{E}_{(u,v)\in\mathcal{E}_{out}} p_\theta(r_{(u,v)}|\mathcal{G})$ of the known homophilic edges and heterophilic edges over graph $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$, i.e.,

$$\mathbf{A}_{+,(u,v)} \sim \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}|\mathcal{G}),\ \mathbf{A}_{-,(u,v)} \sim \mathbb{E}_{(u,v)\in\mathcal{E}_{out}} p_\theta(r_{(u,v)}|\mathcal{G}). \tag{S-2}$$

Lemma 1 ensures the distinguishability of the distributions for both homophilic and heterophilic attributes. Thus far, the optimization problem presented in Eq. (10) in manuscript is only composed of variables $\mathbf{X}$ and $\mathbf{H}$. Solving it directly remains a challenge, as the non-linear operator introduced in p_θ lead to a complex differentiation process.

To tackle this challenge, we can reframe the problem by decoupling the learning of $\mathbf{A}_{+,(u,v)}$ and $\mathbf{A}_{-,(u,v)}$ from the main optimization objective and adopt a staged learning approach. Suppose we have obtained the node features at the l-th iteration, denoted as $\mathbf{H}^{(l)}$. In the first phase, we can calculate $\mathbf{A}^{(l+1)}_{+,(u,v)}$ and $\mathbf{A}^{(l+1)}_{-,(u,v)}$ using $\mathbf{H}^{(l)}$ with Eq. (6) in manuscript, which effectively corresponds to our OOD detecter described in Section 3.2. In the second stage, by incorporating the calculated values of $\mathbf{A}^{(l+1)}_{+,(u,v)}$ and $\mathbf{A}^{(l+1)}_{-,(u,v)}$, we can decompose the Laplacian regularization of the original optimization problem into two distinct components:

$$\arg\min_{\mathbf{H}} \|\mathbf{H} - \mathbf{X}\|_2^2 + \xi \cdot tr\left(\mathbf{H}^\top \mathbf{L}_+ \mathbf{H}\right) + \xi \cdot tr\left(\mathbf{H}^\top \mathbf{L}_- \mathbf{H}\right), \tag{S-3}$$

where $\mathbf{L}_+ = \mathbf{D}_+ - \mathbf{A}_+$ and $\mathbf{L}_- = \mathbf{D}_- + \mathbf{A}_-$ are fixed values. Consequently, iteratively stacking OOD detector and message-passing layers is tantamount to the aforementioned alternative learning approach for addressing the attribute-aware graph denoising problem outlined in Eq. (10) in manuscript. By employing the Adam algorithm (Kingma and Ba 2014), we minimize both the prediction loss and the energy distribution regularization $\mathcal{L}_{reg}$ as defined in Eq. (7) in manuscript. This method not only assigns concrete significance to attribute-aware learning but also ensures the convergence of our proposed alternative learning process.

Theoretical Soundness for Lemma 1

We denote the conditional likelihood $p_\theta(\mathcal{G}|r_{(u,v)} = 1)$ according to EBM as:

$$p_\theta(\mathcal{G}|r_{(u,v)} = 1) = \frac{e^{-E(\mathcal{G};f_\theta)}}{\int_{\mathcal{G}|r_{(u,v)}} e^{-E(\mathcal{G};f_\theta)}}, \tag{S-4}$$

and $Z_{r_{(u,v)}} = \int_{\mathcal{G}|r_{(u,v)}} e^{-E(\mathcal{G};f_\theta)}$ is the normalized densities with respect to $\mathcal{G}$ and f_θ. We subsequently reformulate energy scores of homophilic edges (i.e., in-distribution data) from our OOD detector expression into a likelihood format:

$$E(\mathcal{G}; f_\theta) = -\log\left(\mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G}|r_{(u,v)} = 1) \cdot Z_{r_{(u,v)}}\right), \tag{S-5}$$

$$= -\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G}|r_{(u,v)} = 1) \underbrace{- \log Z_{r_{(u,v)}}}_{Z}. \tag{S-6}$$

Afterwards, by applying the Bayesian theorem transformation for term $\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G}|r_{(u,v)}=1)$, we obtain:

$$E(\mathcal{G}; f_\theta) = -\log \frac{\mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1|\mathcal{G}) \cdot \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G})}{\mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1)} + Z, \tag{S-7}$$

$$= -\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1|\mathcal{G}) \; - \; \log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G}) + \underbrace{Z + \log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1))}_{C}, \tag{S-8}$$

$$= \underbrace{-\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1|\mathcal{G})}_{\downarrow \text{ for homophilic edges}} \underbrace{- \log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G})}_{\downarrow \text{ for homophilic edges}} + C. \tag{S-9}$$

References

Abu-El-Haija, S.; Perozzi, B.; Kapoor, A.; Alipourfard, N.; Lerman, K.; Harutyunyan, H.; Ver Steeg, G.; and Galstyan, A. 2019. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *ICML*, 21–29.

Bo, D.; Wang, X.; Shi, C.; and Shen, H. 2021. Beyond low-frequency information in graph convolutional networks. In *AAAI*, 3950–3957.

Cartwright, D.; and Harary, F. 1956. Structural balance: A generalization of Heider's theory. *Psychological Review*, 63(5): 277.

Chen, J.; Chen, S.; Gao, J.; Huang, Z.; Zhang, J.; and Pu, J. 2024. Exploiting neighbor effect: Conv-agnostic GNN framework for graphs with heterophily. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10): 13383–13396.

Chen, J.; Li, Z.; Zhu, Y.; Zhang, J.; and Pu, J. 2023. From node interaction to hop interaction: New effective and scalable graph learning paradigm. In *CVPR*, 7876–7885.

Chen, M.; Wei, Z.; Huang, Z.; Ding, B.; and Li, Y. 2020. Simple and deep graph convolutional networks. In *ICML*, 1725–1735.

Chien, E.; Peng, J.; Li, P.; and Milenkovic, O. 2021. Adaptive universal generalized pagerank graph neural network. In *ICLR*.

Fan, S.; Wang, X.; Shi, C.; Cui, P.; and Wang, B. 2024. Generalizing graph neural networks on out-of-distribution graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1): 322–337.

Fey, M.; and Lenssen, J. E. 2019. Fast graph representation learning with PyTorch Geometric. In *RLGM Workshop at ICLR*.

Garcia Duran, A.; and Niepert, M. 2017. Learning graph representations with embedding propagation. In *NIPS*, 5125–5136.

Huang, C.; Wang, Y.; Jiang, Y.; Li, M.; Huang, X.; Wang, S.; Pan, S.; and Zhou, C. 2024. Flow2GNN: Flexible two-way flow message passing for enhancing GNNs beyond homophily. *IEEE Transactions on Cybernetics*, 54(11): 6607–6618.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, H.; Wang, X.; Zhang, Z.; and Zhu, W. 2022a. Out-of-distribution generalization on graphs: A survey. *arXiv preprint arXiv:2202.07987*.

Li, H.; Wang, X.; Zhang, Z.; and Zhu, W. 2023. OOD-GNN: Out-of-distribution generalized graph neural network. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7328–7340.

Li, J.; Zheng, R.; Feng, H.; Li, M.; and Zhuang, X. 2024. Permutation equivariant graph framelets for heterophilous graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9): 11634–11648.

Li, X.; Zhu, R.; Cheng, Y.; Shan, C.; Luo, S.; Li, D.; and Qian, W. 2022b. Finding global homophily in graph neural networks when meeting heterophily. In *ICML*, 13242–13256.

Li, Z.; Wu, Q.; Nie, F.; and Yan, J. 2022c. Graphde: A generative framework for debiased learning and out-of-distribution detection on graphs. In *NeurIPS*, 30277–30290.

Liang, L.; Hu, X.; Xu, Z.; Song, Z.; and King, I. 2023. Predicting global label relationship matrix for graph neural networks under heterophily. In *NeurIPS*, 10909–10921.

Lim, D.; Hohne, F.; Li, X.; Huang, S. L.; Gupta, V.; Bhalerao, O.; and Lim, S. N. 2021. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. In *NeurIPS*, 20887–20902.

Liu, Y.; Ding, K.; Liu, H.; and Pan, S. 2023. Good-d: On unsupervised graph out-of-distribution detection. In *ICDM*, 339–347.

Luan, S.; Hua, C.; Lu, Q.; Zhu, J.; Zhao, M.; Zhang, S.; Chang, X.-W.; and Precup, D. 2022. Revisiting heterophily for graph neural networks. In *NeurIPS*, 1362–1375.

Luan, S.; Hua, C.; Xu, M.; Lu, Q.; Zhu, J.; Chang, X.-W.; Fu, J.; Leskovec, J.; and Precup, D. 2023. When do graph neural networks help with node classification? Investigating the homophily principle on node distinguishability. In *NeurIPS*, 28748–28760.

Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch. In *Autodiff Workshop at NeurIPS*.

Popescu, M.-C.; Balas, V. E.; Perescu-Popescu, L.; and Mastorakis, N. 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7): 579–588.

Song, Y.; and Wang, D. 2022. Learning on graphs with out-of-distribution nodes. In *KDD*, 1635–1645.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph attention networks. In *ICLR*.

Welling, M.; and Kipf, T. N. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2023. Beyond low-pass filtering: Graph convolutional networks with automatic filtering. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 6687–6697.

Xu, B.; Shen, H.; Cao, Q.; Qiu, Y.; and Cheng, X. 2019. Graph wavelet neural network. In *ICLR*.

Yu, J.; Liang, J.; and He, R. 2023. Mind the label shift of augmentation-based graph OOD generalization. In *CVPR*, 11620–11630.

Zhu, J.; Yan, Y.; Zhao, L.; Heimann, M.; Akoglu, L.; and Koutra, D. 2020. Beyond homophily in graph neural networks: Current limitations and effective designs. In *NeurIPS*, 7793–7804.