Unsupervised Graph-level Anomaly Detection via Multi-granular Graph Structure Learning

Ge Zhang[1], **Huimei Li**[1], **Guohao Sun**[1*], **Xiu Fang**[1], **Xixun Lin**[2], **Xiaobao Wang**[3], **Pengfei Jiao**[4] and **Liang Yang**[5]

[1]School of Information and Intelligent Science, Donghua University, Shanghai, China
[2]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[3]School of Artificial Intelligence, Tianjin University, Tianjin, China
[4]School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China
[5]School of Artificial Intelligence, Hebei University of Technology, Tianjin, China
gezhang@dhu.edu.cn, 2242853@mail.dhu.edu.cn, ghsun@dhu.edu.cn, xiu.fang@dhu.edu.cn, linxixun@iie.ac.cn, wangxiaobao@tju.edu.cn, pjiao@hdu.edu.cn, yangliang@vip.qq.com

Abstract

Graph-level anomaly detection (GLAD) aims to identify graphs that deviate from the majority in a dataset of graphs. Existing methods typically adopt either a global aggregation perspective that summarizes nodes within a graph into a representation vector, or a subgraph-oriented perspective which regards certain local subgraphs as indicators of anomalous properties. However, both paradigms operate from a fixed perspective and may fail to identify anomalous graphs whose discriminative characteristics manifest at multiple levels of structure granularity, where each level features the coarsened graphs at a specific granularity. In this paper, we propose M-GLAD, an unsupervised GLAD method via multi-granular graph structure learning. M-GLAD is grounded in the Prototype-guided Multi-granular Information Bottleneck (PMIB) principle. PMIB aims to measure the mutual information between coarsened graphs at the specific level of structure granularity and the learnable granularity-specific prototypes that summarize normal patterns of graphs at each granularity. The multiple levels of structure granularity of graphs and prototypes are abstracted by a graph structure abstraction module. This formulation enables granularity-aware anomaly scoring that considers anomalous characteristics across different levels of structure granularity. Extensive experiments on eight real-world graph datasets demonstrate that M-GLAD achieves superior performance over competitive baselines.

1 Introduction

Graph-structured data are ubiquitous in a wide range of domains, including biochemistry [Guo *et al.*, 2022], finance [Rajput and Singh, 2022], and social networks [Xia *et al.*, 2021; Ju *et al.*, 2024]. In many of these domains, it is crucial to identify graphs that exhibit a significant deviation from the majority in the graph collection [Qiao *et al.*, 2025; Yang *et al.*, 2026], a task formalized as Graph-level Anomaly Detection (GLAD). Detecting such anomalous graphs is essential for uncovering rare events, mitigating risks, and driving scientific discovery. Compared to identifying anomalous nodes [Lin *et al.*, 2024; Niu *et al.*, 2025] or edges [Lee *et al.*, 2024] in a single graph, GLAD highlights its significance in understanding the structure features of different graphs.

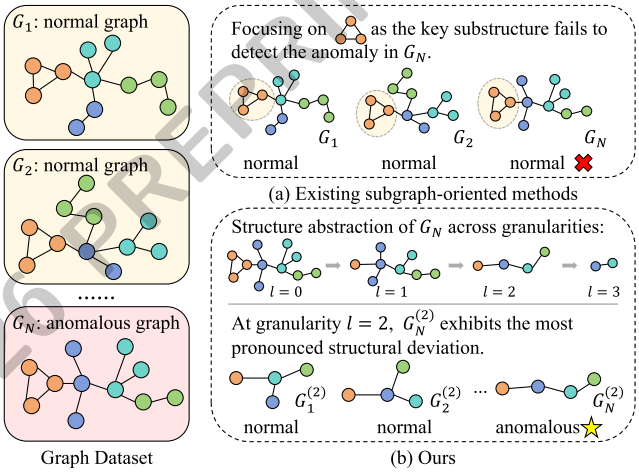


Figure 1: An illustrative comparison between (a) existing subgraph-oriented GLAD approaches and (b) our proposed method M-GLAD. Existing methods primarily focus on local subgraphs, which may fail to account for graph-level anomalies arising from long-range dependencies in graphs, e.g., G_N. In contrast, M-GLAD analyzes graph structures across multiple levels of granularity and highlights structural deviations at specific granularities.

A rich line of GLAD algorithms has recently been developed, among which two paradigms have shown promise. The first paradigm adopts a global perspective, aggregating node representations learned via Graph Neural Networks (GNNs) [Kipf and Welling, 2017; Veličković *et al.*, 2018;

*Corresponding author.

Xu *et al.*, 2019; Lin *et al.*, 2025] into a graph embedding, which is subsequently scored for anomaly detection [Zhao and Akoglu, 2020; Luo *et al.*, 2022; Zhang *et al.*, 2022a; Ma *et al.*, 2023; Zhang *et al.*, 2024; Kim *et al.*, 2024; Chen *et al.*, 2025; Zhang *et al.*, 2026]. The second takes a subgraph-oriented perspective [Liu *et al.*, 2023; Yang *et al.*, 2025; Zhang *et al.*, 2025], aiming to identify subgraphs that capture the essential differences between normal and anomalous graphs. Despite their effectiveness, both paradigms are limited by the fixed perspective on graph structures.

In GLAD, the key differences among anomalous and normal graphs often involve multiple levels of structure granularity. As illustrated in Figure 1 (b), a graph can be progressively coarsened into multiple levels of structure granularity; at particular granularities, certain graphs exhibit anomalous characteristics due to significant structural deviations from other graphs. In contrast, the global aggregation in the first paradigm may obscure discriminative signals that differentiate normal and anomalous graphs. Subgraph-oriented approaches of the second paradigm implicitly assume that the distinction between normal and anomalous graphs can be attributed solely to the presence or absence of certain subgraphs, may fail to account for anomalies arising from long-range dependencies across the graph. As shown in Figure 1 (a), the real-world graphs often violate this assumption, especially for graphs with complex relational structures (e.g., social networks), where subgraphs exhibit high variability and do not uniquely characterize abnormality. Hence, because GLAD can benefit from a multi-granular perspective, it is essential to develop a method that considers such structures.

To effectively distinguish anomalous graphs from normal ones based on the multi-granular structure learning, two core challenges arise. **Challenge #1**: How to extract multiple levels of structure granularity under an unsupervised manner? The scarcity of labeled anomalies in real-world scenarios necessitates unsupervised extraction of multi-granular graph representations. Hierarchical graph pooling methods [Ying *et al.*, 2018; Lee *et al.*, 2019; Gao and Ji, 2019; Duval and Malliaros, 2022] can derive multiple granularities through graph coarsening, but they are primarily designed for supervised classification objectives rather than anomaly-sensitive graph structure learning. **Challenge #2**: How to achieve granularity-aware anomaly scoring? In GLAD, abnormalities may manifest at specific granularities, and anomalous patterns may vary across granularities. Therefore, the model must adaptively learn across multiple levels of structure granularity to quantify the anomaly degree of graphs.

To address the above challenges, we propose M-GLAD, an unsupervised GLAD model via multi-granular graph structure learning. M-GLAD captures multiple levels of structure granularity through a structure abstraction module. For each input graph, it produces a sequence of coarsened graphs at different levels of granularity. To enable granularity-aware anomaly scoring, we introduce the Prototype-guided Multi-granular Information Bottleneck (PMIB), an unsupervised graph representation framework tailored for the GLAD task. The PMIB maximizes the mutual information between the learnable granularity-specific prototype and input graphs at the corresponding granularity. The granularity-specific prototypes are parameterized graphs that summarize the typical normal characteristics of graphs at a specific granularity. During inference, the anomaly score of a test graph is obtained by aggregating deviations at different granularities, where each deviation measures the discrepancy between the coarsened graph and the corresponding learnable granularity-specific prototype. To this end, our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to advance unsupervised GLAD from the exploration of multiple levels of structure granularity.
- We propose M-GLAD, an unsupervised multi-granular GLAD framework equipped with the PMIB principle. The framework learns granularity-specific prototypes to capture the typical characteristics shared by graphs and enables granularity-aware anomaly scoring.
- Extensive experiments on real-world graph datasets show that M-GLAD achieves superior performance and validate the effectiveness of distinguishing between normal and anomalous graphs through multi-granular graph structure learning.

2 Related Work

Global-aggregation based GLAD methods. With the development of GNNs, many recent GLAD methods have moved toward representation learning paradigms that characterize each graph using the graph embedding. For instance, deep one-class classification methods [Zhao and Akoglu, 2020; Qiu *et al.*, 2022; Zhang *et al.*, 2024] optimize hypersphere-based objectives to learn compact normal graph representations, using distances to the hypersphere center as anomaly scores. Reconstruction-based approaches [Luo *et al.*, 2022; Kim *et al.*, 2024; Chen *et al.*, 2025] rely on graph autoencoders or their variational variants, where reconstruction discrepancies indicate anomalies. Knowledge-distillation based methods [Ma *et al.*, 2022] model normal graphs via random distillation, using prediction errors as anomaly scores. There are also other approaches separating the normal and anomalous graphs in the representation space through geometric or spectral objectives [Dong *et al.*, 2024].

Subgraph-oriented GLAD methods. Another line of GLAD methods emphasizes local structural patterns, extracting discriminative subgraphs [Zhang *et al.*, 2021; Zhang and Li, 2021; Zhao *et al.*, 2022] from graphs as evidence for anomaly detection. For example, SIGNET [Liu *et al.*, 2023] learns a subgraph extractor and detects graph-level anomalies based on the abnormality of extracted subgraphs. Building on this paradigm, GLADPro [Yang *et al.*, 2025] further improves the subgraph-level anomaly modeling to better characterize the difference between normal and anomalous graphs.

Despite empirical success, they are generally incapable of capturing graph-level anomalies whose discriminative characteristics emerge at multiple levels of structural granularity. Specifically, global aggregation–based approaches can easily smooth the subtle anomaly signals embedded in fine-grained structures. In contrast, subgraph-based methods typically rely on strong assumptions about structural regularity; however,

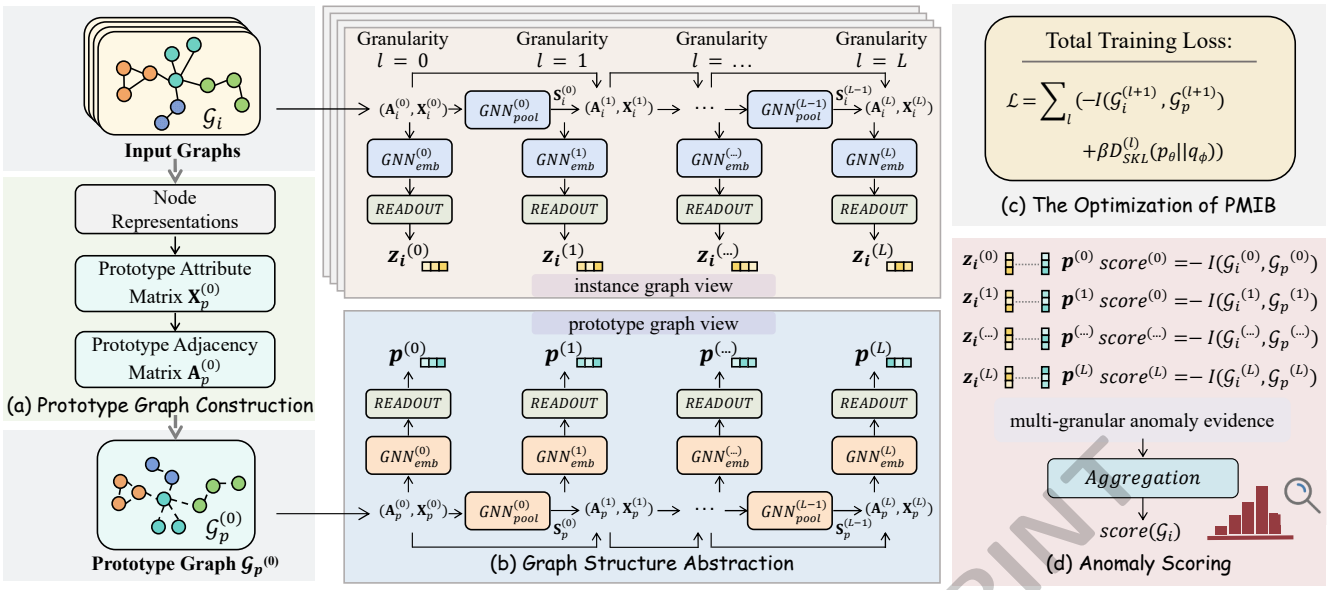


Figure 2: The overview of the proposed method M-GLAD. As illustrated, M-GLAD consists of four main components: (a) Prototype Graph Learning, which constructs a learnable prototype graph from normal graphs; (b) Graph Structure Abstraction, which coarsens the graph and prototype into multiple granularity levels; (c) Optimization of PMIB, which is formulated with a mutual information term and a symmetrized Kullback–Leibler divergence term; and (d) Anomaly Scoring, which aggregates multi-granular anomaly evidence for final anomaly scoring.

the presence of diverse or irregular subgraphs does not necessarily indicate anomalies.

3 Preliminaries

Graph. A graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{A})$, where $\mathcal{V}$ and $\mathcal{E}$ are the node and edge sets, respectively. $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d}$ is the node attribute matrix, and $\mathbf{A} \in \{0,1\}^{|\mathcal{V}| \times |\mathcal{V}|}$ is the adjacency matrix. For a graph $\mathcal{G}$, its multiple levels of structure granularity can be represented by a graph sequence $\{\mathcal{G}^{(l)}\}_{l=0}^{L}$ with $\mathcal{G}^{(0)} = \mathcal{G}$, where $\mathcal{G}^{(l)} = (\mathcal{V}^{(l)}, \mathcal{E}^{(l)}, \mathbf{X}^{(l)}, \mathbf{A}^{(l)})$.

Granularity-specific Prototype. We define the granularity-specific prototype as $\mathcal{G}_p^{(l)} = (\mathcal{V}_p^{(l)}, \mathcal{E}_p^{(l)}, \mathbf{X}_p^{(l)}, \mathbf{A}_p^{(l)})$, for $l = \{0, 1, ..., L\}$. Each granularity-specific prototype is designed to capture the typical patterns shared by normal graphs at the corresponding level of structure granularity. The attribute matrix $\mathbf{X}_p^{(0)}$ and adjacency matrix $\mathbf{A}_p^{(0)}$ are initialized by the method described in Section 4.2. Prototypes at the deeper level of granularities ($l > 0$) are coarsened from the original prototype $\mathcal{G}_p^{(0)}$. $\mathbf{X}_p^{(l)}$ and $\mathbf{A}_p^{(l)}$ for $l > 0$ are learnable, with each element in matrices being an updatable parameter.

Graph Convolution. At the t-th layer of a GNN, the node representation is updated by

$$\mathbf{h}_v^{(t)} = f_{\text{Update}}^{(t-1)}\left(\mathbf{h}_v^{(t-1)}, f_{\text{Agg}}^{(t-1)}\left(\{\mathbf{h}_u^{(t-1)} : u \in \mathcal{N}(v; \mathbf{A})\}\right)\right), \tag{1}$$

where $\mathcal{N}(v; \mathbf{A})$ denotes the neighborhood of node v induced by $\mathbf{A}$, $f_{\text{Agg}}(\cdot)$ aggregates information from the neighbors, and $f_{\text{Update}}(\cdot)$ updates the node representation.

Multi-view Information Bottleneck (MIB). MIB is an unsupervised formulation of the information bottleneck (IB) principle [Tishby *et al.*, 2000; Tishby and Zaslavsky, 2015; Wu *et al.*, 2020]. Specifically, MIB exploits complementary views of the same data instance to learn its representations in an unsupervised fashion [Bachman *et al.*, 2019; Hassani and Khasahmadi, 2020; Xu *et al.*, 2024]. Given two complementary views V_1 and V_2 of the same instance, MIB learns a representation Z_1 from view V_1 that is maximally predictive of view V_2, while suppressing information that is specific to V_1. Formally, MIB can be formulated as the following constrained optimization:

$$\max_{Z_1} \; I(V_2; Z_1) - \beta_1 \, I(V_1; Z_1 \mid V_2), \tag{2}$$

where β_1 is a trade-off coefficient. A symmetric objective for V_2 is defined analogously.

Problem Formulation. Given a training set $\mathcal{D}_{\text{tr}}$ consisting exclusively of normal graphs, GLAD aims to learn a scoring function $f : \mathcal{G} \to \text{score}(\mathcal{G})$ that assigns each input graph an anomaly score. In M-GLAD, the final anomaly score is obtained by aggregating granularity-specific scores across multiple levels: $\text{score}(\mathcal{G}) = \text{Agg}\big(\{\text{score}^{(l)}(\mathcal{G})\}_{l=0}^{L}\big)$, where $\text{score}^{(l)}(\mathcal{G})$ denotes the score at granularity l and $\text{Agg}(\cdot)$ specifies the scoring strategy (formalized in Section 4.5).

4 Methodology

This section describes the methodology of M-GLAD, and Figure 2 provides an illustration of the overall framework.

4.1 PMIB Framework Overview

Inspired by the MIB principle and prototype learning [Zhang *et al.*, 2022b; Dai and Wang, 2025], we propose PMIB, a

Prototype-guided Multi-granular Information Bottleneck objective that guides the optimization of M-GLAD.

To capture shared patterns of normal graphs at a specific level of structure granularity, we introduce learnable granularity-specific prototypes $\mathcal{G}_p^{(l)}$, where $l = \{0, 1, ..., L\}$. The discrepancy between a graph and the prototype quantifies abnormality: normal graphs align with prototypes across granularities, whereas anomalies deviate due to atypical patterns. To operationalize this principle, PMIB defines two complementary views at each granularity l: (i) instance graph view, denoting the coarsened graph $\mathcal{G}^{(l)}$ at granularity l, and (ii) prototype graph view, defined as the corresponding prototype $\mathcal{G}_p^{(l)}$ at the same granularity. For the instance graph view, PMIB enforces an information bottleneck objective to ensure that the coarsened graph $\mathcal{G}^{(l+1)}$ at the next granularity level preserves information consistent with the prototype $\mathcal{G}_p^{(l)}$, while suppressing instance-specific redundancies irrelevant to normality. Formally,

$$\max_{\mathcal{G}^{(l+1)}} \; I\Big(\mathcal{G}_p^{(l)}; \mathcal{G}^{(l+1)}\Big) - \beta_1 \, I\Big(\mathcal{G}^{(l)}; \mathcal{G}^{(l+1)} \mid \mathcal{G}_p^{(l)}\Big), \quad (3)$$

where β_1 is a trade-off parameter. Symmetrically, the prototype view is governed by the dual bottleneck objective:

$$\max_{\mathcal{G}_p^{(l+1)}} \; I\Big(\mathcal{G}^{(l)}; \mathcal{G}_p^{(l+1)}\Big) - \beta_2 \, I\Big(\mathcal{G}_p^{(l)}; \mathcal{G}_p^{(l+1)} \mid \mathcal{G}^{(l)}\Big), \quad (4)$$

where β_2 is defined analogously.

The final PMIB objective is the negative of the sum of the two symmetric bottleneck objectives in Eqs. (3) and (4). Since directly minimizing this objective is intractable, we optimize a tractable variational upper bound instead:

$$\begin{aligned}\mathcal{L}_{\mathrm{PMIB}}^{(l)} = &-I(\mathcal{G}^{(l+1)}; \mathcal{G}_p^{(l+1)}) \\ &+ \beta D_{\mathrm{SKL}}\Big(p_\theta(\mathcal{G}^{(l+1)}|\mathcal{G}^{(l)}) \,\|\, q_\phi(\mathcal{G}_p^{(l+1)}|\mathcal{G}_p^{(l)})\Big),\end{aligned} \quad (5)$$

where $p_\theta(\mathcal{G}^{(l+1)} \mid \mathcal{G}^{(l)})$ and $q_\phi(\mathcal{G}_p^{(l+1)} \mid \mathcal{G}_p^{(l)})$ denote the instance and prototype transition distributions induced by the graph and prototype encoders, respectively, $D_{\mathrm{SKL}}(\cdot)$ denotes the symmetrized Kullback–Leibler (SKL) divergence and β is a trade-off hyper-parameter. Detailed derivations of the upper bound are provided in the supplementary[1]. This formulation encourages the instance and prototype views to align on normal patterns, while allowing deviations to signal anomalies. We denote the mutual information (MI) loss (the first term) as $\mathcal{L}_{\mathrm{MI}}^{(l)}$, and the SKL divergence loss (the second term) as $\mathcal{L}_{\mathrm{SKL}}^{(l)}$. Their concrete computation is detailed in Section 4.4.

In the following, we first describe how to obtain the initial prototype $\mathcal{G}_p^{(0)}$ (Section 4.2). Then, we introduce the graph structure abstraction to coarsen input graphs and prototypes (Section 4.3), followed by PMIB optimization (Section 4.4). Finally, we present granularity-specific anomaly scoring to quantify the degree of graph abnormality (Section 4.5).

[1]https://github.com/Lab821-gz/M-GLAD/blob/main/M-GLAD_supplementary.pdf

4.2 Prototype Graph Construction

At each level of structure granularity, normal graphs are associated with a shared granularity-specific prototype. The initial prototype is represented as $\mathcal{G}_p^{(0)} = (\mathcal{V}_p^{(l)}, \mathcal{E}_p^{(l)}, \mathbf{X}_p^{(0)}, \mathbf{A}_p^{(0)})$. Next, we introduce how to initialize the attribute matrix $\mathbf{X}_p^{(0)}$ and the adjacency matrix $\mathbf{A}_p^{(0)}$.

Attribute matrix initializing. We first obtain node embeddings for normal graphs in the training set $\mathcal{D}_{\mathrm{tr}}$:

$$\mathbf{H} = \mathrm{GNN}_{\mathrm{emb}}(\mathbf{A}, \mathbf{X}), \quad (6)$$

where $\mathrm{GNN}_{\mathrm{emb}}$ is built upon the graph convolution, and each row of $\mathbf{H}$ represents a node embedding in a graph $\mathcal{G}$. Let $\mathbf{U}_{\mathrm{tr}}$ denote the representation matrix of all nodes from graphs in $\mathcal{D}_{\mathrm{tr}}$, we initialize the prototype node attributes by:

$$\mathbf{X}_p^{(0)} = \mathrm{KMeans}(\mathbf{U}_{\mathrm{tr}}, k = |\mathcal{V}_p^{(0)}|), \quad (7)$$

where $\mathrm{KMeans}(\cdot)$ denotes the k-means clustering algorithm, and $k = |\mathcal{V}_p^{(0)}|$ specifies the number of clusters, corresponding to the number of nodes in the prototype $\mathcal{G}_p^{(0)}$. In practice, $|\mathcal{V}p^{(0)}|$ is set to the average number of nodes of the graphs in the training set $\mathcal{D}$tr.

Adjacency matrix initializing. To equip the prototype with structural information [Cai *et al.*, 2025], we initial the adjacency matrix as:

$$\mathbf{A}_p^{(0)} = \mathcal{T}(\mathbf{X}_p^{(0)} \mathbf{X}_p^{(0)\top}), \quad (8)$$

where $\mathcal{T}(\cdot)$ is a symmetric mapping that transforms pairwise latent similarities into non-negative edge weights. This yields an adaptive prototype structure that co-evolves with $\mathbf{X}_p^{(0)}$ and captures shared structural regularities. Prototypes at next levels ($l > 0$) of structure granularity can be obtained through the module of graph structure abstraction.

4.3 Graph Structure Abstraction

To obtain multiple levels of granularities in graph structures, we introduce a graph structure abstraction module that progressively coarsens graphs and prototypes via hierarchical graph pooling [Islam *et al.*, 2023; Wang *et al.*, 2024]. For clarity, we describe the procedure using the instance graph as an example; the same operations are applied to the prototype in parallel. At the level l of the structure granularity, node representations are first updated using the GNN:

$$\mathbf{H}^{(l)} = \mathrm{GNN}_{\mathrm{emb}}^{(l)}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)}), \quad (9)$$

where $\mathbf{H}^{(l)}$ denote the node embeddings of the graph $\mathcal{G}^{(l)}$, with $\mathbf{H}_p^{(l)}$ defined analogously for the prototype. Following [Ying *et al.*, 2018], we employ a separate pooling GNN with the same graph convolution operator as $\mathrm{GNN}_{\mathrm{emb}}$ to generate soft cluster assignments:

$$\mathbf{S}^{(l)} = \mathrm{softmax}\Big(\mathrm{GNN}_{\mathrm{pool}}^{(l)}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)})\Big), \quad (10)$$

where each row of $\mathbf{S}^{(l)}$ defines a categorical distribution over K_l clusters at granularity $l+1$, and $K_l = \lceil r \cdot |\mathcal{V}^{(l)}| \rceil$ is determined by a coarsening ratio r.

In practice, we aim to obtain near one-hot cluster assignments for each node. Since sampling from the categorical distributions in $\mathbf{S}^{(l)}$ to achieve true one-hot assignments is non-differentiable, we instead employ a differentiable relaxation, i.e., the Gumbel-Softmax trick [Jang *et al.*, 2017], which provides a differentiable reparameterization of the categorical sampling process. The resulting matrix, denoted as $\hat{\mathbf{S}}^{(l)}$, serves as a near one-hot approximation of $\mathbf{S}^{(l)}$, enabling gradient-based optimization. The attribute and the adjacency matrix of $\mathcal{G}^{(l+1)}$ are obtained as below:

$$\mathbf{X}^{(l+1)} = \hat{\mathbf{S}}^{(l)\top}\mathbf{H}^{(l)}, \quad \mathbf{A}^{(l+1)} = \hat{\mathbf{S}}^{(l)\top}\mathbf{A}^{(l)}\hat{\mathbf{S}}^{(l)}. \tag{11}$$

By recursively applying the structure abstraction process described above, we obtain multiple levels of granularity for the graph $\mathcal{G}$, denoted as $\{\mathcal{G}^{(l)}\}_{l=0}^{L}$.

The same iterative graph structure abstraction process is applied to the initial prototype graph $\mathcal{G}_p^{(0)}$, thereby constructing coarser prototype series $\{\mathcal{G}_p^{(l)}\}_{l=1}^{L}$ at different levels of structure granularity.

4.4 The Optimization of PMIB

After completing obtaining different levels of structure granularity of the input graphs and learnable prototypes, we optimize the PMIB objective in Eq. (5) by explicitly estimating the two loss terms, $\mathcal{L}_{\mathrm{MI}}^{(l)}$ and $\mathcal{L}_{\mathrm{SKL}}^{(l)}$, at each granularity level. We next instantiate these two terms in a tractable and differentiable way.

The MI Term. Directly computing the MI between graphs is intractable. Therefore, to evaluate $\mathcal{L}_{\mathrm{MI}}^{(l)}$, we first encode the instance graph $\mathcal{G}^{(l)}$ and the prototype $\mathcal{G}_p^{(l)}$ into their graph-level embeddings, $\mathbf{z}^{(l)}$ and $\mathbf{p}^{(l)}$:

$$\mathbf{z}^{(l)} = \mathrm{READOUT}(\mathbf{H}^{(l)}), \quad \mathbf{p}^{(l)} = \mathrm{READOUT}(\mathbf{H}_p^{(l)}), \tag{12}$$

where $\mathrm{READOUT}(\cdot)$ denotes a global pooling operator. To tractably approximate this MI term, we optimize a symmetric InfoNCE loss over graph-prototype pairs.

(i) Graphs-to-prototype. For each instance representation $\mathbf{z}_i^{(l)}$, the corresponding granularity-specific prototype $\mathbf{p}^{(l)}$ serves as the positive sample, while prototypes from other granularities act as negative samples:

$$\mathcal{L}_{\mathrm{g2p}}^{(l)} = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\log\frac{\exp\Big(s(\mathbf{z}_i^{(l)},\mathbf{p}^{(l)})/\tau\Big)}{\sum_{l'}\exp\Big(s(\mathbf{z}_i^{(l)},\mathbf{p}^{(l')})/\tau\Big)}, \tag{13}$$

where $\mathcal{B}$ denotes the mini-batch, $s(\cdot,\cdot)$ is the cosine similarity function, and τ is the temperature hyper-parameter. This term encourages instance representations to align closely with their corresponding prototype while remaining distinguishable from prototypes at other granularities.

(ii) Prototype-to-graphs. Symmetrically, for each prototype $\mathbf{p}^{(l)}$, the instance representations $\mathbf{z}_i^{(l)}$ within the same granularity serve as positive samples, while instances from other granularities are treated as negatives:

$$\mathcal{L}_{\mathrm{p2g}}^{(l)} = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\log\frac{\exp\Big(s\Big(\mathbf{p}^{(l)},\mathbf{z}_i^{(l)}\Big)/\tau\Big)}{\sum_{l'}\sum_{j\in\mathcal{B}}\exp\Big(s\Big(\mathbf{p}^{(l)},\mathbf{z}_j^{(l')}\Big)/\tau\Big)}, \tag{14}$$

where j indexes instances in the mini-batch. This symmetric term ensures that prototypes are also pulled towards their corresponding instance representations.

Finally, the MI term $\mathcal{L}_{\mathrm{MI}}^{(l)}$ is expressed as the average of the two InfoNCE losses:

$$\mathcal{L}_{\mathrm{MI}}^{(l)} = \tfrac{1}{2}\Big(\mathcal{L}_{\mathrm{g2p}}^{(l)} + \mathcal{L}_{\mathrm{p2g}}^{(l)}\Big). \tag{15}$$

The SKL divergence term. For the SKL regularizer, we approximate $\mathcal{L}_{\mathrm{SKL}}^{(l)}$ using a tractable surrogate based on the consistency of cluster usage, i.e., the normalized frequency of node assignments to each cluster, between coarsening processes. To instantiate the SKL term in Eq. (5), we interpret $p_\theta(\mathcal{G}^{(l+1)} \mid \mathcal{G}^{(l)})$ and $q_\phi(\mathcal{G}_p^{(l+1)} \mid \mathcal{G}_p^{(l)})$ as implicit coarsening distributions induced by Gumbel-softmax pooling for the instance and prototype, respectively. As these distributions are intractable, we instead compare their permutation-invariant cluster usage summaries. At granularity l, the pooling module outputs a row-stochastic assignment matrix $\mathbf{S}^{(l)}$. Due to the permutation ambiguity of cluster indices, we summarize assignments by:

$$\mathbf{u}^{(l)} = \frac{1}{|\mathcal{V}^{(l)}|}\mathbf{1}_{|\mathcal{V}^{(l)}|}^{\top}\mathbf{S}^{(l)}, \tag{16}$$

where $\mathbf{u}^{(l)}$ denotes the cluster usage distribution induced by the assignment matrix $\mathbf{S}^{(l)}$. The SKL regularizer at granularity l is then defined as:

$$\mathcal{L}_{\mathrm{SKL}}^{(l)} = \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\frac{1}{2}\Big(\mathbf{u}_i^{(l)\top}\log\frac{\mathbf{u}_i^{(l)}+\epsilon}{\mathbf{u}_p^{(l)}+\epsilon} + \mathbf{u}_p^{(l)\top}\log\frac{\mathbf{u}_p^{(l)}+\epsilon}{\mathbf{u}_i^{(l)}+\epsilon}\Big), \tag{17}$$

where $\mathbf{u}_i^{(l)}$ and $\mathbf{u}_p^{(l)}$ denote the cluster usage distributions of the i-th instance graph and the prototype at granularity l, respectively, and $\epsilon > 0$ is a small constant for numerical stability. This symmetric formulation encourages the instance graphs and prototype graphs to adopt consistent coarsening behaviors while remaining invariant to cluster permutations.

Overall objective. Therefore, upon deriving tractable estimators for the MI and SKL terms, the overall training objective of M-GLAD can be formulated as:

$$\mathcal{L} = \sum_{l=0}^{L}\mathcal{L}_{\mathrm{PMIB}}^{(l)} = \sum_{l=0}^{L} -\mathcal{L}_{\mathrm{MI}}^{(l)} + \beta\,\mathcal{L}_{\mathrm{SKL}}^{(l)}. \tag{18}$$

4.5 Anomaly Scoring

Given a test graph $\mathcal{G}'$, we define its granularity-specific anomaly score as:

$$\mathrm{score}^{(l)}(\mathcal{G}') = -\mathcal{L}_{\mathrm{MI}}^{(l)}(\mathcal{G}'), \tag{19}$$

where larger values indicate lower consistency between the instance and the prototype, and consequently imply a

Metric	Method	PROTEINS-F	AIDS	BZR	COX2	ENZYMES	DD	IMDB-B	MUTAG
AUC	OCGIN	63.13 ± 1.40	97.52 ± 2.06	63.36 ± 4.25	55.23 ± 1.13	60.76 ± 1.78	62.99 ± 2.67	62.52 ± 2.14	76.89 ± 5.23
	OCGTL	60.76 ± 2.16	98.12 ± 0.53	66.03 ± 2.79	58.21 ± 2.27	59.65 ± 2.27	78.16 ± 1.54	63.76 ± 1.89	82.14 ± 3.36
	GOOD-D	67.50 ± 3.59	90.59 ± 1.62	61.14 ± 5.60	50.37 ± 7.38	57.59 ± 3.97	76.67 ± 2.99	62.89 ± 5.38	66.34 ± 4.40
	GLocalKD	60.63 ± 1.42	93.03 ± 3.52	59.09 ± 7.19	58.41 ± 9.20	61.01 ± 5.32	61.81 ± 4.53	65.56 ± 3.50	77.84 ± 6.70
	MUSE	66.59 ± 3.78	86.09 ± 9.76	68.65 ± 8.79	53.74 ± 3.84	47.68 ± 4.86	76.79 ± 3.18	68.74 ± 3.55	88.31 ± 6.26
	DO2HSC	56.11 ± 2.16	81.26 ± 8.23	66.47 ± 2.58	63.38 ± 4.91	58.10 ± 3.61	54.29 ± 4.62	59.42 ± 6.39	85.34 ± 8.79
	SIGNET	69.30 ± 3.76	96.90 ± 1.57	**72.05 ± 9.82**	71.78 ± 6.63	62.40 ± 4.20	71.08 ± 3.08	65.03 ± 4.86	87.12 ± 4.33
	GLADPro	65.97 ± 5.35	78.00 ± 8.35	67.04 ± 5.29	55.67 ± 5.93	52.78 ± 4.49	65.96 ± 3.21	70.62 ± 3.55	82.08 ± 7.67
	M-GLAD	**70.71 ± 1.06**	**99.23 ± 0.51**	69.60 ± 1.93	**72.19 ± 3.07**	**69.86 ± 6.43**	**78.82 ± 2.01**	**71.31 ± 2.98**	**94.37 ± 2.48**
AUPR	OCGIN	72.12 ± 1.01	94.07 ± 1.13	82.76 ± 1.18	77.39 ± 3.61	21.71 ± 0.89	72.12 ± 2.17	60.37 ± 2.58	62.98 ± 4.45
	OCGTL	71.47 ± 1.29	93.83 ± 1.42	81.25 ± 3.71	79.21 ± 0.82	22.97 ± 1.16	78.46 ± 1.73	61.93 ± 2.27	71.07 ± 2.64
	GOOD-D	70.06 ± 4.86	63.12 ± 3.86	80.14 ± 5.26	79.32 ± 4.08	25.21 ± 3.88	77.45 ± 3.19	63.17 ± 8.22	48.19 ± 5.08
	GLocalKD	64.94 ± 1.49	19.38 ± 1.58	83.16 ± 4.03	84.67 ± 4.61	19.69 ± 2.18	57.98 ± 3.06	67.04 ± 3.10	64.23 ± 9.10
	MUSE	71.04 ± 4.54	90.44 ± 8.25	85.70 ± 5.02	77.46 ± 3.39	20.26 ± 5.27	78.64 ± 2.78	65.15 ± 8.65	82.44 ± 8.41
	DO2HSC	54.21 ± 2.36	74.68 ± 4.72	84.96 ± 4.27	83.99 ± 1.65	22.67 ± 1.32	53.14 ± 2.31	63.29 ± 3.72	78.04 ± 10.08
	SIGNET	64.06 ± 2.26	79.74 ± 8.96	84.14 ± 6.23	85.34 ± 3.28	25.52 ± 2.64	75.19 ± 3.46	58.21 ± 9.59	76.36 ± 4.92
	GLADPro	57.73 ± 4.27	80.01 ± 2.34	36.77 ± 8.64	29.66 ± 6.78	24.24 ± 7.92	60.40 ± 6.18	**76.01 ± 2.13**	84.89 ± 9.97
	M-GLAD	**73.49 ± 2.17**	**96.03 ± 1.08**	**87.27 ± 2.44**	**89.31 ± 2.83**	**26.74 ± 2.42**	**79.26 ± 2.18**	67.29 ± 4.26	**86.90 ± 5.04**

Table 1: AUCs and AUPRs (%, mean ± std over 5 trials) of different GLAD methods on real-world datasets. Best results are shown in **bold**.

Method	AIDS	COX2	ENZYMES	IMDB-B	MUTAG
w/o SKL term	98.00	71.90	68.71	69.43	91.73
w/o Coarsening	97.89	65.19	63.13	68.83	90.41
w/o Prototype	96.74	66.58	64.42	67.84	89.36
M-GLAD	**99.23**	**72.19**	**69.86**	**71.31**	**94.37**

Table 2: Ablation study on five datasets w.r.t. AUCs (%).

higher likelihood of being anomalous. We then aggregate $\{\text{score}^{(l)}(\mathcal{G}')\}_{l=0}^{L}$ according to the scoring strategy $\text{Agg}(\cdot)$:

$$\text{score}(\mathcal{G}') = \text{Agg}\Big(\{\text{score}^{(l)}(\mathcal{G}')\}_{l=0}^{L}\Big), \quad (20)$$

where $\text{Agg}(\cdot)$ is instantiated as *avg*, *max*, or a weighted function. For the weighted variant, we adopt geometrically decaying aggregation across different levels of structure granularity. Specifically,

$$\text{Agg}\Big(\{\text{score}^{(l)}(\mathcal{G}')\}_{l=0}^{L}\Big) = \sum_{l=0}^{L} w_l \, \text{score}^{(l)}(\mathcal{G}'), \quad (21)$$

where $w_l = \frac{\gamma^l}{\sum_{k=0}^{L}\gamma^k}$, l and k index the granularity level, and γ^l denotes the exponential weighting factor assigned to granularity l. Larger γ assigns higher weights to the deeper levels of structure granularity. Time complexity analysis can be found in supplementary materials.

5 Experiments

5.1 Experimental Setup

Datasets. We conduct experiments on eight graph datasets from the TU Dataset Collection [Morris *et al.*, 2020]. These datasets span multiple domains, including bioinformatics networks (PROTEINS-F, ENZYMES, DD), molecular networks (AIDS, BZR, COX2, MUTAG), and social networks (IMDB-B), following the setting in [Liu *et al.*, 2023].

Baselines. We compare M-GLAD with the following state-of-the-art GLAD methods: (1) **Global-aggregation based GLAD methods.** These approaches, including OCGIN [Zhao and Akoglu, 2020], OCGTL [Qiu *et al.*, 2022], GOOD-D [Liu *et al.*, 2022], GLocalKD [Ma *et al.*, 2022], MUSE [Kim *et al.*, 2024], and DO2HSC [Zhang *et al.*, 2024], adopt a global perspective by learning node-level representations via GNNs and aggregating them into graph-level embeddings for anomaly scoring. (2) **Subgraph-oriented GLAD methods.** These methods, including SIGNET [Liu *et al.*, 2023] and GLADPro [Yang *et al.*, 2025], aim to identify subgraphs that can distinguish anomalous graphs from normal ones.

Metrics and Implementation. We use GIN [Xu *et al.*, 2019] as the GNN backbone. The default number of granularities is set to $L = 2$, and the trade-off coefficient is $\beta = 0.01$. For GLAD performance, we report AUC and AUPR as evaluation metrics. All experiments are repeated five times with different random seeds, and the average performance and standard deviation are reported. Our code is available at https://github.com/Lab821-gz/M-GLAD. More settings are provided in supplementary materials.

5.2 Main Results

To evaluate the performance of M-GLAD, we compare it with eight state-of-the-art methods on eight real-world datasets. Based on the results in Table 1, we make the following observations. (1) M-GLAD achieves the best performance on 7 out of 8 TU benchmark datasets in terms of both AUC and AUPR. These results demonstrate the effectiveness and robustness of M-GLAD across diverse domains. (2) Global aggregation-based detectors such as GLocalKD, MUSE, and DO2HSC exhibit considerable performance variability across datasets, with large standard deviations on datastes including AIDS, BZR, and MUTAG. (3) Subgraph-oriented methods, particularly GLADPro, show notably lower AUPR scores on

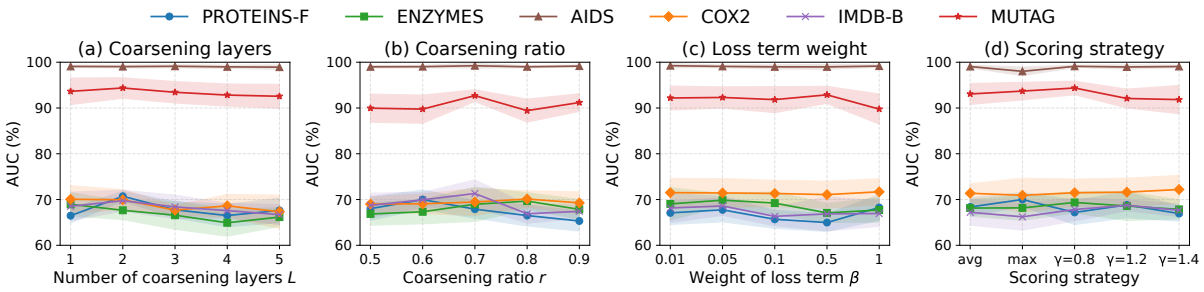
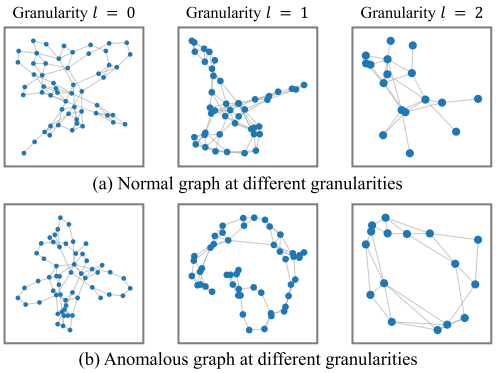


Figure 3: Hyper-parameter sensitivity analysis of M-GLAD.



(a) Normal graph at different granularities

(b) Anomalous graph at different granularities

Figure 4: Multi-granular structural visualizations of normal and anomalous graphs on the PROTEINS-F dataset.

datasets such as PROTEINS-F, BZR, and COX2. The experiment results suggest that anomalous graphs are typically defined by multi-level structural patterns.

5.3 Ablation Study

To evaluate the importance of different components in M-GLAD, we conduct three types of ablation studies. The experimental results are reported in Table 2. (1) We test the impact of the SKL divergence regularization term. Without the SKL regularization that enforces consistency between the prototype view and the instance graph view, the model fails to effectively align the extracted prototype features with the graph information, resulting in consistent performance degradation. (2) We evaluate the role of graph structure abstraction module. Removing the graph structure abstraction module degrades the ability to capture informative structural patterns at different granularities, resulting in inferior GLAD performance across datasets. (3) We examine the effect of prototype modeling by replacing the learned prototypes with randomly initialized ones. This replacement leads to a noticeable performance drop, indicating that data-driven prototype learning is crucial for capturing intrinsic graph structures and summarizing typical normal patterns.

5.4 Parameter Sensitivity

We evaluate the sensitivity to M-GLAD of four key hyper-parameters: the number of coarsening layers L, the coarsening ratio r, the loss weight β, and the scoring strategy $\mathrm{Agg}(\cdot)$. As shown in Figure 3(a), varying L from 1 to 5 results in only minor performance changes, with most datasets such as PROTEINS-F, IMDB-B, and MUTAG achieving optimal or near-optimal performance at $L = 2$. This indicates that moderate graph abstraction is sufficient, as too few coarsening layers may capture insufficient structural information, while overly deep coarsening can introduce noise and degrade performance. Figure 3(b) illustrates the impact of the coarsening ratio r. A moderate ratio provides an effective balance between redundancy removal and structure preservation. On most datasets such as MUTAG, IMDB-B, and COX2, the best performance is achieved at $r = 0.7$ or 0.8, while extreme values degrade performance due to under- or over-coarsening. Regarding the loss weight β, Figure 3(c) shows stable performance over a wide range, while overly large values degrade performance, indicating that the auxiliary loss acts as an effective regularizer without dominating optimization. Finally, Figure 3(d) shows that weighted scoring slightly outperforms *avg* and *max* while remaining stable across different γ values. This suggests that controlled aggregation of anomaly evidence across multiple coarsening levels is more effective, as the most discriminative granularity varies across datasets.

5.5 A Case Study

Figure 4 visualizes how anomalies manifest at different granularities under graph structure abstraction. For $l = 0$ and $l = 1$, the anomalous graph appears similar to the normal one, showing no clear abnormal patterns from the visualized structures. In contrast, at the coarser granularity $l = 2$, a clear discrepancy emerges: the anomalous graph forms a ring-like structure that is absent from the normal graph. This suggests that anomalies can be more distinguishable at certain granularities, thereby motivating multi-granularity modeling.

6 Conclusion

In this paper, we propose M-GLAD, an unsupervised GLAD framework that models anomalies from a multi-granular structural perspective. M-GLAD progressively abstracts graph structures into multiple granularities and adopts the Prototype-guided Multi-granular Information Bottleneck (PMIB) to characterize normal patterns with learnable prototypes and quantify anomaly evidence via cross-granularity discrepancies. By aggregating such evidence, M-GLAD enables granularity-aware anomaly scoring. Extensive experiments on eight real-world graph datasets show that M-GLAD consistently outperforms the competitive methods.

Acknowledgments

This work was supported in part by the Shanghai Pujiang Program (No. 24PJA004), NSFC (No. 92570118, 62376088 and 62372146), Hebei Natural Science Foundation (No. F2024202047), Hebei Yanzhao Golden Platform Talent Gathering Programme Core Talent Project (Education Platform) (No. HJZD202509), Fundamental Research Funds for the Central Universities (No. 26D112010 and 2232025D-33).

References

[Bachman *et al.*, 2019] Philip Bachman, R Devon Hjelm, and William Buchwalter. Learning representations by maximizing mutual information across views. *Advances in neural information processing systems*, 32, 2019.

[Cai *et al.*, 2025] Jinyu Cai, Yunhe Zhang, and Jicong Fan. Self-discriminative modeling for anomalous graph detection. In *Forty-second International Conference on Machine Learning*, 2025.

[Chen *et al.*, 2025] Qingfeng Chen, Haojin Zeng, Jingyi Jie, Shichao Zhang, and Debo Cheng. Denoise: Learning robust graph representations for unsupervised graph-level anomaly detection. *arXiv preprint arXiv:2511.04086*, 2025.

[Dai and Wang, 2025] Enyan Dai and Suhang Wang. Towards prototype-based self-explainable graph neural network. *ACM Transactions on Knowledge Discovery from Data*, 19(2):1–20, 2025.

[Dong *et al.*, 2024] Xiangyu Dong, Xingyi Zhang, and Sibo Wang. Rayleigh quotient graph neural networks for graph-level anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2024.

[Duval and Malliaros, 2022] Alexandre Duval and Fragkiskos Malliaros. Higher-order clustering and pooling for graph neural networks. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pages 426–435, 2022.

[Gao and Ji, 2019] Hongyang Gao and Shuiwang Ji. Graph u-nets. In *international conference on machine learning*, pages 2083–2092, 2019.

[Guo *et al.*, 2022] Zhichun Guo, Bozhao Nan, Yijun Tian, O. Wiest, Chuxu Zhang, and N. Chawla. Graph-based molecular representation learning. In *International Joint Conference on Artificial Intelligence*, 2022.

[Hassani and Khasahmadi, 2020] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *International conference on machine learning*, pages 4116–4126, 2020.

[Islam *et al.*, 2023] Muhammad Ifte Khairul Islam, Max Khanov, and Esra Akbas. Mpool: motif-based graph pooling. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 105–117. Springer, 2023.

[Jang *et al.*, 2017] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *ICLR*, 2017.

[Ju *et al.*, 2024] Wei Ju, Siyu Yi, Yifan Wang, Qingqing Long, Junyu Luo, Zhiping Xiao, and Ming Zhang. A survey of data-efficient graph learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, pages 8104–8113, 2024.

[Kim *et al.*, 2024] Sunwoo Kim, Soo Yong Lee, Fanchen Bu, Shinhwan Kang, Kyungho Kim, Jaemin Yoo, and Kijung Shin. Rethinking reconstruction-based graph-level anomaly detection: Limitations and a simple remedy. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[Lee *et al.*, 2019] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In *International conference on machine learning*, pages 3734–3743, 2019.

[Lee *et al.*, 2024] Jongha Lee, Sunwoo Kim, and Kijung Shin. Slade: Detecting dynamic anomalies in edge streams without labels via self-supervised learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1506–1517, 2024.

[Lin *et al.*, 2024] Xixun Lin, Wenxiao Zhang, Fengzhao Shi, Chuan Zhou, Lixin Zou, Xiangyu Zhao, Dawei Yin, Shirui Pan, and Yanan Cao. Graph neural stochastic diffusion for estimating uncertainty in node classification. In *Forty-first international conference on machine learning*, 2024.

[Lin *et al.*, 2025] Xixun Lin, Qing Yu, Yanan Cao, Lixin Zou, Chuan Zhou, Jia Wu, Chenliang Li, Peng Zhang, and Shirui Pan. Generative causality-driven network for graph multi-task learning. *IEEE transactions on pattern analysis and machine intelligence*, 2025.

[Liu *et al.*, 2022] Yixin Liu, Kaize Ding, Huan Liu, and Shirui Pan. Good-d: On unsupervised graph out-of-distribution detection. *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 2022.

[Liu *et al.*, 2023] Yixin Liu, Kaize Ding, Qinghua Lu, Fuyi Li, Leo Yu Zhang, and Shirui Pan. Towards self-interpretable graph-level anomaly detection. *NeurIPS*, 36:8975–8987, 2023.

[Luo *et al.*, 2022] Xuexiong Luo, Jia Wu, Jian Yang, Shan Xue, Hao Peng, Chuan Zhou, Hongyang Chen, Zhao Li, and Quan Z Sheng. Deep graph level anomaly detection with contrastive learning. *Scientific Reports*, 12(1):19867, 2022.

[Ma *et al.*, 2022] Rongrong Ma, Guansong Pang, Ling Chen, and Anton Van Den Hengel. Deep graph-level anomaly detection by glocal knowledge distillation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, pages 704–714, 2022.

[Ma *et al.*, 2023] Xiaoxiao Ma, Jia Wu, Jian Yang, and Quan Z Sheng. Towards graph-level anomaly detection via deep evolutionary mapping. In *Proceedings of the*

29th ACM SIGKDD conference on knowledge discovery and data mining, pages 1631–1642, 2023.

[Morris *et al.*, 2020] Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond*, 2020.

[Niu *et al.*, 2025] Chaoxi Niu, Hezhe Qiao, Changlu Chen, Ling Chen, and Guansong Pang. Zero-shot generalist graph anomaly detection with unified neighborhood prompts, 2025.

[Qiao *et al.*, 2025] Hezhe Qiao, Hanghang Tong, Bo An, Irwin King, Charu Aggarwal, and Guansong Pang. Deep graph anomaly detection: A survey and new perspectives. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Qiu *et al.*, 2022] Chen Qiu, M. Kloft, Stephan Mandt, and Maja R. Rudolph. Raising the bar in graph-level anomaly detection. In *International Joint Conference on Artificial Intelligence*, 2022.

[Rajput and Singh, 2022] Nitendra Rajput and Karamjit Singh. Temporal graph learning for financial world: Algorithms, scalability, explainability & fairness. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4818–4819, 2022.

[Tishby and Zaslavsky, 2015] Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 ieee information theory workshop (itw)*, pages 1–5. Ieee, 2015.

[Tishby *et al.*, 2000] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

[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] Yuwen Wang, Shunyu Liu, Tongya Zheng, Kaixuan Chen, and Mingli Song. Unveiling global interactive patterns across graphs: Towards interpretable graph neural networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3277–3288, 2024.

[Wu *et al.*, 2020] Tailin Wu, Hongyu Ren, Pan Li, and Jure Leskovec. Graph information bottleneck. *NeurIPS*, 33:20437–20448, 2020.

[Xia *et al.*, 2021] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2):109–127, 2021.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16129–16137, 2024.

[Yang *et al.*, 2025] Zhenyu Yang, Ge Zhang, Jia Wu, Jian Yang, Shan Xue, Amin Beheshti, Hao Peng, and Quan Z Sheng. Global interpretable graph-level anomaly detection via prototype. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 3586–3597, 2025.

[Yang *et al.*, 2026] Zhenyu Yang, Ge Zhang, Shan Xue, Xiaoxiao Ma, Jian Yang, Hao Peng, Amin Beheshti, and Jia Wu. Revisiting graph-level anomaly detection: From partially to fully unsupervised learning. In *Proceedings of the ACM Web Conference 2026*, pages 901–912, 2026.

[Ying *et al.*, 2018] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *Advances in neural information processing systems*, 31, 2018.

[Zhang and Li, 2021] Muhan Zhang and Pan Li. Nested graph neural networks. *Advances in Neural Information Processing Systems*, 34:15734–15747, 2021.

[Zhang *et al.*, 2021] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882, 2021.

[Zhang *et al.*, 2022a] Ge Zhang, Zhenyu Yang, Jia Wu, Jian Yang, Shan Xue, Hao Peng, Jianlin Su, Chuan Zhou, Quan Z Sheng, Leman Akoglu, et al. Dual-discriminative graph neural network for imbalanced graph-level anomaly detection. *Advances in Neural Information Processing Systems*, 35:24144–24157, 2022.

[Zhang *et al.*, 2022b] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Cheekong Lee. Protgnn: Towards self-explaining graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 9127–9135, 2022.

[Zhang *et al.*, 2024] Yunhe Zhang, Yan Sun, Jinyu Cai, and Jicong Fan. Deep orthogonal hypersphere compression for anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2024.

[Zhang *et al.*, 2025] Ge Zhang, Zhenyu Yang, Jia Wu, Pengfei Jiao, Jian Yang, Hao Peng, and Xixun Lin. Learning subgraph-based normality for interpretable graph-level anomaly detection. *IEEE Transactions on Information Forensics and Security*, 21:288–303, 2025.

[Zhang *et al.*, 2026] Ge Zhang, Jiapei Chen, Guohao Sun, Xiu Fang, Zhenyu Yang, Xixun Lin, and Liang Yang. Generalizable graph-level anomaly detection via prompted anomaly expansion and normality extraction. In *WWW*, pages 1068–1079, 2026.

[Zhao and Akoglu, 2020] Lingxiao Zhao and Leman Akoglu. On using classification datasets to evaluate graph outlier detection: Peculiar observations and new insights. *Big Data*, 11:151 – 180, 2020.

[Zhao *et al.*, 2022] Lingxiao Zhao, Wei Jin, Leman Akoglu, and Neil Shah. From stars to subgraphs: Uplifting any GNN with local structure awareness. In *ICLR*, 2022.