# Beyond Energy Smoothing Propagation in Graph Out-of-Distribution Detection

Xiaolong Fan*, Yuxuan Liu†, Maoguo Gong*, Mingyang Zhang*, Xiangming Jiang*,
Jianzhao Li‡, and Shanfeng Wang§
*School of Electronic Engineering, Key Laboratory of Collaborative Intelligence Systems,
Ministry of Education, Xidian University, Xi'an, China
†School of Optoelectronic Engineering, Xidian University, Xi'an, China
‡Guangzhou Institute of Technology, Xidian University, Guangzhou, China
§School of Cyber Engineering, Xidian University, Xi'an, China
Email: xiaolongfan@outlook.com; yuxuanliu@stu.xidian.edu.cn; gong@ieee.com; myzhang@xidian.edu.cn;
xmjiang@xidian.edu.cn; lijianzhao@xidian.edu.cn; sfwang@xidian.edu.cn

***Abstract*—Recently, energy propagation based methods have shown considerable success in graph out-of-distribution (OOD) detection. However, existing approaches predominantly adhere to an energy smoothing propagation framework. A key limitation arises when nodes from two distinct distributions are directly connected, as this can cause the propagation mechanism to fail, thereby compromising detection reliability. To mitigate the limitations of employing energy smoothing, we propose augmenting the propagation framework with an energy difference propagation. By integrating energy smoothing and difference propagation, we propose SDGSafe, a smoothing and difference graph out-of-distribution detection method, for graph out-of-distribution detection. This results in a hybrid propagation scheme where the operation of each central node is determined by its neighbors' distribution heterophily, i.e., energy smoothing propagates across edges linking nodes of the same distribution, while energy difference operates across edges that bridge different distributions. Extensive experiments on benchmark datasets demonstrate that our method substantially outperforms the baseline that relies solely on smooth energy propagation, confirming the effectiveness of the integrated approach.**

***Index Terms*—neural network, graph neural network, out-of-distribution detection, energy propagation**

## I. INTRODUCTION

In recent years, graph neural networks (GNNs) [1], [2] have emerged as a powerful framework for analyzing non-Euclidean graph data, finding broad utility across diverse fields including recommendation systems [3], natural language processing [4], and beyond. GNNs utilize inter-node relationships to model intricate dependencies within graph-structured data, thereby enhancing their efficacy in predictive applications. In practical applications, however, models frequently encounter test data that deviates substantially from the training distribution, a challenge commonly termed out-of-distribution (OOD) generalization. Such distributional shifts may result in predictions that are unjustifiably confident, thereby introducing considerable hazards in safety-critical domains. In knowledge graph applications, for instance, the ongoing discovery of novel entities and relations necessitates their integration into existing graph structures. This process, however, introduces previously unobserved node categories, compelling GNN models to effectively manage the resultant uncertainty.

To enhance the generalizability of graph machine learning models, graph out-of-distribution (OOD) detection [5], [6], [7] has been developed. This task aims to differentiate in-distribution (IND) graphs or nodes from their OOD counterparts, thereby identifying and excluding OOD samples to improve model robustness. Distinct from OOD detection in domains like images or text, the task on graphs presents distinctive challenges. Graph-structured data, which inherently resides in a non-Euclidean space, exhibits complex topologies and dynamic relational dependencies among nodes. Such inherent complexity significantly complicates the discrimination between in-distribution (IND) and out-of-distribution (OOD) samples.

To tackle these distinct challenges, energy based graph out-of-distribution (OOD) detection has emerged as a rapidly expanding research direction. Such methods employ energy based models (EBMs) [8] to compute OOD scores, which are subsequently propagated across the graph structure to enhance detection efficacy. Specifically, the pioneering energy based graph OOD detection framework, GNNSafe [5], derives an OOD discriminator directly from the output logits of a standard graph neural network classifier through a free energy function, thereby circumventing the need for training auxiliary models. To improve detection consensus over the graph topology, the framework further integrates a learning-free energy based belief propagation mechanism. Building upon GNNSafe, NODESAFE [9] addresses the influence of extreme fractional values during aggregation by incorporating optimization terms that constrain negative energy scores and mitigate logarithmic bias, thereby improving result stability and model robustness. In comparison, TopoOOD [10] emphasizes neighbor context, quantifying local structural variation via k-hop Dirichlet energy. Separately, Energy-Def [11] integrates an EBM into the GNN architecture to estimate node-level densities for OOD detection.

This work is supported by the National Natural Science Foundation of China (Grant No. 62476210). Corresponding Author (Xiangming Jiang, email: xmjiang@xidian.edu.cn)

Despite the considerable success of energy propagation models, existing approaches [5], [9] commonly adopt an energy smoothing mechanism to minimize energy discrepancies between adjacent nodes. However, this strategy may become ineffective when graph edges link nodes originating from distinct distributions, such as when an in-distribution central node is directly connected to an out-of-distribution neighbor. A key remaining challenge, therefore, is how to overcome the limitations inherent to conventional energy smoothing propagation in graph OOD detection.

To address this challenge, we augment the propagation framework by introducing an energy difference propagation mechanism to mitigate the limitations of conventional energy smoothing. This approach can be realized by reformulating the standard summation based aggregation into a differential form. In contrast to prior energy smoothing propagation, which reduces energy disparities between adjacent nodes, the proposed energy difference propagation is designed to amplify the energy gap across connected node pairs. When two connected nodes originate from distinct distribution types, e.g., an IND node is connected to an OOD node, the energy difference propagation method effectively mitigates the inter-distribution energy gap. In contrast, energy smoothing propagation often fails under such conditions. Additionally, we integrate the proposed difference energy propagation with established energy smoothing techniques to formulate SDGSafe, a novel energy propagation framework. SDGSafe first assigns initial energy values, interpreting higher-energy nodes as probable OOD instances and lower-energy nodes as likely IND samples. Subsequently, when a central node and its neighbor originate from different distributions, the model applies differential energy propagation. Otherwise, standard energy smoothing propagation is employed. Following multiple propagation iterations, the refined energy scores are used as the criterion for identifying out-of-distribution nodes in the graph.

Experimental results on well-known graph benchmark datasets consistently demonstrate the effectiveness of the developed SDGSafe framework. To summarize, the contributions of this paper are summarized as follows:

1) We analyze the previous energy smoothing propagation and additionally develop the energy difference propagation for graph OOD detection.
2) We integrate the proposed difference energy propagation with previous energy smoothing propagation to develop SDGSafe, a novel energy propagation framework.
3) The proposed method demonstrates consistently superior results across a range of graph benchmark datasets, attesting to its effectiveness.

## II. Related Work

### A. Graph Neural Network

Message passing neural networks [1], [12] provide a unified framework that encapsulates a wide range of graph neural networks, such as GCN [13], [14], which have achieved state-of-the-art results on graph tasks. However, prevailing approaches are predominantly designed under the i.i.d. assumption, where training and test data are presumed to originate from an identical distribution. Consequently, GNNs often treat all inputs as conforming to a single distribution, thereby overlooking the distinct patterns. This limitation highlights the need for developing dedicated graph OOD detection methods that can effectively address such distributional disparities.

### B. Out-of-distribution Detection on Graphs

Out-of-distribution (OOD) detection for non-graph data [8] has received substantial research interest. Nonetheless, prevailing approaches generally rely on the assumption that instances are i.i.d., thereby failing to account for the interdependent relationships that frequently characterize real-world data. In comparison, the problem of OOD detection on graphs remains significantly under explored. Recent literature has introduced a range of approaches, including enhancement-based [6], [15], reconstruction-based [7], [16], propagation-based [5], [9], and classification-based [17], [18] techniques. Among these, this work centers on propagation-based methodologies and develops a novel energy propagation framework.

## III. Preliminaries

This section commences with a formal problem formulation and then shifts to an examination of prior frameworks.

### A. Problem Definition

Consider an attributed graph $\mathrm{G} = (\mathrm{V}, \mathrm{E}, \mathrm{X})$, where V is the node set, E is the edge set, and $\mathrm{X} = \{\mathrm{x}_1, ..., \mathrm{x}_i, ..., \mathrm{x}_n\}$ with $\mathrm{x}_i \in \mathbb{R}^d$ constitutes the node feature. The graph structure is characterized by an adjacency matrix $\mathrm{A} \in \{0,1\}^{n\times n}$, where $\mathrm{A}_{ij} = 1$ signifies the edge $(\mathrm{v}_i, \mathrm{v}_j) \in \mathrm{E}$. A graph encoder $f(\cdot)$ typically maps the input $(\mathrm{A}, \mathrm{X})$ to node embeddings $\mathrm{H} = \{\mathrm{h}_1, ..., \mathrm{h}_i, ..., \mathrm{h}_n\}$ for downstream tasks. This paper investigates the problem of out-of-distribution detection on graphs, treating each node as an individual sample to be evaluated. Graph out-of-distribution detection aims to develop a method that can effectively quantify the uncertainty or deviation of each node from the known training distribution.

**Definition 1** (Graph Out-of-Distribution Detection)**.** *Given in-distribution data $\mathcal{D}_{in}$ and out-of-distribution data $\mathcal{D}_{out}$, graph OOD detection method expects to construct a decision function to access samples from $\mathcal{D}_{in}$ while rejecting samples from $\mathcal{D}_{out}$, i.e.,*

$$h_i = \begin{cases} f(x_i, \mathcal{G}_i), & if \ s_i \in \mathcal{D}_{in} \\ reject, & if \ s_i \in \mathcal{D}_{out} \end{cases} \tag{1}$$

*where $s_i$ denotes a node sample, consisting of the corresponding node feature $x_i$ and its ego-graph $\mathcal{G}_i$.*

### B. Energy Based Graph OOD Detection

Given an initial node feature $x_i$ and its ego-graph $\mathcal{G}_i$, a graph neural network encoder generates node embeddings for downstream tasks. Here, the classical graph convolutional network [13] is employed to update node representation with

$$h_i^{(\ell+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \frac{1}{\sqrt{d_i d_j}} h_i^{(\ell)}\right) \tag{2}$$

where $h_i^{(\ell)}$ is the $i$-th node representation at the $\ell$-th layer, $d_i$ is the degree of the $i$-th node, $\mathcal{N}(i)$ is the neighbor of the $i$-th node, $\sigma$ is the non-linear activation function. After $L$-th updates, the graph convolutional network encoder outputs the final representation $h_i$ as the $i$-th node logit. The logit is then passed through a softmax function to obtain a categorical distribution for classification in form of

$$p(y|x_i, \mathcal{G}_i) = \frac{e^{h_i[y]/T}}{\sum_{c=1}^{C} e^{h_i[c]/T}} \tag{3}$$

where $C$ denotes the class number, $T$ is the temperature, and $h_i[c]$ denotes the $c$-th entry of the node representation vector.

Suppose $E(x_i, \mathcal{G}_i, y) = -h_i[y]$, the previous work [5] establishes a link with the energy based models and then obtain the following Boltzmann distribution in form of

$$p(y|x_i, \mathcal{G}_i) = \frac{e^{-E(x_i,\mathcal{G}_i,y)/T}}{\sum_{c=1}^{C} e^{-E(x_i,\mathcal{G}_i,c)/T}} = \frac{e^{-E(x_i,\mathcal{G}_i,y)/T}}{e^{-E(x_i,\mathcal{G}_i)/T}} \tag{4}$$

where $E(x_i, \mathcal{G}_i, y)$ denotes the energy function. By additionally marginalizes over $y$, the free energy function $E(x_i, \mathcal{G}_i)$ can be expressed in form of

$$E(x_i, \mathcal{G}_i) = -T \cdot \log \sum_{c=1}^{C} e^{h_i[c]/T}. \tag{5}$$

Applying the aforementioned energy function yields an energy value for each input node. This energy value measures the classifier's confidence in the node and serves directly as the criterion for OOD detection.

To further harness the sample dependencies inherent in the graph structure, prior work [5] introduces energy based propagation mechanism by propagating the energy value of each node across the graph structure, i.e.,

$$e_i^{(k+1)} = \alpha e_i^{(k)} + (1-\alpha) \sum_{j\in\mathcal{N}(i)} \frac{1}{d_i} e_j^{(k)} \tag{6}$$

where $\alpha$ is the weight parameter, $d_i$ is the degree of the $i$-th node, $e_i^{(k)}$ is the energy value after $k$-th propagation, and $e_i^{(0)} = E(x_i, \mathcal{G}_i)$. By adopting such energy propagation, the confidence produced by the classifier is reinforced along the graph structure, leveraging global information to enhance the OOD estimation for individual nodes. The energy propagation shifts the updated energy values toward the consensus of the majority among neighbor nodes. Consequently, a node exhibits a stronger tendency to be identified as OOD if the majority of its neighbors are OOD samples.

## IV. Methodology

This section begins with the analysis of the current framework and then proceeds to delineate the proposed solution.

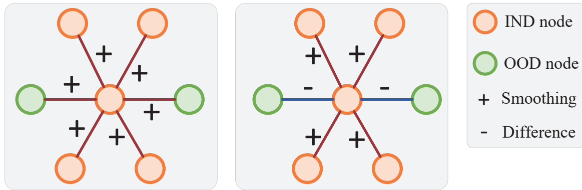



Fig. 1. Energy smoothing propagation vs. energy difference propagation. Energy smoothing propagates across edges linking nodes of the same distribution, while energy difference operates across edges that bridge different distributions.

### A. Analysis of Current Framework

Despite the considerable success achieved by current energy propagation methods in graph OOD detection, we note that the current energy propagation framework, shown in Eqn. 6, belongs to the energy smoothing propagation framework. Hence, we can easily obtain the following proposition.

**Proposition 1.** *Energy smoothing propagation makes the energy value more similar.*

*Proof.* Given two energy values $e_i$ and $e_j$, and assuming $d_e = |e_i - e_j|$, after energy smoothing propagation, the energy difference is obtained as $\hat{d}_e$. We can easily see that after smooth energy smoothing propagation, $\hat{d}_e$ is less than $d_e$. Therefore, we can obtain energy smoothing propagation makes the energy value more similar. □

Proposition 1 tells us that energy smoothing propagation makes the energy value more similar. However, the effectiveness of energy smoothing propagation may be compromised when edges in the graph connect nodes from different types of distributions. For example, the central node is an in-distribution sample, whereas a directly connected neighbor node is an out-of-distribution sample.

To develop an intuitive understanding of this issue, consider two connected nodes, $x_i$ and $x_j$, which are drawn from different data distributions. Specifically, $x_i$ is an in-distribution sample, whereas $x_j$ is an out-of-distribution sample. $e_i$ and $e_j$ are the energy values of nodes $x_i$ and $x_j$ before energy smoothing propagation, respectively. Generally, low energy implies high likelihood, meaning the samples are more likely to be in-distribution, and vice versa. Hence, we have $e_i < e_j$ and thus denote $d_e = e_j - e_i$. A core criterion for OOD detection is that the energy gap between IND and OOD data should be maximized. Proposition 1, however, reveals a fundamental limitation is that the smoothing process inherently diminishes the energy disparity between nodes. This effect runs counter to the requirement for a large energy gap in graph OOD detection, thereby rendering the mechanism ineffective for this specific scenery.

### B. Proposed Method

To address the problem arising from solely employing the energy smoothing propagation, we propose augmenting the propagation framework with an energy difference propagation and further develop SDGSafe, an smoothing and difference graph out-of-distribution detection method, for graph out-of-distribution detection.

*1) Energy Difference Propagation:* Different from the previous energy smoothing propagation, which makes the node energy value more similar, we develop energy difference propagation to amplify energy gap. Specifically, given node $x_i$ and its neighbor node $x_j \in \mathcal{N}(i)$, and their corresponding energy values $e_i$ and $e_j$, energy difference propagation can be described in form of

$$e_i^{(k+1)} = \alpha e_i^{(k)} - (1-\alpha) \sum_{j \in \mathcal{N}(i)} \frac{1}{d_i} e_j^{(k)} \tag{7}$$

where $\alpha$ is the wight parameter, $\mathcal{N}(i)$ and $d_i$ are the neighbor nodes and the degree of the $i$-th node, $e_i^{(k)}$ is the energy value after $k$-th propagation respectively, and $e_i^{(0)} = E(x_i, \mathcal{G}_i)$. Note that the only difference between energy difference propagation and the previous energy smoothing propagation is that the former uses a subtractive differential form while the latter uses an aggregation form. Figure 1 shows a comparison of energy smoothing propagation and energy difference propagation.

*2) Proposed SDGSafe:* Energy smoothing propagation can be used to enhance out-of-distribution data detection when the two nodes connected by an edge belong to the same type of distribution, while the energy difference operator can be used when the two nodes belong to different types of distribution. The next question is how to apply different energy propagation methods to different neighbor nodes of a given central node.

To answer this question, we develop the SDGSafe framework consisting of the following three phases. First, following the previous energy based model, we can obtain the energy value of the $i$-th node in form of

$$e_i^{(0)} = E(x_i, \mathcal{G}_i) = -T \cdot \log \sum_{c=1}^{C} e^{h_i[c]/T} \tag{8}$$

where $C$ denotes the class number, $T$ is the temperature, and $h_i[c]$ denotes the $c$-th entry of the node representation vector. Second, to determine whether a node might be IND or OOD, a direct approach is to consider nodes with higher energy values as most likely OOD nodes, and nodes with lower energy values as most likely IND nodes. Inspired by the previous work [19], for all nodes, we obtain energy thresholds in form of

$$S_{ind}^{(k)} = \{s_i | e_i^{(k)} < \lambda_\alpha\}, \;\; S_{ood}^{(k)} = \{s_i | e_i^{(k)} > \lambda_{100-\alpha}\} \tag{9}$$

where $S_{ind}^{(k)}$ and $S_{ood}^{(k)}$ denote the IND sample set and OOD sample set respectively at the $k$-th propagation, $\lambda_\alpha$ denotes the $\alpha$-th percentile of energy values. Third, we can further apply different energy propagation methods based on the obtained node sample sets $S_{ind}$ and $S_{ood}$, i.e.,

$$\begin{aligned} e_i^{(k+1)} &= \alpha e_i^{(k)} + (1-\alpha) \sum_{j \in \mathcal{N}(i)} \frac{\beta}{d_i} e_j^{(k)} \\ \beta &= \begin{cases} 1, & if\;\; i,j \in S_{ind}^{(k)}/S_{ood}^{(k)} \\ -1, & if\;\; i,j \notin S_{ind}^{(k)}/S_{ood}^{(k)} \end{cases} \end{aligned} \tag{10}$$

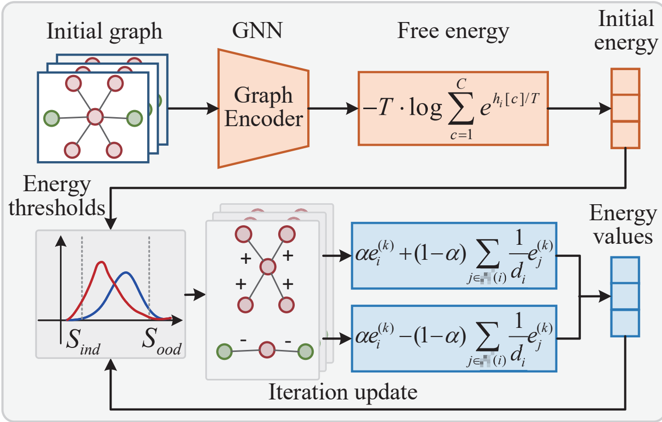



Fig. 2. The overall framework of the proposed SDGSafe, which integrates energy smoothing and difference propagation for graph OOD detection.

---

**Algorithm 1:** The proposed SDGSafe framework

**Input:** Node representation $x_i$ and its ego-graph $\mathcal{G}_i$
**Output:** OOD detection result for $i$-the node
$k \leftarrow 0$;
Obtaining initial energy $e_i^{(0)}$ via Eqn 8;
**while** $k < K$ **do**
  Obtaining energy thresholds via Eqn. 9;
  Applying energy propagation via Eqn. 10;
  $k \leftarrow k + 1$;
**end**
Deriving a decision criterion for OOD detection by applying a classification threshold to the energy score;

---

where $i, j \in S_{ind}^{(k)}/S_{ood}^{(k)}$ denotes that node $i$ and $j$ both belong to sample set $S_{ind}^{(k)}$ or $S_{ood}^{(k)}$. Similarly, we use the propagation iteration method, and use the energy value obtained by $K$ iterations as the final OOD detection discrimination score. The proposed SDGSafe results in a hybrid propagation scheme where the operation each central node is determined by its neighbors' distribution heterophily, i.e., energy smoothing propagates across edges linking nodes of the same distribution, while energy difference operates across edges that bridge different distributions.

*3) Dicussion and analysis:* The overall framework is shown in Figure 2 and the pseudocode of SDGSafe is shown in Algorithm 1. The proposed SDGSafe is a post-hoc out-of-

TABLE I
STATISTICS OF GRAPH BENCHMARK DATASETS. # C IS THE NUMBER OF NODE CLASSES.

| DATASET | CONTEXT | # NODES | # EDGES | # FEATURES | # C |
|---|---|---|---|---|---|
| Cora | Citation network | 2,708 | 10,556 | 1,433 | 7 |
| CiteSeer | Citation network | 3,327 | 9,104 | 3,703 | 6 |
| PubMed | Citation network | 19,717 | 88,648 | 500 | 3 |
| Ratings | Amazon network | 24,492 | 186,100 | 300 | 5 |
| Roman | Wikipedia network | 22,662 | 65,854 | 300 | 18 |
| ArXiv | Citation network | 169,343 | 152,226 | 128 | 40 |

TABLE II

PERFORMANCE COMPARISON WITH BASELINE METHODS IN TERMS OF AUROC (↑).

| METHOD | Cora@AUROC(↑) | | | CiteSeer@AUROC(↑) | | | PubMed@AUROC(↑) | | | Ratings@AUROC(↑) | | | Roman@AUROC(↑) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label |
| MSP | 78.38±0.19 | 89.02±0.14 | 78.71±0.09 | 76.10±0.16 | 77.18±0.19 | 70.15±5.91 | 80.57±1.19 | 88.48±0.74 | 52.69±2.46 | 67.95±0.20 | 74.29±0.48 | 49.83±0.46 | 31.02±0.29 | 57.31±0.31 | 61.56±0.40 |
| Energy | 76.87±1.70 | 88.22±1.35 | 79.09±1.57 | 73.88±2.47 | 75.00±3.40 | 61.42±1.65 | 80.71±2.08 | 88.02±2.21 | 52.85±2.53 | 75.56±1.85 | 82.47±2.65 | 57.72±1.92 | 26.72±0.61 | 69.94±0.16 | 73.79±0.38 |
| ODIN | 24.60±0.67 | 17.52±0.93 | 20.62±0.88 | 27.50±0.05 | 23.54±0.20 | 38.83±2.26 | 23.89±2.52 | 14.94±1.54 | 46.17±6.63 | 21.23±1.35 | 20.66±0.73 | 43.18±1.80 | 69.79±0.25 | 30.42±0.30 | 25.98±0.25 |
| Mahalanobis | 38.31±0.49 | 46.35±0.42 | 71.18±0.49 | 38.67±0.66 | 43.70±1.20 | 67.60±2.41 | 55.88±1.84 | 63.36±2.61 | 53.14±3.21 | 53.36±0.09 | 82.67±0.03 | 53.49±0.08 | 30.47±0.07 | 67.14±0.32 | 56.57±0.22 |
| GKDE | 49.24±1.07 | 49.06±1.18 | 67.21±2.58 | 47.34±1.48 | 47.35±1.39 | 68.92±8.87 | 50.47±0.83 | 50.39±0.89 | 63.30±3.76 | 50.15±0.46 | 50.07±0.68 | 48.20±1.92 | 49.09±0.46 | 49.14±0.55 | 54.98±2.55 |
| GPN | 49.39±1.35 | 49.47±0.42 | 71.25±2.23 | 47.92±0.46 | 46.38±2.80 | 70.86±2.71 | 49.69±1.36 | 49.68±1.37 | 64.27±6.14 | 50.31±0.45 | 50.43±0.24 | 49.74±0.95 | 49.76±0.86 | 49.75±0.71 | 51.34±3.11 |
| GNNSafe | **94.28±0.32** | 95.00±1.31 | 71.53±2.85 | 88.85±1.41 | 89.60±1.61 | 70.28±3.67 | **97.54±0.81** | 97.04±1.63 | 65.05±2.41 | 84.78±2.58 | 86.69±2.47 | 60.17±2.36 | 38.28±0.55 | 74.67±0.18 | 70.95±0.52 |
| SDGSafe | 94.07±0.18 | **95.26±0.91** | **81.58±1.75** | **88.95±1.42** | **89.72±1.72** | **76.74±4.74** | 97.15±0.81 | **97.59±1.62** | **69.78±2.71** | **85.53±2.54** | **87.45±2.50** | **62.20±2.43** | **39.28±0.55** | **74.74±0.19** | **70.97±0.52** |

TABLE III

PERFORMANCE COMPARISON WITH BASELINE METHODS IN TERMS OF AUPR (↑).

| METHOD | Cora@AUPR(↑) | | | CiteSeer@AUPR(↑) | | | PubMed@AUPR(↑) | | | Ratings@AUPR(↑) | | | Roman@AUPR(↑) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label |
| MSP | 53.00±0.19 | 76.85±0.47 | 47.42±0.16 | 43.79±0.03 | 70.69±0.46 | 55.35±7.06 | 22.88±0.48 | 46.18±1.07 | 13.63±1.64 | 24.39±1.09 | 34.50±0.82 | 2.02±0.02 | 6.06±0.02 | 9.75±0.08 | 5.69±0.06 |
| Energy | 49.85±1.46 | 75.56±2.11 | 52.48±3.11 | 41.07±1.48 | 68.17±4.20 | 25.14±1.52 | 23.17±1.34 | 46.39±5.29 | 11.99±1.29 | 34.65±0.26 | 52.75±2.04 | 2.40±0.12 | 5.76±0.04 | 14.42±0.10 | 7.69±0.14 |
| ODIN | 17.55±0.11 | 16.37±0.13 | 15.12±0.12 | 15.30±0.01 | 14.64±0.03 | 14.00±0.62 | 2.90±0.08 | 2.67±0.03 | 8.56±1.36 | 5.45±0.09 | 5.39±0.04 | 1.62±0.06 | 3.32±0.07 | 5.99±0.02 | 2.22±0.01 |
| Mahalanobis | 23.47±0.24 | 33.89±0.69 | 41.65±1.06 | 19.61±0.40 | 26.79±1.43 | 47.42±5.40 | 12.12±0.54 | 14.49±2.13 | 10.69±0.47 | 12.34±0.09 | 50.57±0.14 | 2.19±0.01 | 5.97±0.01 | 13.58±0.23 | 4.93±0.06 |
| GKDE | 9.13±0.41 | 9.09±0.42 | 58.32±6.90 | 8.65±0.39 | 8.64±0.39 | 35.53±12.28 | 9.30±0.21 | 9.29±0.21 | 18.29±2.76 | 9.22±0.10 | 9.21±0.17 | 1.88±0.11 | 5.91±0.04 | 8.95±0.06 | 4.57±0.64 |
| GPN | 8.83±0.26 | 8.84±0.18 | 58.27±3.49 | 8.50±0.25 | 8.50±0.25 | 39.31±7.31 | 9.06±0.38 | 9.29±0.21 | 16.89±9.53 | 9.09±0.12 | 9.19±0.08 | 2.16±0.15 | 6.05±0.23 | 9.05±0.18 | 4.30±1.00 |
| GNNSafe | 87.56±0.38 | 88.84±2.17 | 41.33±5.21 | 77.96±2.48 | 79.78±2.41 | 28.11±3.48 | **87.44±1.44** | 86.49±3.56 | 21.75±3.40 | 67.40±2.11 | 64.00±2.72 | 2.58±0.16 | **6.72±0.05** | 16.38±0.27 | 8.54±0.23 |
| SDGSafe | **87.78±0.26** | **89.38±1.66** | **60.96±5.32** | **78.96±2.48** | **80.11±2.50** | **38.49±4.17** | 87.24±1.40 | **86.51±3.52** | **22.09±3.65** | **68.17±2.09** | **64.77±2.75** | **2.59±0.17** | **6.72±0.05** | **16.43±0.28** | **8.03±0.26** |

distribution detection method that operates without interfering with the underlying node classification process. Consequently, the classification accuracy of in-distribution data remains unaffected by our detection mechanism when using graph convolutional networks. Additionally, compared to prior approaches, the introduced energy difference propagation only modifies the aggregation scheme, thereby avoiding significant increases in time and space complexity.

## V. EXPERIMENT

In this section, we conduct validation experiments to verify the effectiveness of the proposed method.

### A. Datasets

We use six publicly available benchmark datasets, including Cora, CiteSeer, PubMed, Ratings, Roman, and ArXiv, for validation. Here, Cora [20], CiteSeer [20], and PubMed [20] are citation network datasets. Ratings [21] and Roman [21] are two heterophilous graph datasets. Arxiv [22] is the citation network provided by OGB package [22]. For ArXiv dataset, we define papers published before 2015 as in-distribution data, and those published in 2018, 2019, and 2020 as out-of-distribution data. For other datasets, we generate OOD data through structure manipulation, feature interpolation, and label leave-out, following the practice in GNNSafe [5]. The statistics of all these datasets are summarized in Table I.

### B. Baselines

We perform a comparative analysis across two distinct categories of OOD detection methods. The first category includes baseline models originally designed for the visual domain, which operate under the i.i.d. assumption, such as MSP [23], Energy [8], ODIN [24], and Mahalanobis [25]. The second category encompasses methods specifically tailored for node-level OOD detection on graphs, including GKDE [26], GPN [27], and GNNSafe [5].

### C. Implementation Details

We implement both our model and the baselines conducting all experiments on a 24GB NVIDIA GeForce RTX card. The source code is publicly available at the GitHub site[1]. For all datasets, we first train graph convolutional network and save the best model weight according to the valid set. Here, we utilize a 2-layer GCN model as graph encoder. For We employ AUROC (↑), AUPR (↑), and FPR95 (↓) as evaluation metrics for OOD detection, in line with established conventions. We run the experiments three times and record the average and standard deviation of the evaluation metrics. For in-distribution classification performance, we evaluate using accuracy on the test nodes. The baseline implementations are sourced from their respective original publications to ensure a fair comparison.

### D. Experimental Results

The main experimental results are shown in Tables II, III, IV, and V. From these tables, we have the following observations. First, in comparison to existing baselines, including both those originally designed for the visual domain, including MSP, Energy, ODIN, Mahalanobis, and recent node-level graph OOD detection methods, including GKDE, GPN, the proposed SDGSafe model consistently achieves substantial and superior performance gains across all experimental settings. These results validate the overall design efficacy of SDGSafe and underscore the advantage of integrating energy difference propagation into the OOD detection framework. Second, compared to previous GNNSafe methods based on energy smoothing propagation, the proposed SDGSafe model

[1] https://github.com/xiaolongo/SDGSafe

TABLE IV

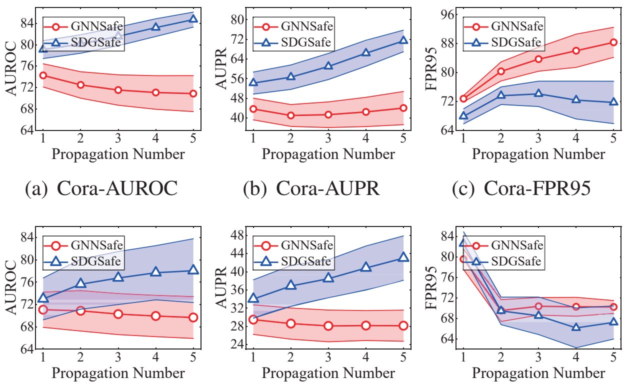
PERFORMANCE COMPARISON WITH BASELINE METHODS IN TERMS OF FPR95 (↓).

| METHOD | Cora@FPR95(↓) | | | CiteSeer@FPR95(↓) | | | PubMed@FPR95(↓) | | | Ratings@FPR95(↓) | | | Roman@FPR95(↓) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label | Structure | Feature | Label |
| MSP | 70.41±2.05 | 47.18±1.87 | 87.64±2.54 | 72.53±2.52 | 53.65±2.15 | 79.96±2.66 | 67.99±3.42 | 49.58±2.25 | 97.25±0.65 | 97.68±1.50 | 82.80±1.05 | 95.45±0.74 | 96.26±0.10 | 84.96±0.29 | 85.17±1.17 |
| Energy | 73.38±3.69 | 54.65±1.69 | 81.23±4.00 | 76.87±5.69 | 64.11±7.42 | 89.84±1.96 | 71.34±4.29 | 61.16±7.63 | 95.86±0.58 | 96.21±2.10 | 85.04±7.27 | 87.16±2.77 | 97.16±0.34 | 74.03±1.19 | 81.03±2.12 |
| ODIN | 98.61±0.11 | 99.63±0.06 | 98.51±0.05 | 99.56±0.16 | 99.86±0.05 | 97.77±0.46 | 99.90±0.03 | 99.92±0.02 | 97.08±1.11 | 99.67±0.02 | 99.89±0.02 | 94.83±0.29 | 60.09±0.43 | 94.12±0.33 | 98.60±0.03 |
| Mahalanobis | 99.45±0.09 | 99.31±0.03 | 79.31±2.59 | 99.57±0.13 | 99.65±0.16 | 70.57±4.51 | 96.94±0.84 | 95.02±0.58 | 95.37±2.18 | 99.55±0.04 | 79.21±0.38 | 91.22±0.34 | 98.28±0.04 | 88.92±0.62 | 90.05±0.28 |
| GKDE | 95.78±0.92 | 95.62±0.90 | 89.50±2.68 | 96.10±1.13 | 96.16±1.26 | 71.18±19.88 | 94.83±0.37 | 94.91±0.60 | 95.00±1.80 | 94.97±0.52 | 94.78±0.48 | 95.38±2.11 | 94.83±0.35 | 95.21±0.05 | 92.87±1.96 |
| GPN | 95.27±1.63 | 96.23±0.57 | 81.00±5.53 | 96.28±0.37 | 96.28±0.37 | 71.70±9.38 | 94.96±0.22 | 94.97±0.27 | 95.21±4.89 | 94.62±0.71 | 94.76±0.40 | 95.17±1.46 | 94.95±0.14 | 95.00±0.24 | 95.13±0.89 |
| GNNSafe | 31.07±1.76 | 24.05±7.93 | 83.67±3.32 | 59.31±2.43 | **52.79±8.23** | 70.37±1.72 | 14.94±9.28 | **19.44±5.16** | 94.36±0.27 | 94.29±3.56 | 76.13±2.67 | 84.77±3.62 | 79.57±0.11 | 61.07±1.13 | 76.50±2.35 |
| SDGSafe | **29.65±2.75** | **22.60±5.86** | **74.07±3.51** | **58.31±2.43** | 53.37±9.98 | **68.52±3.66** | **14.29±9.60** | 19.57±6.39 | **94.12±1.10** | **94.13±2.82** | **76.11±2.28** | **84.67±3.74** | **79.56±0.10** | **60.93±1.12** | **76.48±2.36** |

TABLE V

PERFORMANCE COMPARISON WITH BASELINE METHODS IN TERMS OF AUROC (↑) / AUPR (↑) / FPR95 (↓) ON ARXIV DATASET.

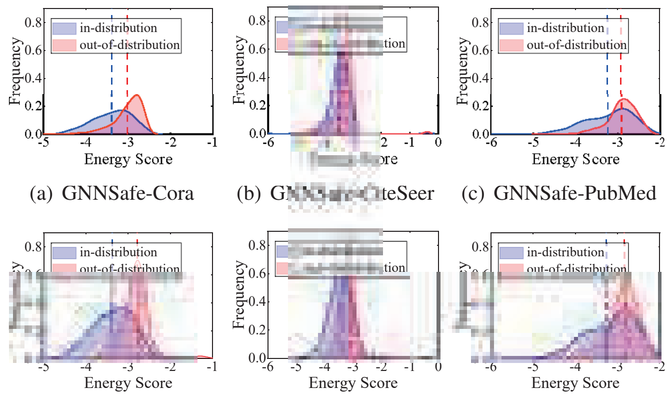
| METHOD | ArXiv-2018 | | | ArXiv-2019 | | | ArXiv-2020 | | |
|---|---|---|---|---|---|---|---|---|---|
| | AUROC(↑) | AUPR(↑) | FPR95(↓) | AUROC(↑) | AUPR(↑) | FPR95(↓) | AUROC(↑) | AUPR(↑) | FPR95(↓) |
| MSP | 61.33±0.12 | 24.28±0.32 | 91.37±0.58 | 62.48±0.31 | 20.41±0.13 | 90.75±0.67 | 64.56±0.11 | 55.53±0.61 | 89.87±0.98 |
| Energy | 62.25±0.43 | 26.11±0.34 | 91.54±0.64 | 63.53±0.56 | 22.13±0.51 | 90.98±0.93 | 66.89±0.27 | 59.44±0.45 | 90.22±0.88 |
| ODIN | 36.27±0.41 | 11.48±0.15 | 99.27±0.07 | 35.22±0.60 | 8.71±0.15 | 99.24±0.09 | 31.17±0.15 | 28.43±0.24 | 99.88±0.02 |
| Mahalanobis | 56.89±0.70 | 19.19±0.32 | 95.26±0.32 | 55.81±0.70 | 14.64±0.23 | 96.28±0.40 | 56.22±0.67 | 43.79±0.41 | 97.82±0.34 |
| GKDE | 45.90±2.35 | 14.41±1.23 | 95.72±0.54 | 46.75±0.33 | 12.16±0.53 | 96.58±0.60 | 44.54±0.04 | 36.31±0.83 | 96.57±1.04 |
| GPN | OOM | OOM | OOM | OOM | OOM | OOM | OOM | OOM | OOM |
| GNNSafe | 67.00±0.29 | 33.72±0.28 | 90.54±0.43 | 68.98±0.39 | 31.14±0.97 | 89.41±0.74 | 78.12±0.23 | 75.22±0.15 | 85.35±1.28 |
| SDGSafe | **67.28±0.37** | **34.39±0.37** | **90.46±0.67** | **69.36±0.44** | **32.05±1.04** | **89.22±0.94** | **78.22±0.30** | **75.48±0.29** | **85.38±1.55** |




(a) Cora-AUROC (b) Cora-AUPR (c) Cora-FPR95

(d) CiteSeer-AUROC (e) CiteSeer-AUPR (f) CiteSeer-FPR95

Fig. 3. Impact of propagation numbers.




(a) GNNSafe-Cora (b) GNNSafe-CiteSeer (c) GNNSafe-PubMed

(d) SDGSafe-Cora (e) SDGSafe-CiteSeer (f) SDGSafe-PubMed

Fig. 4. Comparison of energy distribution.

achieves a significant performance improvement. For example, when the OOD type is defined by label shifts, SDGSafe achieves a 10% improvement in AUROC value compared to GNNSafe on Cora. These results strongly demonstrate the effectiveness of introducing energy difference propagation.

### E. Comprehensive Analysis

To further analysis the proposed SDGSafe model, we perform additional experiments in this subsection.

*1) Impact of propagation number:* To investigate the impact of the propagation number, we perform a comparative analysis by evaluating values ranging from 1 to 5 for the out-of-distribution (OOD) type defined by label shifts, on both the Cora and CiteSeer datasets. The corresponding results are illustrated in Figure 3. As shown in the figure, increasing the number of energy propagations leads to consistent improvements in AUROC, AUPR, and FPR95 metrics for the proposed SDGSafe method. In contrast, the baseline GNNSafe model exhibits performance degradation under the same conditions. These findings further validate the effectiveness of the proposed energy differential propagation mechanism in enhancing OOD detection robustness.

*2) Comparison of energy distribution:* To visualize and compare the energy distributions, we statistically analyze the frequency density plots of negative energy scores for GNNSafe and SDGSafe across the Cora, CiteSeer, and PubMed datasets. The OOD type is defined by label shifts. The results are presented in Figure 4. As qualitatively observed in the figure, the proposed SDGSafe model produces a more distinct separation between in-distribution and out-of-distribution data compared to GNNSafe, demonstrating its enhanced discriminative capability. This improvement can be attributed primarily to the incorporation of the energy difference propagation mechanism within the SDGSafe framework.

## VI. Conclusion

In this work, we propose a hybrid energy propagation framework for graph out-of-distribution (OOD) detection. To overcome the limitation of pure employing smoothed energy propagation, which fails when nodes from different distributions are directly connected, we augment it with an energy difference propagation mechanism. The resulting method, SDGSafe, adaptively applies smoothing or difference propagation based on the distribution heterophily of a node's neighbors. Experiments show that SDGSafe substantially outperforms smoothed energy-only baselines, confirming the effectiveness of the integrated approach. Future work may extend this paradigm to dynamic graphs and explore adaptive mechanisms for automatically identifying distribution boundaries.

## References


[1] X. Fan, M. Gong, Y. Wu, A. Qin, and Y. Xie, "Propagation enhanced neural message passing for graph representation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 02, pp. 1952–1964, 2023.

[2] X. Fan, M. Gong, Y. Wu, Z. Tang, and J. Liu, "Neural gaussian similarity modeling for differential graph structure learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 11 919–11 926.

[3] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He *et al.*, "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Transactions on Recommender Systems*, pp. 1–51, 2023.

[4] L. Wu, Y. Chen, K. Shen, X. Guo, H. Gao, S. Li, J. Pei, B. Long *et al.*, "Graph neural networks for natural language processing: A survey," *Foundations and Trends in Machine Learning*, vol. 16, no. 2, pp. 119–328, 2023.

[5] Q. Wu, Y. Chen, C. Yang, and J. Yan, "Energy-based out-of-distribution detection for graph neural networks," in *Proceedings of the International Conference on Learning Representations*, 2023.

[6] Y. Liu, K. Ding, H. Liu, and S. Pan, "GOOD-D: On unsupervised graph out-of-distribution detection," in *Proceedings of the ACM International Conference on Web Search and Data Mining*, 2023, pp. 339–347.

[7] X. Shen, Y. Wang, K. Zhou, S. Pan, and X. Wang, "Optimizing OOD detection in molecular graphs: A novel approach with diffusion models," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 2640–2650.

[8] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 33, 2020, pp. 21 464–21 475.

[9] S. Yang, B. Liang, A. Liu, L. Gui, X. Yao, and X. Zhang, "Bounded and uniform energy-based out-of-distribution detection for graphs," 2025.

[10] T. Bao, Q. Wu, Z. Jiang, Y. Chen, J. Sun, and J. Yan, "Graph out-of-distribution detection goes neighborhood shaping," in *Proceedings of the International Conference on Machine Learning*, 2024, pp. 2923–2943.

[11] Z. Gong and Y. Sun, "An energy-centric framework for category-free out-of-distribution node detection in graphs," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 908–919.

[12] X. Fan, M. Gong, Y. Wu, M. Zhang, H. Li, and X. Jiang, "CCGIB: A cross-channel graph information bottleneck principle," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 5, pp. 9488–9499, 2024.

[13] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the International Conference on Learning Representations*, 2017.

[14] C. Huang, M. Li, F. Cao, H. Fujita, Z. Li, and X. Wu, "Are graph convolutional networks with random weights feasible?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 2751–2768, 2022.

[15] L. Wang, D. He, H. Zhang, Y. Liu, W. Wang, S. Pan, D. Jin, and T.-S. Chua, "Goodat: Towards test-time graph out-of-distribution detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 14, 2024, pp. 15 537–15 545.

[16] M. Wu, S. Pan, and X. Zhu, "OpenWGL: Open-world graph learning," in *Proceedings of the IEEE International Conference on Data Mining*, 2020, pp. 681–690.

[17] Y. Guo, C. Yang, Y. Chen, J. Liu, C. Shi, and J. Du, "A data-centric framework to endow graph neural networks with out-of-distribution detection ability," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 638–648.

[18] X. Luo, Y. Zhao, Y. Qin, W. Ju, and M. Zhang, "Towards semi-supervised universal graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 1, pp. 416–428, 2023.

[19] L. Ma, Y. Sun, K. Ding, Z. Liu, and F. Wu, "Revisiting score propagation in graph out-of-distribution detection," *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 37, pp. 4341–4373, 2024.

[20] Z. Yang, W. Cohen, and R. Salakhudinov, "Revisiting semi-supervised learning with graph embeddings," in *Proceedings of the International Conference on Machine Learning*, 2016, pp. 40–48.

[21] O. Platonov, D. Kuznedelev, M. Diskin, A. Babenko, and L. Prokhorenkova, "A critical look at the evaluation of gnns under heterophily: Are we really making progress?" in *Proceedings of the International Conference on Learning Representations*, 2023.

[22] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec, "Open graph benchmark: Datasets for machine learning on graphs," in *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 33, 2020, pp. 22 118–22 133.

[23] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proceedings of the International Conference on Learning Representations*, 2017.

[24] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *Proceedings of the International Conference on Learning Representations*, 2018.

[25] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Proceedings of the International Conference on Neural Information Processing Systems*, 2018, pp. 7167–7177.

[26] X. Zhao, F. Chen, S. Hu, and J.-H. Cho, "Uncertainty aware semi-supervised learning on graph data," *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 33, 2020.

[27] M. Stadler, B. Charpentier, S. Geisler, D. Zügner, and S. Günnemann, "Graph posterior network: Bayesian predictive uncertainty for node classification," *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 34, pp. 18 033–18 048, 2021.