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

Yi Wang[1], **Changqin Huang**[2,1*], **Ming Li**[1], **Tingyi Cai**[3], **Zhonglong Zheng**[3], **Xiaodi Huang**[4]
[1]Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University
[2]College of Education, Zhejiang University
[3]School of Computer Science and Technology, Zhejiang Normal University
[4]School of Computing Mathematics and Engineering, Charles Sturt University
wangyi@zjnu.edu.cn, cqhuang@zju.edu.cn, {mingli, tingyicai, zhonglong}@zjnu.edu.cn, xhuang@csu.edu.au

Abstract

Heterophilic graph neural networks (GNNs) have gained prominence for their ability to learn effective representations in graphs with diverse, attribute-aware relationships. While existing methods leverage attribute inference during message passing to improve performance, they often struggle with challenging heterophilic graphs. This is due to edge distribution shifts introduced by diverse connection patterns, which blur attribute distinctions and undermine message-passing stability. This paper introduces H_2OGNN, a novel framework that reframes edge attribute inference as an out-of-distribution (OOD) detection problem. H_2OGNN introduces a simple yet effective symbolic energy regularization approach for OOD learning, ensuring robust classification boundaries between homophilic and heterophilic edge attributes. This design significantly improves the stability and reliability of GNNs across diverse connectivity patterns. Through theoretical analysis, we show that H_2OGNN addresses the graph denoising problem by going beyond feature smoothing, offering deeper insights into how precise edge attribute identification boosts model performance. Extensive experiments on nine benchmark datasets demonstrate that H_2OGNN not only achieves state-of-the-art performance but also consistently outperforms other heterophilic GNN frameworks, particularly on datasets with high heterophily.

1 Introduction

Heterophilic graph neural networks (GNNs) [Luan *et al.*, 2024; Ma *et al.*, 2022; Li *et al.*, 2023b; Gong *et al.*, 2024] have garnered significant attention due to their ability to process both homophilic and heterophilic graphs. Unlike traditional GNNs, such as GCN [Welling and Kipf, 2017], GAT [Veličković *et al.*, 2018], and GCNII [Chen *et al.*, 2020], heterophilic GNNs have demonstrated remarkable performance in real-world tasks, particularly when handling structured data with heterophilic attributes.

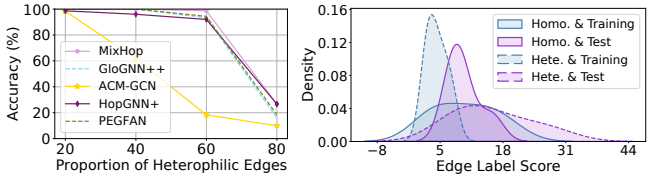

Figure 1: The left presents a trend graph that demonstrates how the performance of existing methods significantly declines as the number of heterophilic edges increases. The inflection point occurs approximately when the ratio of heterophilic edges exceeds 60%, with performance reaching its lowest point when the ratio hits or exceeds 80%. The right further investigates a potential cause for this decline, revealing a shift in edge label semantics that leads to confusion in the distribution. Insignificant differences in the edge-attribute distributions between the training and testing stages on graphs with more than 80% heterophilic edges may contribute significantly to the substantial performance drop observed in current methods.

A central challenge in the development of heterophilic GNNs is effectively integrating graph structure and node features to infer edge attributes and enhance the propagation of local homophilic features[Abu-El-Haija *et al.*, 2019; Zhu *et al.*, 2020; Li *et al.*, 2022; Chen *et al.*, 2024; Bi *et al.*, 2024]. Enhancing local homophilic message passing has thus emerged as a fundamental objective. For instance, methods like MixHop [Abu-El-Haija *et al.*, 2019] and H_2GNN [Zhu *et al.*, 2020] address this challenge by aggregating information from higher-order neighbors with similar attributes, improving the transmission of homophilic features. Other approaches, such as GloGNN++ [Li *et al.*, 2022], CAGNN [Chen *et al.*, 2024] and DHGR [Bi *et al.*, 2024], have designed guidance mechanisms based on node-level features to directly estimate edge attributes in local structures. These mechanisms enhance the perception and propagation of homophilic features during message passing. Recent advancements have expanded the scope of heterophilic GNNs by exploring diverse learning paradigms. For exam-

*Corresponding Author (cqhuang@zju.edu.cn)

ple, heterophilic GNNs based on contrastive learning [He *et al.*, 2023; Li *et al.*, 2023a; Wang *et al.*, 2024] leverage the power of representation learning to strengthen local message passing. Similarly, spectral domain appraoches on heterophilic GNNs [Bo *et al.*, 2021; Luan *et al.*, 2022; Li *et al.*, 2024] provide new insights from a frequency-domain perspective.

Although these methods have shown success, they primarily rely on node-level representations to infer edge attributes, which poses significant challenges in scenarios involving high heterophily. As the proportion of heterophilic edges increases, the graph's connectivity patterns become more complex and diverse. This complexity disrupts the stability and reliability of message passing, resulting in deteriorated performance, particularly in graphs with a higher proportion of heterophilic edges. As shown in Figure 1(a), the performance of these algorithms significantly declines as the number of heterophilic edges increases. Upon further investigation, we identify the underlying cause of this performance degradation: *insignificant distribution shift*. Figure 1(b) illustrates how graphs with a high ratio of heterophilic edges exhibit unclear distinctions in the disbrituion of homophilic and heterophilic edge semantics between training and testing phases. This unclear distribution shift arises due to the increased diversity in the classes of connected nodes, making it difficult to compress edge semantics, determined by node representations, into a binary attribute space (homophily vs. heterophily).

To address this limitation, we propose a novel approach that shifts focus from node-level to edge-level representations. However, distinguishing between homophilic and heterophilic edge in the presence of insignificant attribute distribution shifts remains a significant challenge. We frame this problem as an Out-of-Distribution (OOD) detection issue, where OOD samples, which differ in distribution from in-distribution (IND) data, must be identified [Song and Wang, 2022; Liu *et al.*, 2023; Cai *et al.*, 2025b]. By reframing the problem in this way, we can design more reliable and stable message-passing paradigms that can handle heterophilic graphs more effectively.

In summary, this paper introduces a novel heterophilic GNN framework, H_2OGNN, which employs an edge-level OOD detection strategy to improve performance. The main contributions of this paper are as follows.

- **New Problem Formulation:** By defining heterophilic edges as OOD samples, we propose a new problem formulation from the perspective of OOD detection. This formulation lays the foundation for heterophilic graph representation learning, enabling a clearer separation between homophilic and heterophilic edge distributions.
- **New Learning Framework:** We present H_2OGNN, a novel framework that incorporates a signed energy regularization OOD method for inferring edge-level attribute semantics. This method improves the generalization of GNNs across diverse connection patterns and establishes clear boundaries between homophilic and heterophilic edge attributes. As a result, H_2OGNN ensures robust edge attribute identification, making it highly effective for a variety of graph learning tasks.
- **Theoretical Insight:** To further validate H_2OGNN, we provide a theoretical proof that it is equivalent to solving the graph denoising problem with signed smoothing. Additionally, we offer an enhanced explanation of the graph denoising process from the OOD detection perspective, which deepens our understanding of the underlying mechanism of H_2OGNN.

2 Preliminaries

Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a vertex set $\mathcal{V}$ with $N = |\mathcal{V}|$ vertices, and an edge set $\mathcal{E}$ with $M = |\mathcal{E}|$ edges. The vertex feature matrix is denoted by $\mathbf{X} \in \mathbb{R}^{N \times m}$. Let $\mathbf{Y} \in \mathbb{R}^N$ denote the labels. The fundamental task of node classification is to find the optimal mapping g_θ^* from $\mathbf{X}$ to $\mathbf{Y}$ such that $\mathbf{Y} = g_\theta^*(\mathbf{X})$, where g_θ^* is typically fitted by a GNNs model with parameters θ. A key aspect of this task involves designing an effective message-passing (MP) scheme to generate node representations that align with the task's requirements. Traditional MP methods often struggle to aggregate useful information effectively, especially in the presence of numerous heterophilic edges in heterophilic graphs. For heterophilic graphs, it is crucial to differentiate between homophilic and heterophilic connections. In the following, we introduce the definition of homophilic/heterophilic edges (See Definition 1) along with commonly used quantification metrics for the homophily ratio of edges (See Definition 2).

Definition 1. *(Homophilic/Heterophilic Edge) A **homophilic** edge is a link between two nodes that share the same type, while **heterophilic edge** connects two nodes of different types.*

Edge Homophily/Heterophily. The homophily of a graph, from the perspective of its edges, is typically measured as the proportion of homophilic edges. Conversely, the heterophily of a graph is calculated by subtracting its homophily from 1. For simplicity, only edge homophily is considered in the subsequent section. The specific calculation method is given in Definition 2.

Definition 2. *The edge homophily ($\mathcal{H}$) on a graph is defined as:*

$$\mathcal{H} = \frac{|\{(v,u) \in \mathcal{E} : y_v = y_u\}|}{|\mathcal{E}|}, \tag{1}$$

where y_v and y_u are the labels of nodes v and u, respectively.

3 Methods

This section provides an in-depth analysis of our proposed framework, H_2OGNN. It includes the problem formulation (see Section 3.1), a detailed explanation of the framework (see Section 3.2), and its theoretical underpinnings (see Section 3.3). Further insights into the design motivation are provided in **Appendix C**[1].

3.1 Problem Formulation

We aim to model the attribute distribution of edges and propose a novel problem formulation for representation learning

[1]https://kellysylvia77.github.io/H2OGNN/Appendix.pdf

on heterophilic graphs from an OOD detection perspective, as described in Problem 1. Additionally, we incorporate the more challenging and realistic Hypothesis 3.1 into the model design.

Problem 1. *(OOD Detection Perspective) Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, a label set $\mathcal{Y} = \{y_v\}_{v\in\mathcal{V}}$ and a training set $\mathcal{D} = \{\mathbf{x}_v, y_v\}_{v\in\mathcal{V}_{train}}$ drawn from $\mathcal{G}$, from which the known homophilic edge set $\mathcal{E}_{in}$ and heterophilic edge set $\mathcal{E}_{out}$ on $\mathcal{G}$ can be determined. The objective is to jointly optimize an edge attribute discriminator p^*_θ and a graph predictor g^*_ϕ that achieves optimal generalization representation under significant differences in attribute distributions, i.e., $p(\mathcal{E}_{in}) \neq p(\mathcal{E}_{out})$:*

$$p^*_\theta, g^*_\phi = \arg\min_{\theta,\phi} \mathbb{E}_{v\in\mathcal{V}_{train}}[\mathcal{L}(g_\phi(\{p_\theta(u,v)\mathbf{x}_u : u \in \mathcal{N}(v)\}), y_v)], \quad (2)$$

where $p_\theta(u,v)$ denotes the estimated attribute of edge (u,v) and $\mathcal{N}(v)$ is the neighbors set of node v.

***Remark**: From the perspective of OOD detection, it is essential that $p(\mathcal{E}_{in}) \neq p(\mathcal{E}_{out})$. However, several factors complicate the learning of differentiated attribute distributions. These include the limited availability of labeled training data, the challenge of learning edges with similar semantics but different attributes, and the difficulty of handling edges with distant semantics but identical attributes. As a result, designing an optimal edge attribute discriminator p^*_θ becomes the most challenging aspect of this problem formulation.*

Two dissimilar nodes in a graph are often connected due to shared similarities in specific knowledge domains, which may also influence the linking patterns observed among similar nodes.

3.2 The H_2OGNN Framework

Guided by the problem formulation, we introduce a novel framework for heterophilic GNNs. This framework addresses the challenges of learning effective representations for heterophilic graphs while incorporating OOD detection. By unifing generalization and detection, it enhances edge-aware inference, thereby improving representation learning for heterophilic graphs. We name this framework H_2OGNN, as illustrated in Figure 2.

Aligning edge-attribute Inference with OOD Detection to enhance edge-aware learning. Accurate edge-attribute inference plays a crucial role in facilitating effective knowledge propagation during local message passing. We hypothesize that aligning edge-attribute inference with the OOD detection task provides an effective approach. To achieve this, we **first** sample in-distribution (IND) and OOD data from the graph structure based on edges with known attributes, supporting subsequent OOD detection for edge attribute identification. Specifically, We define heterophilic edges as OOD samples and homophilic edges as IND samples, as follows:

$$\mathcal{E}_{ind} = \{(u,v) : y_u = y_v\},\ \mathcal{E}_{out} = \{(u,v) : y_u \neq y_v\}, \quad (3)$$

where (u,v) is sampled from the training data. The edge set $\mathcal{E}$ is then divided into three subsets: $\mathcal{E} = \mathcal{E}_{ind} \cup \mathcal{E}_{out} \cup \mathcal{E}_{un}$, where $\mathcal{E}_{un}$ is the set of edges with unknown attributes. The objective of aligning edge-attribute identification with OOD detection is to model the distribution boundary between samples from $\mathcal{E}_{ind}$ and $\mathcal{E}_{out}$, enabling effective inference of edge attributes in $\mathcal{E}_{un}$. **Next**, a key challenge arises in ensuring the accurate identification of homophilic edges across different attributes while maintaining robust detection of heterophilic edges under the more realistic conditions described in Hypothesis 3.1.

Energy-based modeling for edge attribute distribution to establish boundaries between homophilic and heterophilic edges. To address this challenge, we propose an effective solution: a unified approach to generalization and detection in edge identification. Specifically, we use logits $p_\theta(r_{(u,v)}|\mathcal{G})$ to derive the distribution of edge attributes, enabling direct estimation of the attribute $r_{(u,v)}$ for any edge $(u,v) \in \mathcal{E}$, based on the graph structure $\mathcal{G} = (\mathbf{X}, \mathbf{A})$. Formally, the probability distribution is expressed as:

$$p_\theta(r_{(u,v)}|\mathcal{G}) = \frac{e^{f_\theta(\mathcal{G},r)}}{\sum_{r\prime\in\mathcal{A}} e^{f_\theta(\mathcal{G},r\prime)}}. \quad (4)$$

Here, $\mathcal{A} = \{0,1\}$ denotes an indicator set. $f_\theta(\mathcal{G}, r)$ is a feature transformer that fuses the node features and adjacency matrix into a representation for node-pairs. To further enhance our approach, we seek to train an effective boundary for the edge attribute distribution using energy-based methods. By defining the relationship between the energy function and probability density through an EBM model [Ranzato, 2007; Cai *et al.*, 2025a], we define an energy form $E(\mathcal{G}, r; f_\theta) = -f_\theta(\mathcal{G}, r)$. Additionally, the free energy function $E(\mathcal{G}, r; f_\theta)$ marginalizes over $r_{(u,v)}$ and is given by the denominator in Eq. (4):

$$E(\mathcal{G}, r; f_\theta) = -\log \sum_{r\prime\in\mathcal{A}} e^{f_\theta(\mathcal{G},r')}. \quad (5)$$

It is important to note that the energy function is derived directly from the predicted logits, with no modifications to the parameterization. The non-probabilistic energy score in Eq. (5) incorporates information from the specific edge instance, as well as potential dependencies with other instances based on the observed structures. While optimizing the energy scores effectively models the distribution boundary, ensuring the generalizability of the in-distribution data, specifically for the homophilic edge distribution, remains challenging—a common issue in existing OOD detection methods. To address this, we propose further symbolizing the energy and applying signed energy inference to align information from edges with identical attributes to enhance the generalizability of the in-distribution data. The signed inference rule is as follows:

$$\tilde{E}_{(u,v)} = \mathsf{Sign}(E_{(u,v)} < \tau),\ (u,v) \in \mathcal{E}, \quad (6)$$

where $E_{(u,v)}$ represents the energy function $E(\mathcal{G}, r; f_\theta)$ and $\mathsf{Sign}(\cdot)$ is a signed discriminative function that compares the energy with a threshold τ. This function enforces similarity for edges with the same attributes and differentiation for edges with different attributes.

***Remark**: We introduce hard constraints on the energy gap through a regularization loss $\mathcal{L}_{reg}$, which bounds the energy value for in-distribution data (i.e., labeled instances in $\mathcal{E}_{in}$*

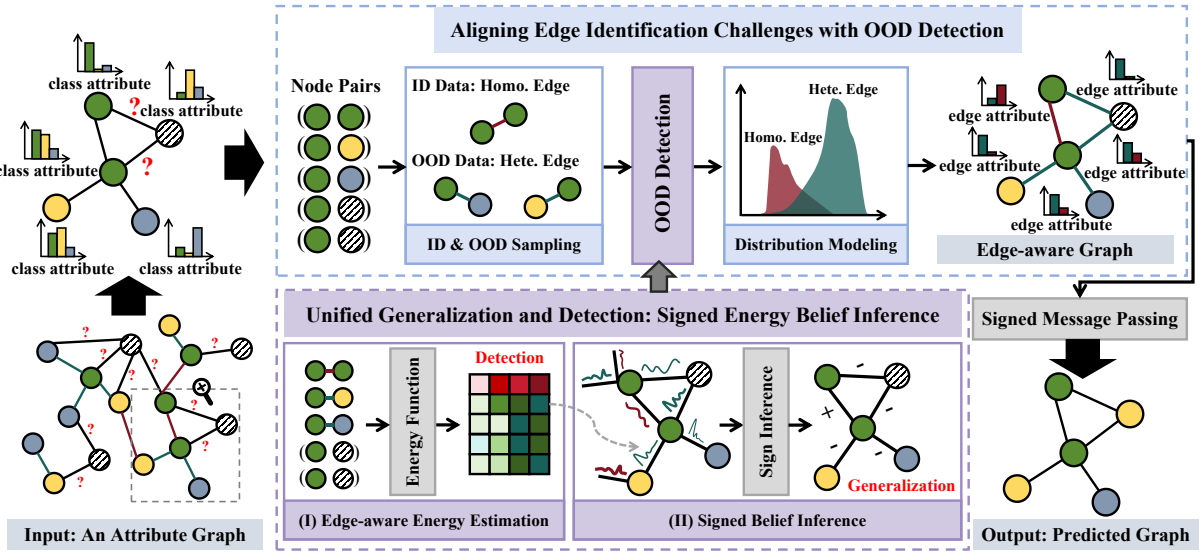

Figure 2: Schematic diagram of H_2OGNN: A unified OOD framework for heterophilic graph representation learning.

and for OOD data (i.e., auxiliary training instances in $\mathcal{E}_{out}$ from a different distribution). The regularization term $\mathcal{L}_{reg}$ offers great design flexibility. We define it as bounding constraints on the signed energy values, as follows:

$$\mathcal{L}_{reg} = \frac{1}{|\mathcal{E}_{ind}|} \sum_{(u,v)\in\mathcal{E}_{ind}} \Big(\mathsf{ReLU}\Big(\tilde{E}(\mathcal{G}, r_{(u,v)}; f_\theta) - \rho_{ind}\Big)\Big)^2 + \frac{1}{|\mathcal{E}_{out}|} \sum_{(u,v)\in\mathcal{E}_{out}} \Big(\mathsf{ReLU}\Big(\rho_{out} - |\tilde{E}(\mathcal{G}, r_{(u,v)}; f_\theta)|\Big)\Big)^2, \quad (7)$$

where bounds the values within the range $[\rho_{ind}, \rho_{out}]$ to decrease for in-distribution samples and increase for out-of-distribution samples. This approach encourages the distinction between homophilic edges and heterophilic edges, while simultaneously pulling together edges with the same attributes.

Signed message passing for effective representation learning. We use signed GNNs that align with the weak balance theory assumption to learn node representations. This not only optimizes the overall objective of graph representation learning but also enhances the performance of edge attribute distribution modeling. Specifically, we employ signed message passing to distinguish between homophily and heterophily attributes. A formal description is provided below:

$$\mathbf{H}_v^{(l+1)} = g_\phi\left(\left\{\tanh(\tilde{E}_{(u,v)}^{(l)})\mathbf{H}_u^{(l)} : u \in \mathcal{N}(v)\right\}\right), \quad (8)$$

where $\mathbf{H}^{(l)}$ represents the feature derived from the l-th layer of message passing, with the specific initial condition $\mathbf{H}^{(0)} = \mathbf{X}$. $g_\phi(\cdot)$ is a signed message passing function. Regardless of the aggregation strategy used in the message-passing process, the primary step in signed message passing is to define a rule for assigning appropriate symbolized coefficients. $\tanh(\cdot)$ further normalizes the energy values, preventing gradient explosion in this model.

***Remark**: The sign inference based on edge identification is reasonable, as it implicitly implements the structural balance theory, similar to the discussion in [Liang* et al.*, 2023], following the idea that "the homophilic neighbor of my heterophilic neighbor is a heterophilic node." This suggests that our strategy can accurately predict the label relationships between direct neighbors, thereby supporting signed message passing to achieve effective representation learning.*

3.3 Theoretical Analysis

Existing research [Ma *et al.*, 2021; Zhu *et al.*, 2021] shows that most standard GNNs can be reduced to a graph information denoising problem, formulated as

$$\arg\min_{\mathbf{H}} \|\mathbf{H} - \mathbf{X}\|_2^2 + \xi \cdot tr\left(\mathbf{H}^\top \mathbf{L}\mathbf{H}\right), \quad (9)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the graph Laplacian matrix, and ξ is a constant coefficient. The term $tr\left(\mathbf{H}^\top \mathbf{L}\mathbf{H}\right)$ represents Laplacian regularization, enforcing feature smoothness between connected nodes. This implies that features of heterophilic nodes become overly smoothed in GNNs, reducing class separability—a key reason why standard GNNs perform poorly on heterophilic graphs, where most connections are between heterophilic nodes. However, in open-world scenarios, heterophilic nodes often connect due to shared knowledge, and smoothing over this knowledge can enhance node representations. Hypothesis 3.1 suggests aggregating shared knowledge along edge attributes, dividing graph filtering into two types: feature smoothing along homophilic edges and differentiated feature smoothing along heterophilic edges. This generalizes the traditional graph signal denoising problem in Eq. (9) as follows.

Theorem 1. *The proposed H_2OGNN is equivalent to tackling the attribute-aware graph denoising probelm, which can be*

Datasets Edge Hom. $\mathcal{H}$	Cora 0.81	Citeseer 0.74	Pubmed 0.80	Actor 0.22	Cornell 0.30	Texas 0.11	Wisconsin 0.21	Chameleon 0.23	Squirrel 0.22	Rank
MLP	75.69±2.00	74.02±1.90	87.16±0.37	36.53±0.70	81.89±6.40	80.81±4.75	85.29±3.31	46.21±2.99	28.77±1.56	11
GCN	86.98±1.27	76.50±1.36	88.42±0.50	27.32±1.10	60.54±5.30	55.14±5.16	51.76±3.06	64.82±2.24	53.43±2.01	13
GAT	87.30±1.10	76.55±1.23	86.33±0.48	27.44±0.89	61.89±5.05	52.16±6.63	49.41±4.09	60.26±2.50	40.72±1.55	14
GCNII	**88.37±1.25**	**77.33±1.48**	**90.15±0.43**	**37.44±1.30**	77.86±3.79	77.57±3.83	80.39±3.40	63.86±3.04	38.47±1.58	6
MixHop	87.61±0.85	76.26±1.33	85.31±0.61	32.22±2.34	73.51±6.34	77.84±7.73	75.88±4.90	60.50±2.53	43.80±1.48	12
H_2GCN	87.87±1.20	77.11±1.57	89.49±0.38	35.70±1.00	82.70±5.28	84.86±7.23	**87.65±4.98**	60.11±2.15	36.48±1.86	8
LINKX	84.64±1.13	73.19±0.99	87.86±0.77	36.10±1.55	77.84±5.81	74.60±8.37	75.49±5.72	68.42±1.38	61.81±1.80	10
GPR-GNN	87.95±1.18	77.13±1.67	87.54±0.38	34.63±1.22	80.27±8.11	78.38±4.36	82.94±4.21	46.58±1.71	31.61±1.24	9
ACM-GCN	87.91±0.95	**77.32±1.70**	**90.00±0.52**	36.28±1.09	85.14±6.07	**87.84±4.40**	**88.43±3.22**	66.93±1.85	54.40±1.88	**3rd**
GloGNN++	**88.33±1.09**	**77.22±1.78**	89.24±0.39	**37.70±1.40**	**85.95±5.10**	84.05±4.90	**88.04±3.22**	**71.21±1.84**	57.88±1.76	**2nd**
CAGNN	87.28±1.01	76.03±1.16	89.74±0.55	35.83±0.73	81.35±5.47	85.13±5.73	82.55±4.17	69.16±1.90	61.82±1.45	7
HopGNN+	87.57±1.33	76.69±1.56	**90.28±0.42**	**37.09±0.97**	84.05±4.48	82.97±5.12	85.69±5.43	**71.21±1.45**	**64.23±1.33**	4
PEGFAN	87.16±1.31	76.92±1.57	89.56±0.30	35.48±0.94	**86.22±4.75**	**86.22±3.30**	86.67±4.28	**80.31±1.10**	**75.06±1.72**	4
H_2OGNN	**88.09±0.85**	77.17±1.25	89.37±0.26	**37.09±0.71**	**86.76±5.85**	**89.73±3.97**	86.08±3.09	**78.82±1.28**	**75.06±2.15**	**1st**

Table 1: Comparative Performance of various GNNs on homophilic graphs and heterophilic graphs. The best-performing model is highlighted in **red**, the second-best in **blue**, and the third-best in **green**.

generalized as follows:

$$\begin{aligned} &\arg\min_{\mathbf{H}} \|\mathbf{H}-\mathbf{X}\|_2^2 + \xi \cdot tr\left(\mathbf{H}^\top \mathbf{L}_+ \mathbf{H}\right) + \xi \cdot tr\left(\mathbf{H}^\top \mathbf{L}_- \mathbf{H}\right), \\ &\textit{where } \mathbf{L}_+ = \mathbf{D}_+ - \mathbf{A}_+, \mathbf{L}_- = \mathbf{D}_- + \mathbf{A}_-, \\ &\quad s.t.\ \mathbf{D}_+ + \mathbf{D}_- = \mathbf{D}, \mathbf{A}_+ - \mathbf{A}_- = \mathbf{A}, \\ &\qquad \mathbf{A}_{+,(i,j)} \in [0,1], \mathbf{A}_{-,(i,j)} \in [-1,0]. \end{aligned} \tag{10}$$

The first term, $tr\left(\mathbf{H}^\top \mathbf{L}_+ \mathbf{H}\right)$, enforces feature smoothing along homophilic edges, enhancing similarity among similar nodes. In contrast, the second term, $tr\left(\mathbf{H}^\top \mathbf{L}_- \mathbf{H}\right)$, promotes differentiated feature smoothing along heterophilic edges, capturing key information from dissimilar but knowledgeably connected nodes.

The energy $E(\mathcal{G}; f_\theta)$ allows for the distinction between homophilic and heterophilic edge attribute distributions. This is achieved by modeling the distributions as two terms, represented in terms of log data density, which decrease as density increases. This formulation aligns with the ideal conditions for OOD detection. Formally, the energy function is expressed as:

$$\begin{aligned} E(\mathcal{G}; f_\theta) = &\underbrace{-\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(r_{(u,v)}=1|\mathcal{G})}_{\downarrow \text{ for homophilic edges}} \\ &\underbrace{-\log \mathbb{E}_{(u,v)\in\mathcal{E}_{ind}} p_\theta(\mathcal{G})}_{\downarrow \text{ for homophilic edges}} + C. \end{aligned} \tag{11}$$

The above equation shows that $E(\mathcal{G}; f_\theta)$ promotes attribute inference by effectively distinguishing IND and OOD data. The second term benefits energy-based OOD detection by aligning with the principle that higher probability densities (homophilic samples) correspond to lower scores, meeting the requirements for OOD detection. The first term further enhances the discriminability between IND and OOD data, as OOD data is expected to have a lower conditional likelihood. A key advantage of our approach is that it avoids the computationally expensive task of directly estimating joint likelihoods with generative models, making it more practical for deployment on complex datasets.

For any energy distribution regularization term $\mathcal{L}_{reg}$, the edge attribute discriminator p_θ^* that minimizes $\mathcal{L}_{reg}$ produces energy distribution scores that are upper-bounded by ρ_{ind} for homophilic edges and lower-bounded by ρ_{out} for heterophilic edges, where $\rho_{ind} < \rho_{out}$ are two margin parameters.

The implications of this proposition are significant, as it demonstrates that our chosen regularization term, $\mathcal{L}_{reg}$, ensures optimal energy distribution scores for identifying heterophilic edges across the graph while maintaining alignment with the supervised training of the graph predictor. As a result, this approach provides a training objective that guarantees precise discrimination between homophilic and heterophilic edges in the training set, while simultaneously enhancing the quality of representation learning.

All theoretical proofs are provided in **Appendix D**.

4 Experiments

In this section, we present a series of experiments designed to evaluate the performance of H_2OGNN in addressing the challenges of homophilic/heterophilic graph learning from OOD detection perspective. The primary objectives of these experiments are to answer the following key questions:

Q1: Can H_2OGNN effectively handle heterophilic graphs? (Section 4.3)

Q2: Do OOD detection strategies enable effective edge attribute identification? (Section 4.4)

Q3: How does H_2OGNN demonstrate its power and robustness? (Section 4.5 and Section 4.6)

Additional details and experiments can be found in **Appendix A** and **Appendix B**.

4.1 Datasets

To comprehensively evaluate the performance of H_2OGNN, we conducted experiments on nine real-world datasets, including comparisons, effect studies, and efficiency analysis. We grouped these datasets into two categories, *homophily* and *heterophily*, by edge homophily, as defined in Definition 2. Specifically, datasets with homophily ratios greater than 0.5 are classified as **homophily**, while those with ratios less than or equal to 0.5 are classified as **heterophily**. For all benchmarks, we used the feature vectors, class labels, and

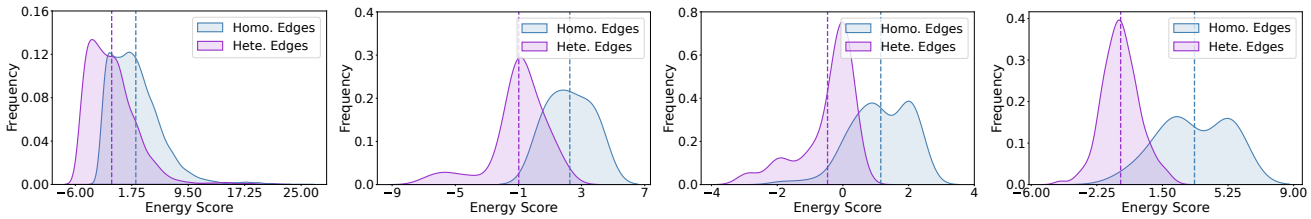

Figure 3: Visualization of the edge identification performance on homophilic and heterophilic graphs. From left to right are the results on datasets Pubmed, Wisconsin, Cornell, and Texas. The results indicate that energy-based method contributes to the effective expression of the energy margin between homophily (Homo.) and heterophily (Hete.) edges.

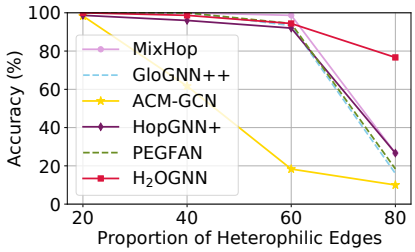

Figure 4: A trend graph showing the performance variations of existing methods and H_2OGNN as the proportion of heterophilic edges increases.

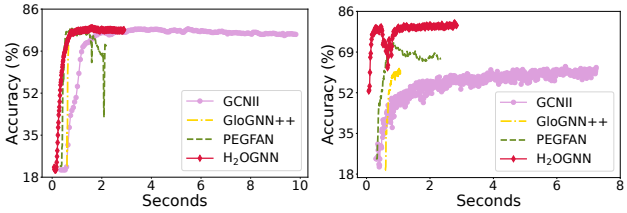

Figure 5: Efficiency analysis on the Citeseer (left) and Chameleon (right) datasets. The x-axis represents training time, while the y-axis denotes the accuracy score on the validation set.

ten random splits (48%/32%/20% of nodes per class for training/validation/testing) from [Pei *et al.*, 2020]. The dataset details are summarized in **Appendix A**.

4.2 Experiment Setup

We evaluate the performance of H_2OGNN by comparing it against both conventional GNNs and specialized heterophilic GNNs. The conventional GNNs include classical models such as GCN [Welling and Kipf, 2017], GAT [Veličković *et al.*, 2018], and GCNII [Chen *et al.*, 2020]. For heterophilic GNNs, we benchmark against several representative methods, including MixHop [Abu-El-Haija *et al.*, 2019], H_2GCN [Zhu *et al.*, 2020], PR-GNN [Chien *et al.*, 2021], LINKX [Lim *et al.*, 2021], ACM-GCN [Luan *et al.*, 2022], GloGNN++ [Li *et al.*, 2022], CAGNN [Chen *et al.*, 2024], HopGNN+ [Chen *et al.*, 2023], and PEGFAN [Li *et al.*, 2024].

Implementation Details. We implement H_2OGNN using PyTorch and conducted all experiments on a single NVIDIA RTX A6000 GPU with 48GB of memory, employing Adam [Kingma and Ba, 2014] as the optimizer. Hyperparameters were tuned through grid search, with specific values such as learning rates in $\{5e-4, 5e-3, 5e-2, 1e-2\}$, weight decay in $\{5e-7, 5e-6, 5e-5, 1e-4, 5e-4, 1e-3, 1e-2\}$, hidden dimensions in $\{32, 64, 128, 256, 512, 1024, 2048\}$, and layer numbers range in $[1, 128]$. As results for most baseline methods on our benchmark datasets are publicly available, we report these directly. For cases where results were unavailable, we ran the original code provided by the authors.

4.3 Comparative Experiments: Heterophilic Challenge (RQ1)

We evaluate H_2OGNN on both homophilic and heterophilic graphs, with results shown in Table 1. Significant observations from the analysis are as follows:

Observation 1: Most heterophilic GNNs struggle with complex heterophilic graph data. Many heterophilic GNNs struggle with all types of heterophilic graphs, particularly in datasets like Actor, Chameleon, and Squirrel, which are characterized by high proportion of structural imbalance. In these cases, most methods fail to achieve satisfactory performance.

Observation 2: H_2OGNN excels on both homophilic and heterophilic graphs. H_2OGNN outperforms other heterophilic GNNs overall and performs well on complex datasets like Actor, Chameleon, and Squirrel.

4.4 Effect of Integrating OOD Detection (RQ2)

The H_2OGNN framework introduces a novel OOD detection method based on signed energy regularization. This approach distinguishes homophilic and heterophilic edges by assigning different energy scores and minimizing shifts for edges with the same attribute by using consistent signs. Figure 3 shows the energy distribution of homophilic and heterophilic edges across datasets, highlighting how our strategy captures attribute differences while avoiding excessive energy shifts. Additionally, Figure 4 illustrates H_2OGNN's performance trend with OOD detection as the proportion of heterophilic edges increases. Two key observations are summarized below.

Observation 3: The OOD detection strategy effectively models the classification margin of edge attributes. As shown in Figure 3, the energy margin gaps reveal a clear distinction between homophilic and heterophilic edges, indicating that the energy-based edge representation is discriminative and aids in edge attribute identification. Homophilic edges mainly concentrate in the positive energy range, while

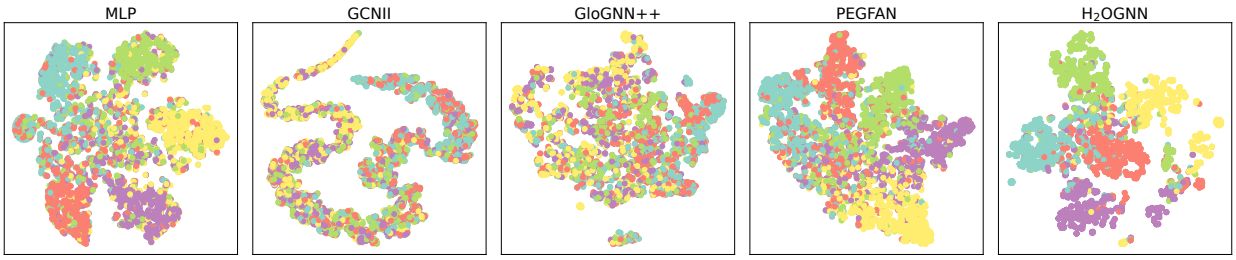

Figure 6: Visualization of the deep aggregation feature in several baselines and H_2OGNN on Squirrel dataset.

heterophilic edges are concentrated in the negative range. This boundary not only keeps homophilic edges within a focused energy region but also minimizes semantic deviations within the same attribute, enhancing generalization.

Observation 4: The OOD detection-based H_2OGNN maintains good performance on various heterophilic graphs. As shown in Figure 4, H_2OGNN, with its OOD detection strategy, consistently outperforms other methods, maintaining good performance even on graphs with a high proportion ($\geq 60\%$) of heterophilic edges, while others experience significant performance declines.

In summary, the introduced OOD detection strategy in this paper demonstrates good edge recognizability and can effectively distinguish between homophilic and heterophilic edge attributes.

4.5 Efficiency Analysis (RQ3)

Simplicity is a key advantage of H_2OGNN. To substantiate this with quantitative evidence, we demonstrate its efficiency by showing that it achieves rapid convergence on both homophilic and heterophilic datasets, as depicted in Figure 5.

Observation 5: H_2OGNN converges relatively fast on both homophilic and heterophilic datasets. On heterophilic graphs, this is a relatively obvious observation; On the homophilic graph Citeseer, although H_2OGNN converges slightly slower than the state-of-the-art work PEGFAN, we achieve relatively higher performance and stability.

Observation 6: H_2OGNN demonstrates significant advantages in both speed and performance on heterophilic graphs. On the heterophilic graph Chameleon, it shows a clear trend of rapid convergence compared to other baselines. Although there was a brief performance decline in the training process, it quickly stabilized at a better accuracy result subsequently.

4.6 Generalization Study (RQ3)

Higher generalization is another key advantage of H_2OGNN. We demonstrate its strength by showcasing the clustering effects of the features captured by H_2OGNN and its performance in facing a classic challenge for GNNs—over-smoothing. The experimental results can be seen in Figures 6 and 7.

Observation 7: H_2OGNN can capture better feature representations. In comparison with the feature visualization of the baselines on heterophilic graphs, H_2OGNN achieves good category clustering effects within its deep features, demonstrating better intra-class clustering and inter-class separability.

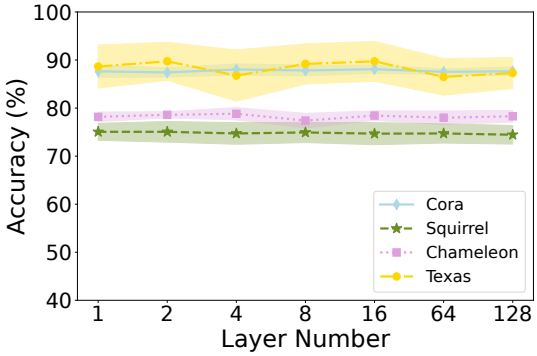

Figure 7: The performance of H_2OGNN across both homophilic and heterophilic datasets as the number of layers increases.

Observation 8: H_2OGNN can alleviate the over-smoothing. Over-smoothing is a key challenge in graph representation learning. A robust and generalizable GNN should combat this issue, ensuring performance remains stable as the number of layers increases. Many standard GNNs suffer significant performance degradation after just two layers. In contrast, H_2OGNN maintains stable performance on both homophilic and heterophilic graphs, even with up to 128 layers.

In summary, H_2OGNN possesses excellent feature representation learning capabilities and robust and generalizable performance, which sufficiently demonstrate the strong of H_2OGNN.

5 Conclusion

This paper introduces H_2OGNN, a novel framework for heterophilic graph representation learning (GRL) that utilizes OOD detection to infer edge attributes and enhance local message passing. The proposed OOD strategy effectively balances generalization on IND samples while preserving distributional differences between homophilic and heterophilic edges. This not only boosts message passing efficiency but also enables more distinctive representations. Extensive experiments demonstrate H_2OGNN's superior performance on heterophilic graphs and its effectiveness in edge identification. Additionally, we provide theoretical insights into how H_2OGNN improves performance on heterophilic graphs, underscoring its potential to advance the field.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62337001, No. 62037001), and the Jinhua Science and Technology Plan (No. 2023-3-003a).

Contribution Statement

Changqin Huang (email: cqhuang@zju.edu.cn) is the corresponding author of this work.

References

[Abu-El-Haija *et al.*, 2019] Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Nazanin Alipourfard, Kristina Lerman, Hrayr Harutyunyan, Greg Ver Steeg, and Aram Galstyan. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *ICML*, pages 21–29, 2019.

[Bi *et al.*, 2024] Wendong Bi, Lun Du, Qiang Fu, Yanlin Wang, Shi Han, and Dongmei Zhang. Make heterophilic graphs better fit gnn: A graph rewiring approach. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8744–8757, 2024.

[Bo *et al.*, 2021] Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *AAAI*, pages 3950–3957, 2021.

[Cai *et al.*, 2025a] Tingyi Cai, Yunliang Jiang, Ming Li, Changqin Huang, Yi Wang, and Qionghao Huang. ML-GOOD: Towards multi-label graph out-of-distribution detection. In *AAAI*, volume 39, pages 15650–15658, 2025.

[Cai *et al.*, 2025b] Tingyi Cai, Yunliang Jiang, Yixin Liu, Ming Li, Changqin Huang, and Shirui Pan. Out-of-distribution detection on graphs: A survey. *arXiv preprint arXiv:2502.08105*, 2025.

[Chen *et al.*, 2020] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *ICML*, pages 1725–1735, 2020.

[Chen *et al.*, 2023] Jie Chen, Zilong Li, Yin Zhu, Junping Zhang, and Jian Pu. From node interaction to hop interaction: New effective and scalable graph learning paradigm. In *CVPR*, pages 7876–7885, 2023.

[Chen *et al.*, 2024] Jie Chen, Shouzhen Chen, Junbin Gao, Zengfeng Huang, Junping Zhang, and Jian Pu. Exploiting neighbor effect: Conv-agnostic GNN framework for graphs with heterophily. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13383–13396, 2024.

[Chien *et al.*, 2021] Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *ICLR*, 2021.

[Gong *et al.*, 2024] Chenghua Gong, Yao Cheng, Jianxiang Yu, Can Xu, Caihua Shan, Siqiang Luo, and Xiang Li. A survey on learning from graphs with heterophily: Recent advances and future directions. *arXiv preprint arXiv:2401.09769*, 2024.

[He *et al.*, 2023] Dongxiao He, Jitao Zhao, Rui Guo, Zhiyong Feng, Di Jin, Yuxiao Huang, Zhen Wang, and Weixiong Zhang. Contrastive learning meets homophily: Two birds with one stone. In *ICML*, pages 12775–12789, 2023.

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

[Li *et al.*, 2022] Xiang Li, Renyu Zhu, Yao Cheng, Caihua Shan, Siqiang Luo, Dongsheng Li, and Weining Qian. Finding global homophily in graph neural networks when meeting heterophily. In *ICML*, pages 13242–13256, 2022.

[Li *et al.*, 2023a] Wen-Zhi Li, Chang-Dong Wang, Hui Xiong, and Jian-Huang Lai. Homogcl: Rethinking homophily in graph contrastive learning. In *KDD*, pages 1341–1352, 2023.

[Li *et al.*, 2023b] You Li, Bei Lin, Binli Luo, and Ning Gui. Graph representation learning beyond node and homophily. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4880–4893, 2023.

[Li *et al.*, 2024] Jianfei Li, Ruigang Zheng, Han Feng, Ming Li, and Xiaosheng Zhuang. Permutation equivariant graph framelets for heterophilous graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9):11634–11648, 2024.

[Liang *et al.*, 2023] Langzhang Liang, Xiangjing Hu, Zenglin Xu, Zixing Song, and Irwin King. Predicting global label relationship matrix for graph neural networks under heterophily. In *NeurIPS*, pages 10909–10921, 2023.

[Lim *et al.*, 2021] Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. In *NeurIPS*, pages 20887–20902, 2021.

[Liu *et al.*, 2023] Yixin Liu, Kaize Ding, Huan Liu, and Shirui Pan. Good-d: On unsupervised graph out-of-distribution detection. In *ICDM*, pages 339–347, 2023.

[Luan *et al.*, 2022] Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Revisiting heterophily for graph neural networks. In *NeurIPS*, pages 1362–1375, 2022.

[Luan *et al.*, 2024] Sitao Luan, Chenqing Hua, Qincheng Lu, Liheng Ma, Lirong Wu, Xinyu Wang, Minkai Xu, Xiao-Wen Chang, Doina Precup, Rex Ying, et al. The heterophilic graph learning handbook: Benchmarks, models, theoretical analysis, applications and challenges. *arXiv preprint arXiv:2407.09618*, 2024.

[Ma *et al.*, 2021] Yao Ma, Xiaorui Liu, Tong Zhao, Yozen Liu, Jiliang Tang, and Neil Shah. A unified view on graph neural networks as graph signal denoising. In *CIKM*, pages 1202–1211, 2021.

[Ma *et al.*, 2022] Yao Ma, Xiaorui Liu, Neil Shah, and Jiliang Tang. Is homophily a necessity for graph neural networks? In *ICLR*, 2022.

[Pei *et al.*, 2020] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *ICLR*, 2020.

[Ranzato, 2007] Boureau Y. L. Chopra S. LeCun Y. Ranzato, M. A. A unified energy-based framework for unsupervised learning. In *AISTATS*, pages 371–379, 2007.

[Song and Wang, 2022] Yu Song and Donglin Wang. Learning on graphs with out-of-distribution nodes. In *KDD*, pages 1635–1645, 2022.

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

[Wang *et al.*, 2024] Chenhao Wang, Yong Liu, Yan Yang, and Wei Li. HeterGCL: Graph contrastive learning framework on heterophilic graph. In *IJCAI*, 2024.

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

[Zhu *et al.*, 2020] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *NeurIPS*, pages 7793–7804, 2020.

[Zhu *et al.*, 2021] Meiqi Zhu, Xiao Wang, Chuan Shi, Houye Ji, and Peng Cui. Interpreting and unifying graph neural networks with an optimization framework. In *WWW*, pages 1215–1226, 2021.