HyperGOOD: Towards Out-of-Distribution Detection in Hypergraphs

Tingyi Cai[1,2,4], **Yunliang Jiang**[1,2,4,5*], **Ming Li**[1], **Changqin Huang**[3], **Yujie Fang**[1,4], **Chengling Gao**[1,4], **Zhonglong Zheng**[4]

[1]Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China
[2]China-Mozambique Belt and Road Joint Laboratory on Smart Agriculture, Zhejiang Normal University, Jinhua, China
[3]College of Education, Zhejiang University, Hangzhou, China
[4]School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China
[5]School of Information Engineering, Huzhou University, Huzhou, China
{tingyicai, mingli, yjfang, chl_gao, zhonglong}@zjnu.edu.cn, jyl2022@zjnu.cn, cqhuang@zju.edu.cn

Abstract

Out-of-distribution (OOD) detection plays a critical role in ensuring the robustness of machine learning models in open-world settings. While extensive efforts have been made in vision, language, and graph domains, the challenge of OOD detection in hypergraph-structured data remains unexplored. In this work, we formalize the problem of hypergraph out-of-distribution (HOOD) detection, which aims to identify nodes or hyperedges whose high-order relational contexts differ significantly from those seen during training. We propose HyperGOOD, a unified energy-based detection framework that integrates multi-scale spectral decomposition with structure-aware uncertainty propagation. By preserving both low- and high-frequency signals and diffusing uncertainty across the hypergraph, HyperGOOD effectively captures subtle and relationally entangled anomalies. Experimental results on nine hypergraph datasets demonstrate the effectiveness of our approach, establishing a new foundation for robust hypergraph learning under distributional shifts.

Code — https://github.com/ca1man-2022/HyperGOOD

1 Introduction

Hypergraphs provide a powerful and flexible representation for modeling high-order relationships among entities, where a single hyperedge can connect more than two nodes (Murgas, Saucan, and Sandhu 2022; Sun et al. 2023). This modeling capability makes hypergraphs well-suited for a variety of real-world systems, including biological networks, recommendation systems, and medical knowledge graphs. Hypergraph neural networks (HNNs) (Feng et al. 2024; Gao et al. 2024; Antelmi et al. 2023), which generalize graph neural networks to operate on hypergraph structures, have shown promising performance in a wide range of tasks by effectively capturing high-order dependencies among nodes.

Most existing HNNs (Huang and Yang 2021; Wang et al. 2023; Chien et al. 2022), however, are developed under the closed-world assumption—namely, that the training and testing data are drawn from the same distribution. This assumption is frequently violated in real-world scenarios, where data can evolve over time or exhibit unexpected variations. As depicted in the left panel of Figure 1, models trained under closed-world settings may become overly confident when encountering unfamiliar inputs in open-world environments, resulting in severe performance degradation (Wang et al. 2024c; Ma et al. 2023). While out-of-distribution (OOD) detection has been extensively explored in image, text, and graph data (Yang et al. 2024a; Wang et al. 2024a; Cai et al. 2025b), no existing method has been developed to handle the unique structural and semantic characteristics of hypergraphs. This gap is particularly concerning given the rising adoption of HNNs in high-stakes applications. To address this gap, we introduce the task of **hypergraph out-of-distribution (HOOD) detection**, which aims to identify nodes or hyperedges whose high-order relational contexts deviate from those observed during training. This task presents unique challenges that distinguish it from conventional OOD detection. First, nodes in hypergraphs often participate in overlapping and semantically diverse hyperedges, making it difficult to isolate distributional shifts due to entangled relational contexts. Second, the feature aggregation process in HNNs tends to smooth over local deviations (Li et al. 2025b), suppressing subtle anomalies. Third, as illustrated in the right panel of Figure 1, OOD signals in real-world hypergraphs, such as rare syndromes or unconventional treatment combinations in a medical case-sharing network (Wu et al. 2023a; Solomon 2006), may not be localized to single nodes but instead emerge from inconsistencies

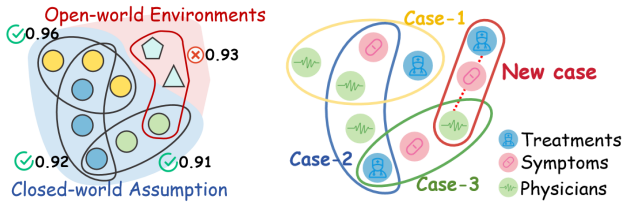

Figure 1: Illustration of HOOD detection scenarios in real-world hypergraphs. *Left:* Risks under the closed-world assumption. *Right:* In a medical case sharing hypergraph, OOD cases may involve rare syndromes or atypical treatment plans.

*Corresponding author.

across multiple structural scales.

In this paper, we propose **HyperGOOD**, a unified framework for HOOD detection that integrates multi-scale hypergraph representation learning with structure-aware uncertainty modeling. HyperGOOD begins by applying framelet-based spectral decomposition to extract multi-scale features, preserving both low-frequency semantic regularities and high-frequency structural variations that are critical for capturing subtle distributional mismatches. Rather than collapsing these signals into a single representation, HyperGOOD maintains their multi-scale form and applies an energy-based scoring mechanism to estimate the uncertainty of each node. To further enhance detection, we introduce a structure-aware diffusion module that propagates energy scores through the hypergraph, allowing relational reinforcement to highlight weak or ambiguous anomalies that may not be detectable in isolation. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness of HyperGOOD in identifying OOD instances across diverse hypergraph scenarios. Extensive experiments on nine real-world hypergraph datasets demonstrate that HyperGOOD consistently outperforms several baselines in detecting out-of-distribution samples under diverse settings.

In summary, our primary contributions are three-fold:

- **New Problem Formulation.** We formally define HOOD detection, i.e., out-of-distribution detection for hypergraphs, and analyze its unique challenges arising from high-order relational shifts.
- **Unified Detection Framework.** We propose HyperGOOD, a principled multi-scale detection framework that combines framelet-based spectral decomposition with uncertainty estimation and structure-aware propagation for HOOD detection.
- **Structure-aware Uncertainty Modeling.** We design an energy scoring mechanism that leverages both low- and high-pass framelet responses and propagates these scores through the hypergraph to uncover distributed and relationally entangled anomalies.

2 Related Works

Hypergraph neural networks. Hypergraph learning extends graph models to capture high-order dependencies via hyperedges. Early works like HGNN (Feng et al. 2019) and HyperGCN (Yadati et al. 2019) adapt spectral and message-passing methods, while recent models such as AllSet (Chien et al. 2022), UniGNN (Huang and Yang 2021), and HyperSAGE (Arya et al. 2024) enhance aggregation and generalization. As hypergraph methods evolve, they are applied across various domains, including bioinformatics, recommender systems, and visual scene understanding, where group-wise interactions naturally occur (Yan et al. 2025; Xie et al. 2025; Bai, Zhang, and Torr 2021; Feng et al. 2019).

However, they all assume that training and test data follow the same distribution. This assumption does not hold in many real-world scenarios where unseen structures or cases often appear. So far, no prior work addresses the problem of detecting OOD data in hypergraphs.

Out-of-distribution detection. OOD detection aims to identify inputs that deviate from the training distribution, which is essential in high-stakes domains like healthcare, autonomous systems, and scientific discovery. Traditional methods (Yang et al. 2024a; Xue et al. 2024; Zhang et al. 2024) in vision and NLP rely on statistical distances, uncertainty estimation, or generative models. More recently, OOD detection has been explored in graph learning (Cai et al. 2025b; Ju et al. 2024), focusing on node- or graph-level tasks over pairwise graphs. These methods (Wu et al. 2023b; Shen et al. 2024; Liu et al. 2023; Wang et al. 2024b; Yang et al. 2024b) leverage embedding distances, energy functions, or confidence calibration. However, most operate on simple graphs and often reduce hypergraphs to pairwise forms, leading to substantial structural information loss (Millán et al. 2025; Wang and Kleinberg 2024).

To bridge this gap, we propose the first OOD detection framework specifically tailored to the unique structure of hypergraph data.

3 Problem Formulation

Motivating scenario. Many real-world applications involve high-order interactions among entities, which are naturally modeled by hypergraphs. In a medical case sharing platform (Solomon 2006), for example, each case forms a hyperedge connecting symptoms, treatments, and physicians (as shown in Figure 1), capturing typical relational patterns seen during training. However, in open-world scenarios, test-time cases may deviate from these patterns, such as rare symptom combinations or atypical treatments, leading to relational contexts not observed during training. These OOD instances often stem from subtle changes in high-order compositions that cannot be detected via node features or pairwise links alone. This motivates HOOD detection, which seeks to identify nodes or hyperedges whose relational contexts diverge from the training distribution.

In realistic open-world environments, HOOD detection serves two important purposes. First, it enables the discovery of novel and meaningful structural patterns, such as emerging disease presentations or uncommon treatment pathways. Second, it improves the safety and reliability of deployed systems by identifying structurally inconsistent inputs that could otherwise lead to overconfident or erroneous predictions.

Notation. We consider a single hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{H})$, where $\mathcal{V} = \{v_1, v_2, \cdots, v_n\}$ denotes the set of n nodes, and $\mathcal{E} = \{e_1, e_2, \cdots, e_m\}$ is the set of hyperedges. The incidence matrix $\mathbf{H} \in \{0,1\}^{n \times m}$ specifies node–hyperedge membership, where $H_{ij} = 1$ if node $v_i \in e_j$. Each node v_i has a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ and an optional label $y_i \in \{1, 2, \cdots, C\}$. We define the local structural context of node v_i as the set of incident hyperedges $\mathcal{N}_i = \{e_j \in \mathcal{E} \mid H_{ij} = 1\}$. Rather than considering distributions over entire hypergraphs, we model the training data as a set of node–context pairs $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, \mathcal{N}_i)\}_{i \in \mathcal{V}_{\text{train}}}$ sampled from an in-distribution (IND) D_{in}. At test time, the model observes samples from $\mathcal{D}_{\text{test}} = \{(\mathbf{x}_j, \mathcal{N}_j)\}_{j \in \mathcal{V}_{\text{test}}}$,

which may include OOD nodes drawn from an unknown distribution D_{out}.

Definition 1 (HOOD Detection). *Given a hypergraph-structured dataset possibly containing OOD samples, the goal is to learn a decision function that accepts inputs from the training distribution while rejecting those from unknown distributions. Formally, the detection objective is:*

$$\hat{y} = \begin{cases} f(\mathbf{x}, \mathcal{N}), & \text{if } \mathbf{x} \sim D_{\text{in}}, \\ \text{reject}, & \text{if } \mathbf{x} \sim D_{\text{out}}, \end{cases} \tag{1}$$

where $f(\cdot)$ denotes a predictive model trained on IND data. The OOD distribution D_{out} represents samples with semantic shifts relative to D_{in}.

To the best of our knowledge, we are the first to formally define and address the problem of HOOD detection, which aims to identify node-level distributional anomalies under high-order relational structures.

Remark. HOOD detection goes beyond standard classification, requiring sensitivity to subtle distribution shifts embedded in overlapping high-order structures. Furthermore, real-world hypergraph applications frequently exhibit complex OOD shifts including heterophilic, heterogeneous, and temporally evolving structures (Kim et al. 2024; Khan et al. 2025). These remain underexplored in current HOOD research and present important challenges for future work.

4 Methodology

In this section, we introduce our proposed novel **Hyperg**raph **O**ut-**o**f-**D**istribution Detection (**HyperGOOD**) method, which leverages spectral framelet decomposition and structure-aware energy propagation. The overall architecture consists of three key components: (1) Framelet-based multi-scale representation to capture high- and low-frequency patterns. (2) Energy-based uncertainty modeling to quantify node-level OOD likelihoods. (3) Structure-aware energy propagation over hypergraphs to exploit high-order consistency for robust OOD detection. An overview is illustrated in Figure 2.

4.1 Multi-scale Hypergraph Representation

This section presents a multi-scale hypergraph representation approach based on framelet transforms, following the framework proposed by Li et al. (2025a). To facilitate understanding of the construction of hypergraph framelets, we briefly revisit the necessary theoretical foundations. For further details, we refer the reader to the original work.

Let $\mathcal{L} = I - \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{D}_e^{-1}\mathbf{H}^\top\mathbf{D}_v^{-1/2}$ denote the normalized hypergraph Laplacian, where $\mathbf{D}_v$ and $\mathbf{D}_e$ are the degree matrices of the nodes and hyperedges, respectively. The Laplacian admits an eigendecomposition $\mathcal{L} = \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^\top$, where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$ contains the orthonormal eigenvectors and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_n)$ stores the corresponding eigenvalues. Let $s \in \{1, \ldots, S\}$ be the scale level for framelet decomposition, for a reference node $v_p \in \mathcal{V}$, the framelet containing both low-pass and high-pass components are defined as:

$$\Phi_{s,v_p}(v_i) = \sum_{j=1}^{n} \widehat{\phi}\left(\frac{\sigma_j}{2^s}\right) q_j(v_p)\, q_j(v_i), \tag{2}$$

$$\Psi^{(l)}_{s,v_p}(v_i) = \sum_{j=1}^{n} \widehat{\psi}_l\left(\frac{\sigma_j}{2^s}\right) q_j(v_p)\, q_j(v_i), l = 1, \ldots, L. \tag{3}$$

where $\widehat{\phi}$ and $\widehat{\psi}$ are spectral filters derived from a filter bank $\mathcal{F} = \{g_0; g_1, \ldots, g_L\}$, with $g_0(\cdot)$ representing is the low-pass filter and $g_l(\cdot)$ $(l = 1, \ldots, L)$ the l-th high-pass filter. These filters satisfy the refinement relations in the Fourier domain:

$$\widehat{\phi}(2\xi) = \widehat{g}_0(\xi)\widehat{\phi}(\xi), \widehat{\psi}_l(2\xi) = \widehat{g}_l(\xi)\widehat{\phi}(\xi), \forall \xi \in \mathbb{R}. \tag{4}$$

Given a hypergraph signal $\mathbf{X} \in \mathbb{R}^{n\times d}$, the low-pass and high-pass framelet coefficients are computed as:

$$\mathbf{z}_0 = \mathbf{Q} \cdot \widehat{\phi}\left(\frac{\mathbf{\Sigma}}{2}\right) \cdot \mathbf{Q}^\top\mathbf{X}, \tag{5}$$

$$\mathbf{w}_s^{(l)} = \mathbf{Q} \cdot \widehat{\psi}_l\left(\frac{\mathbf{\Sigma}}{2^{s+1}}\right) \cdot \mathbf{Q}^\top\mathbf{X}, \quad l = 1, \ldots, L. \tag{6}$$

Let $\mathcal{D}$ and $\mathcal{D}_{l,s}$ denote the decomposition operators such that $\mathbf{z}_0 = \mathcal{D}_{0,S}\mathbf{X}$ and $\mathbf{w}_s^{(l)} = \mathcal{D}_{l,s}\mathbf{X}$. The associated transform matrices are given by:$\mathcal{D}_{0,S} = \mathbf{Q}\,\widehat{g}_l(\frac{\mathbf{\Sigma}}{2^M})\,\mathbf{Q}^\top$ and $\mathcal{D}_{l,s} = \mathbf{Q}\,\widehat{g}_l(\frac{\mathbf{\Sigma}}{2^{M+s-1}})\,\widehat{g}_0(\frac{\mathbf{\Sigma}}{2^{M+s-2}})\cdots\widehat{g}_0(\frac{\mathbf{\Sigma}}{2^M})\mathbf{Q}^\top$. Here, M is chosen sufficiently large to ensure the spectrum of $\mathcal{L}$ lies within the domain of the filters, i,e, $\sigma \le 2^M\pi$.

In practice, to avoid expensive eigendecomposition, we approximate the spectral filtering via Chebyshev polynomials. Let $\mathcal{T}_0, \ldots, \mathcal{T}_r$ denote the Chebyshev polynomial approximations for the filter bank $\mathcal{F}$, the decomposition operators are then approximated as:expressed as:

$$\begin{aligned} \mathcal{D}_{0,S} &= \mathcal{T}_0(\frac{\mathcal{L}}{2^{M+s-1}})\cdots\mathcal{T}_0(\frac{\mathcal{L}}{2^M}), \\ \mathcal{D}_{l,1} &= \mathcal{T}_l\left(\frac{\mathcal{L}}{2^M}\right), \\ \mathcal{D}_{l,s} &= \mathcal{T}_l(\frac{\mathcal{L}}{2^{M+s-1}})\mathcal{T}_0(\frac{\mathcal{L}}{2^{M+s-2}})\cdots\mathcal{T}_0(\frac{\mathcal{L}}{2^M}). \end{aligned} \tag{7}$$

With these components, we define a multi-scale hypergraph convolution operator that integrates framelet-based representations into the node feature propagation process. Inspired by residual propagation strategies (Chen et al. 2020), the ℓ-th layer of the hypergraph neural operator is formulated as:

$$\begin{aligned} \mathbf{X}^{(\ell+1)} = \eta[&(1-\alpha_\ell)\sum_{(l,s)\in\Lambda} \mathcal{D}_{l,s}^\top \,\text{diag}(\boldsymbol{\theta}_{l,s})\,\mathcal{D}_{l,s}\,\mathbf{X}^{(\ell)} \\ &+ \alpha_\ell\mathbf{X}^{(0)}] \cdot \left[(1-\beta_\ell)\mathbf{I} + \beta_\ell\mathbf{\Theta}^{(\ell)}\right], \end{aligned} \tag{8}$$

where $\boldsymbol{\theta}_{l,s} \in \mathbb{R}^n$ is a learnable node-wise filter weight, η denotes a nonlinear activation function, and α controls the residual connection to the initial input. The set $\Lambda = \{(l, s) :$

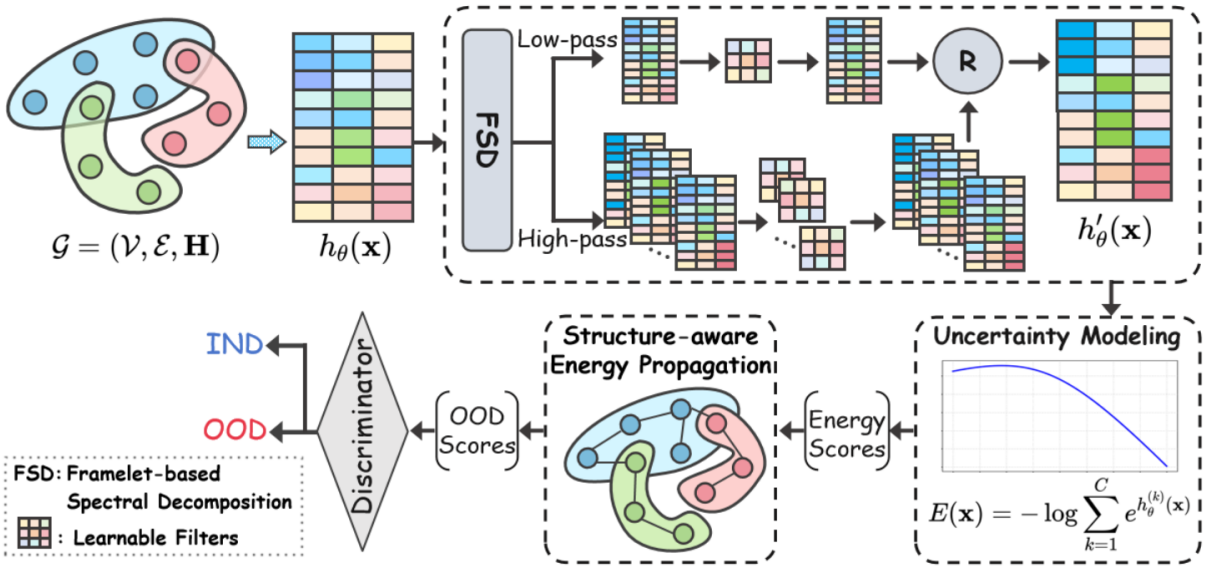

Figure 2: The framework of HyperGOOD. The overall architecture consists of three key components: (1) Framelet-based multi-scale representation to capture high- and low-frequency patterns. (2) Energy-based uncertainty modeling to quantify node-level OOD likelihoods. (3) Structure-aware energy propagation over hypergraphs to exploit high-order consistency for robust OOD detection.

$l = 1, \ldots, L,\ s = 1, \ldots, S\} \cup \{(0, S)\}$ indexes all framelet components used in the aggregation.

This operator forms the core of our model for capturing multi-scale and frequency-sensitive patterns over hypergraphs. By selectively aggregating information across different spectral bands and scales, the model is equipped to detect non-local anomalies and distributional shifts, which are characteristics central to the task of HOOD detection.

In summary, our multi-scale hypergraph representation learning leverages both low-pass and high-pass filters to capture complementary information: coarse-grained semantic trends and fine-grained local inconsistencies. This property is crucial for effectively identifying OOD samples embedded within high-order relational structures.

4.2 Energy-based Uncertainty Modeling

We extract multi-scale representations for each node through a multi-layer framelet hypergraph network. Let $h_\theta(\mathbf{x}_i, \mathcal{N}_i) \in \mathbb{R}^C$ represent the predicted logits output by node v_i, where C is the number of classes and θ denotes trainable parameters. The standard prediction is obtained via the softmax function:

$$p(y_i = c \mid \mathbf{x}_i, \mathcal{N}_i) = \frac{\exp(h_\theta^{(c)}(\mathbf{x}_i, \mathcal{N}_i))}{\sum_{k=1}^{C} \exp(h_\theta^{(k)}(\mathbf{x}_i, \mathcal{N}_i))}. \quad (9)$$

Energy-based models (EBMs) (Yann et al. 2006; Liu et al. 2020) are statistical frameworks that moedl variable configurations through an energy function. Following the formulation of EBMs, we define a probability density $p(\mathbf{x}_i, \mathcal{N}_i)$ using logits through the Boltzmann distribution:

$$\begin{aligned} p(y_i = c \mid \mathbf{x}_i, \mathcal{N}_i) &= \frac{\exp(-E(\mathbf{x}_i, \mathcal{N}_i, c))}{\exp(-E(\mathbf{x}_i, \mathcal{N}_i))} \\ &= \frac{\exp(-E(\mathbf{x}_i, \mathcal{N}_i, c))}{\sum_{k=1}^{C} \exp(-E(\mathbf{x}_i, \mathcal{N}_i, k))}. \end{aligned} \quad (10)$$

Thus, the equivalence between EBM and our hypergraph classifier is established by setting the energy as the predicted logit value, that is, energy function:

$$E(\mathbf{x}_i, \mathcal{N}_i, c) = -h_\theta^{(c)}(\mathbf{x}_i, \mathcal{N}_i). \quad (11)$$

To quantify the total uncertainty of a node–context pair regardless of label, we define the *free energy* by marginalizing over possible labels:

$$\begin{aligned} E(\mathbf{x}_i, \mathcal{N}_i) &= -\log \sum_{k=1}^{C} \exp(-E(\mathbf{x}_i, \mathcal{N}_i, k)) \\ &= -\log \sum_{k=1}^{C} \exp(h_\theta^{(k)}(\mathbf{x}_i, \mathcal{N}_i)). \end{aligned} \quad (12)$$

This free energy score reflects the total confidence the model assigns to the input. A low score indicates high model certainty, while a high value reflects OOD behavior.

We adopt the free energy in Eq. (12) as our energy scoring function:

$$\text{Energy Score}(v_i) := -\log \sum_{k=1}^{C} \exp(h_\theta^{(k)}(\mathbf{x}_i, \mathcal{N}_i)). \quad (13)$$

This score is simple to compute, requires no additional training objective, and serves as a natural indicator of epistemic uncertainty in hypergraph-structured data.

4.3 Structure-aware Energy Propagation

While the free energy score $E(\mathbf{x}_i, \mathcal{N}_i)$ captures node-wise uncertainty, many OOD signals in hypergraphs arise from subtle structural inconsistencies that are not locally isolatable. To address this, we introduce a *structure-aware energy propagation* mechanism designed to enhance and reveal such inconsistencies by diffusing uncertainty scores across the hypergraph topology.

Let $\mathbf{s}^{(t)} \in \mathbb{R}^n$ denote the energy vector at iteration t. This mechanism iteratively updates node-level energy scores through neighborhood-based refinement:

$$\mathbf{s}^{(t+1)} = \gamma \cdot \mathbf{s}^{(t)} + (1-\gamma) \cdot \mathbf{P}\mathbf{s}^{(t)}, \tag{14}$$

where $\mathbf{P} := \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{D}_e^{-1}\mathbf{H}^\top\mathbf{D}_v^{-1/2}$ denotes the normalized hypergraph propagation matrix, and γ balances self-retention and structural smoothing. The goal is to reinforce local consensus in energy while preserving deviations from structurally inconsistent regions, thereby exposing distributed OOD patterns.

We initialize $\mathbf{s}^{(0)}$ with the node-wise energy scores derived from framelet-based representations. After T iterations of propagation, we obtain the final structure-aware energy scores $\mathbf{s}^{(T)}$, and define:

$$\text{OOD Score}(v_i) := s_i^{(T)}. \tag{15}$$

Lemma 1. *Let $\mathbf{s}^{(t)}$ be the node-wise energy vector at iteration t, and define the local deviation for node v_i as:*

$$\delta_i^{(t)} := s_i^{(t)} - \sum_j P_{ij} s_j^{(t)}.$$

Then under the update rule in Eq. (14)*, the deviation $\delta_i^{(t)}$ shrinks over iterations:*

$$|\delta_i^{(t+1)}| < |\delta_i^{(t)}|, \quad \textit{for all } \gamma \in (0,1).$$

That is, node-level energy scores progressively align with their structural neighborhood over propagation.

The propagation operator in Eq. (14) encourages local consistency by continuously reducing the discrepancy between each node's energy and the aggregated energy of its neighbors. For IND nodes residing in coherent structural regions, this results in stable or more confident energy estimates. In contrast, OOD nodes whose energy patterns are misaligned with their local structure retain larger residuals during propagation. Over time, these differences are accentuated, effectively amplifying the energy contrast between IND and OOD nodes. The proof process is provided in Appendix [1] B.

Remark. Although such propagation processes may risk *oversmoothing* (Rusch, Bronstein, and Mishra 2023), especially in densely connected hypergraphs, our multi-scale representation method preserves high-frequency details prior to propagation. This allows the propagation step to reinforce structural consistency without erasing subtle local shifts.

[1]Appendix is available at https://github.com/ca1man-2022/HyperGOOD/blob/main/Appendix_HyperGOOD.pdf.

4.4 Optimization Objective

Our model is trained under a supervised node classification setting, using the standard cross-entropy loss over IND samples. To enhance the reliability of energy scores and suppress extreme values, we introduce a lightweight regularization term $\mathcal{L}_{\text{UB}}$ (Yang et al. 2024b). This term constrains the variation in the norm and total mass of prediction logits, promoting boundedness and discouraging saturation across both IND and potential OOD samples. The regularizer encourages the logits to exhibit consistent magnitude and distributional balance, thereby stabilizing energy estimation and improving downstream uncertainty-based decisions.

The final optimization objective is given by:

$$\begin{aligned}\mathcal{L} =& \mathcal{L}_{\text{NLL}} + \lambda \mathcal{L}_{reg} \\ =& \sum_{i \in D_{in}} \left(-h_\theta^{(y_i)}(\mathbf{x}_i, \mathcal{N}_i) + \log \sum_{k=1}^{C} \exp(h_\theta^{(k)}(\mathbf{x}_i, \mathcal{N}_i)) \right) \\ &+ \lambda \cdot \mathcal{L}_{\text{UB}},\end{aligned} \tag{16}$$

where λ controls the strength of regularization.

5 Experiments

In this section, we evaluate the effectiveness of HyperGOOD on the HOOD detection task using nine benchmark datasets. The performance of HyperGOOD is compared against several classic OOD detection methods and graph OOD detection methods.

5.1 Datasets and Experimental Setup

As the first method designed for the HOOD task, we conduct comprehensive experiments across nine diverse real-world hypergraph datasets (`Cora`, `Citeseer`, `PubMed`, `Cora-CA`, `Senate`, `House`, `NTU2012`, `ModelNet40`, `Mushroom`) to validate the generality and effectiveness of our approach. These datasets cover domains such as citation networks, co-authorship, political voting, 3D vision, and semantic classification (Yadati et al. 2019; Fowler 2006; Chodrow, Veldt, and Benson 2021; Chen et al. 2003; Wu et al. 2015; Dua, Graff et al. 2017).

For each dataset, OOD data is generated via standard perturbation strategies, including feature and label noise (Wu et al. 2023b; Cai et al. 2025a). We report area under curve (AUC) scores for all methods and run all experiments on an NVIDIA RTX A6000 GPU with 48 GB memory. Detailed dataset statistics and setup are provided in Appendix A.

We compare HyperGOOD with four representative generic baselines (MSP (Hendrycks and Gimpel 2016), ODIN (Liang, Li, and Srikant 2018), Mahalanobis (Lee et al. 2018), and Energy (Liu et al. 2020)) and two recent graph-based OOD detection methods (GNNSafe (Wu et al. 2023b) and NODESAFE (Yang et al. 2024b)). For fair comparisons, all generic methods are implemented with a standard multi-layer perceptron as the backbone, except for Energy, where we adopt HGNN (Feng et al. 2019) to allow fairer comparison under hypergraph inputs. For graph-based baselines, which assume 2-graph inputs, we convert each hypergraph

Datasets	Cora-CA	Cora	Citeseer	Pubmed	Senate	House	NTU2012	ModelNet40	Mushroom
MSP	0.4502	0.4508	0.4379	0.4623	0.1492	0.1418	0.4322	0.4081	0.4099
ODIN	0.4382	0.4372	0.4291	0.5328	0.8521	0.8506	0.5051	0.5900	0.5858
Mahalanobis	0.4746	0.4751	0.4776	0.4470	0.1325	0.1203	0.4210	0.3979	0.4464
Energy	0.5184	0.5033	0.5006	0.4947	0.5373	0.5172	0.5106	0.5090	0.4947
GNNSafe	0.5520	0.5047	0.4956	**0.5397**	0.5003	0.4962	0.6179	0.6540	0.6364
NODESAFE	0.5699	0.4980	0.4924	0.5353	0.5177	0.6563	0.6016	0.6349	0.6241
HyperGOOD	**0.6337**	**0.5792**	**0.5917**	0.5323	**0.9954**	**0.9832**	**0.6741**	**0.7148**	**0.7984**

Table 1: OOD detection results in terms of AUC (↑). The best and second-best results are highlighted with bold and underline, respectively. (*Feature-perturbation)*

Datasets	Cora-CA	Cora	Citeseer	Pubmed	Senate	House	NTU2012	ModelNet40	Mushroom
MSP	0.5000	0.5000	0.5000	0.5015	0.5000	0.5000	0.5000	0.5000	0.5000
ODIN	0.4002	0.4002	0.3819	0.4944	0.5154	0.5101	0.4448	0.4999	0.4970
Mahalanobis	**0.5402**	**0.5402**	**0.5452**	0.5055	0.4877	0.4508	0.5112	0.4878	0.5027
Energy	0.5103	0.4988	0.4941	0.4958	0.5373	0.5172	0.5079	0.5078	0.4980
GNNSafe	0.4914	0.4727	0.4609	0.5032	0.5004	0.4956	0.4961	0.4987	0.3073
NODESAFE	0.4949	0.4581	0.4416	0.4979	0.5003	0.5011	0.4939	0.4994	0.5016
HyperGOOD	0.5313	0.5126	0.5101	**0.5063**	**0.9563**	**0.8877**	**0.5298**	**0.5205**	**0.5480**

Table 2: OOD detection results in terms of AUC (↑). The best and second-best results are highlighted with bold and underline, respectively. (*Label-perturbation)*

dataset into a pairwise graph format by flattening hyperedges into clique connections, i.e., connecting all node pairs that co-occur in the same hyperedge.

5.2 Results and Discussion

Table 1 reports the OOD detection performance under the feature perturbation setting. HyperGOOD consistently achieves the best performance across eight datasets, demonstrating strong generalization to distribution shifts in both sparse and dense hypergraph structures. *Unlike these competitors, which largely rely on shallow relational modeling, HyperGOOD uniquely combines high-order hypergraph context with framelet-based spectral decomposition (FSD).* This enables more precise uncertainty estimation across structural for HOOD detection. On the Pubmed dataset, HyperGOOD performs slightly below graph-based methods like GNNSafe and NODESAFE. We attribute this to the relatively simple hypergraph structure of Pubmed. The dataset contains 7,963 hyperedges over 19,717 nodes, resulting in sparse and relatively uniform connectivity. Consequently, the benefit of our multi-scale hypergraph modeling becomes less pronounced, and methods that emphasize node-level feature regularization are more competitive.

Table 2 reports OOD detection performance under the label perturbation setting. Compared to feature corruption, this setting poses a more subtle challenge as the input features remain intact while the label signals are rendered noisy or mismatched. This weakens the effectiveness of supervision and is particularly challenging for methods like HyperGOOD that rely heavily on learning discriminative feature representations. Nevertheless, HyperGOOD maintains strong performance across most datasets, suggesting that its structure-aware and spectral modeling capabilities confer robustness even under degraded label quality. We also observe that traditional uncertainty scoring methods such as MSP and ODIN fail to effectively capture such semantic shifts, often yielding near-random performance around 50%. Although Mahalanobis distance shows slightly better performance on Cora-family datasets, it performs poorly on Senate, House, and ModelNet40. Further experimental

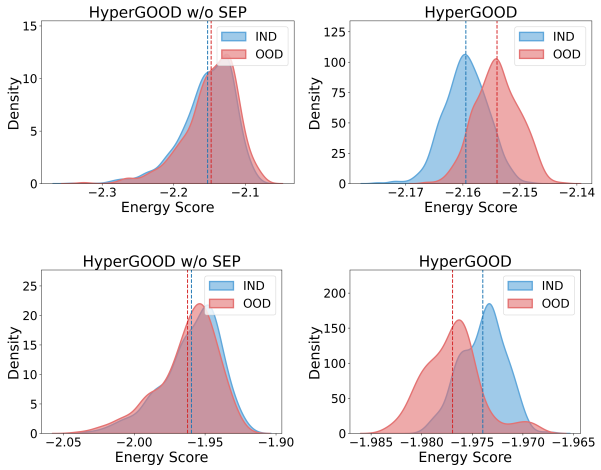

Figure 3: Visualization of energy score distributions used for HOOD detection: top shows results on House with feature perturbation, and bottom shows results on Senate with label perturbation.

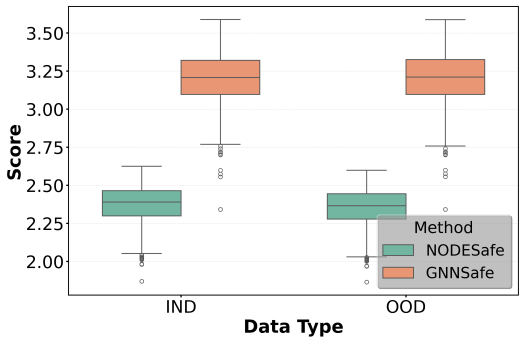

Figure 4: Score distribution of NODESAFE and GNNSafe on IND and OOD data.

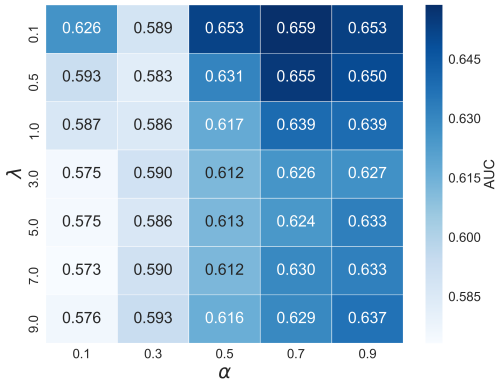

Figure 5: Sensitivity analysis of HyperGOOD with respect to λ and α on Cora.

results and analysis are provided in Appendix A.

Remark on hypergraph-level HOOD detection. In preliminary experiments, we also considered perturbing the hypergraph structure (e.g., modifying hyperedges or connectivity patterns) as a potential form of OOD signal. However, we found that such structural changes often led to unpredictable and non-reproducible behaviors, and lacked consistent correlation with OOD likelihood under our framework. We therefore focus our study on feature-level and semantic OOD shifts under a fixed hypergraph structure, which are more well-defined and aligned with existing literature. We believe the formalization and modeling of structural OOD in hypergraphs remains an open and important research direction.

5.3 Ablation Study

To evaluate the contribution of structure-aware energy propagation (SEP), we visualize the energy score distributions of IND and OOD nodes before and after propagation. These energy scores serve as OOD detection signals in our framework. As shown in Figure 3, the left plot displays energy scores computed directly from our spectral uncertainty estimation, without structure-aware energy propagation. In contrast, the right plot shows the scores after applying our structure-aware propagation over the hypergraph. The propagated energy scores exhibit significantly improved separability between IND and OOD nodes, resulting in a more bimodal and distinguishable distribution.

This result confirms that incorporating relational context allows our model to amplify weak or ambiguous OOD signals that might otherwise be masked in the original feature space. While our multi-scale representation module has already demonstrated effectiveness in the main experiment (Sec. 5.2), this ablation isolates and highlights the added benefit of structure-aware uncertainty refinement. The multiscale characterization provides a complete topological representation of hypergraphs, while structure-guided uncertainty propagation ensures consistent OOD signal amplification across different relational contexts, yielding detection robustness that surpasses feature-space-only approaches.

Observation on graph-based OOD detectors. Figure 4 shows the confidence scores produced by GNNSafe (Wu et al. 2023b) and NODESAFE (Yang et al. 2024b) on the House with feature perturbation. Our visualization reveals a critical limitation of existing graph-based OOD detectors: When applied to hypergraph-converted data through standard graph convolution operators, both GNNSafe and NODESAFE fail to maintain clear separation between IND and OOD distributions. This phenomenon is indicative of the special (different from traditional graph) challenges of performing OOD detection tasks on hypergraphs.

5.4 Parameter Sensitivity Analysis

To examine the robustness of our method with respect to key hyperparameters, we conduct a sensitivity analysis on the Cora dataset under the feature perturbation setting. Specifically, we vary the propagation regularization weight $\lambda \in \{0.1, 0.5, 1, 3, \cdots, 9\}$ and the residual ratio $\alpha \in \{0.1, 0.3, \cdots, 0.9\}$, and report the AUC scores in Figure 5. Overall, HyperGOOD exhibits stable performance across a wide range of hyperparameter choices. The AUC values fluctuate within a relatively narrow band, indicating that the model is not overly sensitive to specific λ or α values. This stability suggests that the structure-aware propagation and multi-scale uncertainty modeling components work synergistically and robustly, without requiring fine-tuned balancing.

6 Conclusion

This paper introduces the novel task of HOOD detection, extending out-of-distribution detection to hypergraph-structured data where high-order relational dependencies present unique challenges. We propose HyperGOOD, a multi-scale, structure-aware framework that combines spectral decomposition and uncertainty propagation to effectively identify OOD samples in hypergraphs. Since, to the best of our knowledge, this is the first work that formally addresses HOOD detection, it is expected to motivate more follow-up research in this direction. Future work may explore HOOD detection in dynamic or temporal hypergraphs, its integration into downstream tasks such as medical analysis and recommendation, and the design of general OOD benchmarks tailored for hypergraph learning.

Acknowledgments

This work is supported by National Key Research and Development Program of China (2024YFE0214000), the National Natural Science Foundation of China (Grant Nos. U22A20102, 62337001, 62536006, 62172370).

References

Antelmi, A.; Cordasco, G.; Polato, M.; Scarano, V.; Spagnuolo, C.; and Yang, D. 2023. A survey on hypergraph representation learning. *ACM Computing Surveys*, 56(1): 1–38.

Arya, D.; Gupta, D. K.; Rudinac, S.; and Worring, M. 2024. Adaptive neural message passing for inductive learning on hypergraphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1): 19–31.

Bai, S.; Zhang, F.; and Torr, P. H. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110: 107637.

Cai, T.; Jiang, Y.; Li, M.; Huang, C.; Wang, Y.; and Huang, Q. 2025a. ML-GOOD: Towards multi-label graph out-of-distribution detection. In *AAAI*, 15, 15650–15658.

Cai, T.; Jiang, Y.; Liu, Y.; Li, M.; Huang, C.; and Pan, S. 2025b. Out-of-distribution detection on graphs: A survey. *arXiv preprint arXiv:2502.08105*.

Chen, D.-Y.; Tian, X.-P.; Shen, Y.-T.; and Ouhyoung, M. 2003. On visual similarity based 3D model retrieval. In *Computer Graphics Forum*, volume 22, 223–232.

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

Chien, E.; Pan, C.; Peng, J.; and Milenkovic, O. 2022. You Are AllSet: A multiset function framework for hypergraph neural networks. In *ICLR*.

Chodrow, P. S.; Veldt, N.; and Benson, A. R. 2021. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, 7(28): eabh1303.

Dua, D.; Graff, C.; et al. 2017. UCI machine learning repository. URL: http://archive. ics. uci. edu/ml.

Feng, Y.; Han, J.; Ying, S.; and Gao, Y. 2024. Hypergraph isomorphism computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5): 3880–3896.

Feng, Y.; You, H.; Zhang, Z.; Ji, R.; and Gao, Y. 2019. Hypergraph neural networks. In *AAAI*, 3558–3565.

Fowler, J. H. 2006. Legislative cosponsorship networks in the US House and Senate. *Social Networks*, 28(4): 454–465.

Gao, Y.; Ji, S.; Han, X.; and Dai, Q. 2024. Hypergraph computation. *Engineering*.

Hendrycks, D.; and Gimpel, K. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*.

Huang, J.; and Yang, J. 2021. UniGNN: A unified framework for graph and hypergraph neural networks. In *IJCAI*, 2563–2569.

Ju, W.; Yi, S.; Wang, Y.; Xiao, Z.; Mao, Z.; Li, H.; Gu, Y.; Qin, Y.; Yin, N.; Wang, S.; et al. 2024. A Survey of graph neural networks in real world: Imbalance, noise, privacy and OOD challenges. *arXiv preprint arXiv:2403.04468*.

Khan, B.; Wu, J.; Yang, J.; and Ma, X. 2025. Heterogeneous hypergraph neural network for social recommendation using attention network. *ACM Transactions on Recommender Systems*, 3(3): 1–22.

Kim, S.; Lee, S. Y.; Gao, Y.; Antelmi, A.; Polato, M.; and Shin, K. 2024. A survey on hypergraph neural networks: An in-depth and step-by-step guide. In *KDD*, 6534–6544.

Lee, K.; Lee, K.; Lee, H.; and Shin, J. 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *NeurIPS*, 7167–7177.

Li, M.; Fang, Y.; Wang, Y.; Feng, H.; Gu, Y.; Bai, L.; and Lio, P. 2025a. Deep hypergraph neural networks with tight framelets. In *AAAI*, 18385–18392.

Li, M.; Gu, Y.; Wang, Y.; Fang, Y.; Bai, L.; Zhuang, X.; and Lio, P. 2025b. When hypergraph meets heterophily: New benchmark datasets and baseline. In *AAAI*, 17, 18377–18384.

Liang, S.; Li, Y.; and Srikant, R. 2018. Enhancing the reliability of out-of-distribution image detection in neural networks. In *ICLR*.

Liu, W.; Wang, X.; Owens, J.; and Li, Y. 2020. Energy-based out-of-distribution detection. In *NeurIPS*, 21464–21475.

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

Ma, J.; Li, F.; Zhang, R.; Xu, Z.; Cheng, D.; Ouyang, Y.; Zhao, R.; Zheng, J.; Zheng, Y.; and Jiang, C. 2023. Fighting against organized fraudsters using risk diffusion-based parallel graph neural network. In *IJCAI*, 6138–6146.

Millán, A. P.; Sun, H.; Giambagli, L.; Muolo, R.; Carletti, T.; Torres, J. J.; Radicchi, F.; Kurths, J.; and Bianconi, G. 2025. Topology shapes dynamics of higher-order networks. *Nature Physics*, 1–9.

Murgas, K. A.; Saucan, E.; and Sandhu, R. 2022. Hypergraph geometry reflects higher-order dynamics in protein interaction networks. *Scientific Reports*, 12(1): 20879.

Rusch, T. K.; Bronstein, M. M.; and Mishra, S. 2023. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993*.

Shen, X.; Wang, Y.; Zhou, K.; Pan, S.; and Wang, X. 2024. Optimizing OOD detection in molecular graphs: A novel approach with diffusion models. In *KDD*, 2640–2650.

Solomon, J. 2006. Case studies: Why are they important? *Nature Clinical Practice Cardiovascular Medicine*, 3(11): 579–579.

Sun, X.; Cheng, H.; Liu, B.; Li, J.; Chen, H.; Xu, G.; and Yin, H. 2023. Self-supervised hypergraph representation learning for sociological analysis. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11860–11871.

Wang, F.; Liu, Y.; Liu, K.; Wang, Y.; Medya, S.; and Yu, P. S. 2024a. Uncertainty in graph neural networks: A survey. *Transactions on Machine Learning Research*. URL: https://openreview.net/forum?id=0e1Kn76HM1.

Wang, L.; He, D.; Zhang, H.; Liu, Y.; Wang, W.; Pan, S.; Jin, D.; and Chua, T.-S. 2024b. GOODAT: Towards test-time graph out-of-distribution detection. In *AAAI*, 15537–15545.

Wang, M.; Yang, H.; Huang, J.; and Cheng, Q. 2024c. Moderate message passing improves calibration: A universal way to mitigate confidence bias in graph neural networks. In *AAAI*, 21681–21689.

Wang, P.; Yang, S.; Liu, Y.; Wang, Z.; and Li, P. 2023. Equivariant hypergraph diffusion neural operators. In *ICLR*.

Wang, Y.; and Kleinberg, J. 2024. From graphs to hypergraphs: Hypergraph projection and its reconstruction. In *ICLR*.

Wu, J.; He, K.; Mao, R.; Li, C.; and Cambria, E. 2023a. MEGACare: Knowledge-guided multi-view hypergraph predictive framework for healthcare. *Information Fusion*, 100: 101939.

Wu, Q.; Chen, Y.; Yang, C.; and Yan, J. 2023b. Energy-based out-of-distribution detection for graph neural networks. In *ICLR*.

Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; and Xiao, J. 2015. 3D ShapeNets: A deep representation for volumetric shapes. In *CVPR*, 1912–1920.

Xie, L.; Gao, S.; Liu, J.; Yin, M.; and Jin, T. 2025. K-hop hypergraph neural network: A comprehensive aggregation approach. In *AAAI*, 20, 21679–21687.

Xue, F.; He, Z.; Zhang, Y.; Xie, C.; Li, Z.; and Tan, F. 2024. Enhancing the power of OOD detection via sample-aware model selection. In *CVPR*, 17148–17157.

Yadati, N.; Nimishakavi, M.; Yadav, P.; Nitin, V.; Louis, A.; and Talukdar, P. 2019. HyperGCN: A new method for training graph convolutional networks on hypergraphs. In *NeurIPS*, 1511–1522.

Yan, Y.; Yang, C.; Chen, Y.; Cai, R.; and Ng, M. 2025. Hypergraph learning for unsupervised graph alignment via optimal transport. In *AAAI*, 20, 21913–21921.

Yang, J.; Zhou, K.; Li, Y.; and Liu, Z. 2024a. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, 132(12): 5635–5662.

Yang, S.; Liang, B.; Liu, A.; Gui, L.; Yao, X.; and Zhang, X. 2024b. Bounded and uniform energy-based out-of-distribution detection for graphs. In *ICML*, 56216–56234.

Yann, L.; Sumit, C.; Raia, H.; Marc'Aurelio, R.; and Fu-Jie, H. 2006. *A tutorial on energy-based learning*. MIT Press.

Zhang, B.; Zhu, J.; Wang, Z.; Liu, T.; Du, B.; and Han, B. 2024. What if the input is expanded in OOD detection? 21289–21329.