GRAPH FOUNDATION MODEL FOR CROSS-GRANULARITY ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Foundation Models (GFMs) for Graph Anomaly Detection (GAD) have gained massive attention, playing important roles in academic and industrial circles. Existing models are mainly divided into two types: granularity-specific cross-domain models (GSCD) and cross-granularity domain-specific models (CGDS), which focus on improving cross-domain ability or cross-granularity ability. However, they fail to explore GAD under cross-granularity cross-domain settings. The aim of this study is to provide a unified solution for GFMs in cross-granularity GAD scenarios. To solve this problem, we propose a novel framework, CgGFAD, specializing in two challenges in this problem. On the one hand, to capture granularity-specific semantics and enhance cross-granularity detection ability, CgGFAD unifies the representations for different granularities and employs maximum mean discrepancy based on the Wasserstein distance to evaluate the similarity between granularities. On the other hand, to capture comprehensive granularity-specific representations and achieve cross-domain generalization, CgGFAD develops a cluster-based self-distillation and prototype generalization module, which leverages deep clustering strategies to derive latent semantic patterns and deploys prototype approximation to achieve cross-domain detection. Extensive experiments on diverse public datasets, compared with state-of-the-art baselines, demonstrate the advancement and effectiveness of CgGFAD.

1 INTRODUCTION

Graph Anomaly Detection (GAD) seeks to identify graph entities whose structural topologies or attribute distributions exhibit significant deviations from the dominant patterns of normal data. As a result, it has played an indispensable role in numerous real-world applications, such as rumor detection (Zhang et al., 2026b), social media mining (Mohawesh et al., 2026), and financial systems (Zhang et al., 2025). However, conventional deep learning methods are confined to single-task and dataset-specific modeling, which fail to meet the rapidly evolving detection demands, especially for multi-granularity samples and generalist detection. This largely limits the scalability and adaptability of GAD models in complex scenarios and diverse applications.

Recent studies have increasingly focused on developing Graph Foundation Models (GFMs) (Shi et al., 2026; Yuan et al., 2026) to improve the scalability and adaptability of graph-based tasks. Rather than specializing in a single dataset or predefined objective, GFMs exploit large-scale pretraining over diverse graph structures and self-supervised tasks to acquire reusable and transferable graph representations, thereby improving their adaptability to downstream applications. GFMs have now been successfully extended to GAD, which are mainly divided into two directions: granularity-specific cross-domain models (GSCD) and cross-granularity domain-specific models (CGDS).

GSCD emphasizes cross-domain generalizability by learning generalized graph representations, such as ARC (Liu et al., 2026a) and AnomalyGFM (Qiao et al., 2025). However, they are designed for only a single-granularity anomaly detection task, especially for the node level. Such a setting overlooks cross-granularity detection, failing to achieve transferability between different granularities. In practice, node-, edge-, and graph-level anomalies are often closely coupled, where abnormal local behaviors may propagate to edge irregularities and further accumulate into graph-level deviations. For instance, the pathological diagnosis of cells is often jointly determined by abnormalities in particular proteins and their associations. Ignoring these cross-granularity dependencies limits

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the model's ability to capture inadequate anomaly semantics and consequently weakens its robustness in complex cross-granularity scenarios. CGDS extends GAD to multi-granularity settings but sacrifices cross-domain generalization, such as UniFORM (Song et al., 2025) and GFM-UAD (Han et al., 2026). Apparently, these methods achieve unification across different granularities; however, they are typically designed for specific datasets, showing limited knowledge transferability to other unseen datasets and poor cross-domain adaptability.

Although CGDS and GSCD have achieved promising progress towards these two directions, they fail to function in cross-granularity, cross-domain GAD scenarios, which limits the scalability of GAD in complex applications and tasks. In this study, we initiate an exploration of a unified framework for GFMs in cross-granularity GAD scenarios. Solving this problem requires tackling twofold challenges. First, it entails acquiring transferable knowledge across multiple granularities in an unsupervised manner. Second, a well-trained model on the source domain is required to achieve generalization towards unseen target domains where anomalies may exhibit inherently distinct characteristics and behavior patterns.

To address these challenges, we propose a cross-granularity GFM for GAD (CgGFAD). For the first challenge, CgGFAD introduces Maximum Mean Discrepancy (MMD) (Yang et al., 2025) with Wasserstein distance to amplify discrimination of granularity-specific representations, enhancing the discrimination of different granularities. Besides, CgGFAD designs a unified alignment strategy based on SimSiam contrastive learning towards node-level representations to maintain latent semantics. For the second challenge, CgGFAD develops a self-distillation framework based on deep clustering to capture comprehensive normal patterns, and then employs prototype optimization to improve cross-domain generalization ability. In the inference stage, CgGFAD is frozen and deployed to obtain the latent representations for a testing granularity and calculates the anomaly scores based on the distillation deviation and the prototype deviation. Our contributions are as follows:

- **A new problem**. We make the first attempt to study GFMs for GAD in cross-granularity scenarios in an unsupervised manner, which enhances the cross-domain generalization to unseen datasets and the discrimination of cross-granularity anomalies.
- **A novel framework**. We propose CgGFAD that utilizes Wasserstein-based MMD to obtain discriminative semantics across granularity, and then integrates self-distillation and prototype guidance to capture latent normal patterns and achieve cross-domain generalization.
- **Advanced performance**. Extensive experiments on diverse public datasets demonstrate the superior generalization of CgGFAD and cross-granularity detection ability compared to state-of-the-art baselines.

2 Related Work

Traditional GAD. Existing graph anomaly detection methods are mainly classified into supervised and unsupervised methods. Supervised methods utilize labels to discriminate complex graph attributes and topological characteristics. For example, GE-GNN (Zhang & Luo, 2025a) introduces augmented edge information and a gate mechanism to capture anomaly features. Unsupervised GAD primarily includes contrastive learning-based methods (Zhang & Luo, 2026) and reconstruction-based methods (Chen et al., 2026b). Some studies also introduce advanced technologies such as diffusion (Wei et al., 2026; Xiao et al., 2025) and adversarial learning (Chen et al., 2026a) to capture the latent representations. *Although they achieve promising progress, they focus on specific tasks, especially node-level anomaly detection, and suffer from limited detection ability on unseen datasets.*

Graph Foundation Models for GAD. GFMs for GAD aim to enhance cross-dataset anomaly detection on graph data. Some approaches introduce in-context learning, such as ARC (Liu et al., 2026a) and SAARCS (Liu et al., 2026b). Besides, many studies develop elaborate tuning strategies to enhance cross-domain detection. For example, UNPrompt (Niu et al., 2025), DPEAD Ran et al. (2025), and AffinityTune Chen et al. (2025) introduce diverse prompt strategies to achieve generalized detection ability. Reconstruction is a primary branch of GFMs for GAD, including spectral signal reconstruction in Zero-GAD (Zheng et al., 2025) and invariant feature discovery in Ia-GGAD (Zhang et al., 2026c). In addition, DP-DGAD Zheng et al. (2026) exploits prototypes to achieve dataset generalization. *However, existing GFMs for GAD fail to provide a unified so-*

lution for cross-granularity anomaly detection tasks, limiting the scalability of GFMs in complex application scenarios.

Multi-granularity Graph Anomaly Detection. Several studies have attempted to address multi-granularity GAD. BOURNE (Liu et al., 2024b) detects node- and edge-level anomalies simultaneously by constructing dual hypergraph representations. UniGAD (Lin et al., 2024) generates a unified subgraph domain via maximum Rayleigh quotient subgraph sampling to unify multi-granularity GAD. UniFORM (Song et al., 2025) introduces an energy-based encoder to characterize anomaly uncertainty across diverse granularities. GFM-UAD (Han et al., 2026) further unifies multi-granularity GAD with cross-domain generalization detection through adversarial learning and fine-tuning. More recently, GRAND (Wang et al., 2026) presents a unified framework for node- and edge-level anomaly detection by combining dual diffusion processes with Transformers as the denoising backbone. *However, they only focus on a specific domain, overlooking the cross-domain generalization ability for anomaly detection on unseen datasets, resulting in limited scalability.*

Section A analyzes some similar studies on graphs to illustrate the difference.

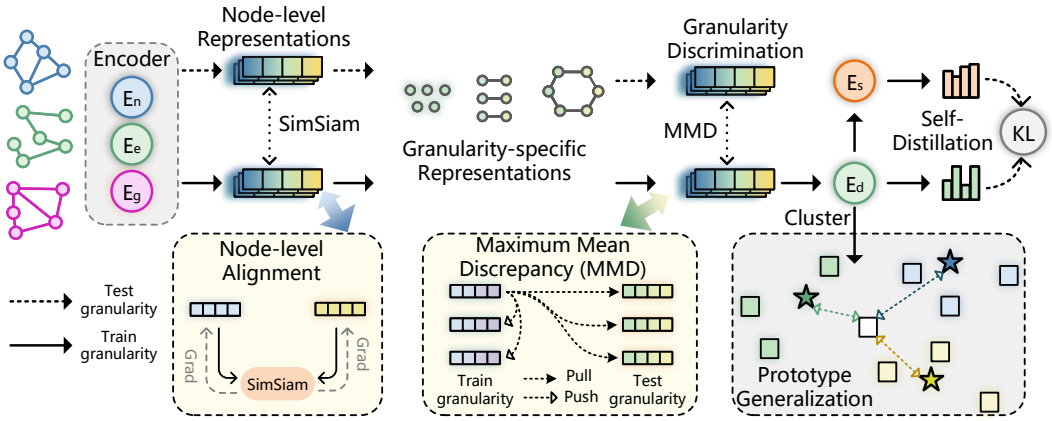

Figure 1: The framework of CgGFAD.

3 PRELIMINARIES

We denote a graph as $G = (\mathbf{X}, \mathbf{A})$, where $\mathbf{X} \in \mathbb{R}^{N \times d}$ is the node feature matrix, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix. N is the number of nodes and d is the dimension of node attributes. $\mathbf{x}_i \in \mathbb{R}^d = \mathbf{X}[i, :]$ denotes the features of the i-th node. $a_{ij} \in \mathbf{A} = 1$ denote there is an edge e_{ij} from i-th node to j-th node v_j. Define the task set as $\mathcal{T} = \{t_n, t_d, t_g\}$, where t_n, t_d, and t_g respectively represent node-, edge-, and graph-granularity learning tasks. Given a training graph G_{tr} on a specific granularity task $t_{\text{tr}} \in \mathcal{T}$, the aim is to learn a unified model to predict anomalies on an untrained graph G_{te} at granularity t_{te}, where $t_{\text{te}} \neq t_{\text{tr}}, t_{\text{te}} \in \mathcal{T}$. In this problem setting, the representations of training and testing granularities in G_{tr} is accessible. G_{te} is inaccessible during the training stage.

4 METHODOLOGY

In this section, we introduce the details of CgGFAD. The framework is shown in Fig. 1. Specifically, CgGFAD first implements granularity-independent encoding using different encoders to obtain node-level representations and makes unified alignment using SimSiam. Then, CgGFAD obtains granularity-specific representations using node-level representations. Based on these specific representations, CgGFAD employs the 2-Wasserstein distance to calculate MMD for training and testing granularities for granularity discrimination. To achieve unsupervised training and cross-domain detection ability, CgGFAD develops a cluster-based self-distillation and prototype generalization module, which evaluates the KL divergence between prediction and target distributions and calculates the distance between prototypes and the assigned samples. In the inference stage, CgGFAD estimates the anomaly scores via the distillation deviation and prototype deviation without retraining on unseen datasets.

4.1 NODE-LEVEL ALIGNMENT AND GRANULARITY ENCODING

To learn the distinctive semantic information for different granularities, CgGFAD first obtains unified node-level representations for each granularity using independent graph encoders and then derives granularity-specific representations based on node-level embeddings. Notably, utilizing granularity-unique encoders can flexibly model latent feature spaces, enhancing semantic disentanglement and cross-granularity discrimination. Specifically, node-level representations for different granularities can be generated as:

$$\mathbf{H}_{t_n} = \mathbf{E}_n(G_{\text{tr}}), \mathbf{H}_{t_d} = \mathbf{E}_d(G_{\text{tr}}), \mathbf{H}_{t_g} = \mathbf{E}_g(G_{\text{tr}}), \tag{1}$$

where $\mathbf{H}_{t_n}, \mathbf{H}_{t_d}, \mathbf{H}_{t_g} \in \mathbb{R}^{N \times d'}$ are node-level representations for node, edge, and graph granularities, $\mathbf{E}_n$, $\mathbf{E}_d$, and $\mathbf{E}_g$ are graph encoders for node, edge, and graph granularities, and d' is the hidden dimension. $\mathbf{H}_{t_n}$ can be directly regarded as node-granularity representations $\hat{\mathbf{H}}_{t_n} \in \mathbb{R}^{N_{t_n} \times d'}$, where $N_{t_n} = N$ is the number of samples of node granularity. Edge-granularity representations from i-th node to j-th node can be obtained with $\hat{\mathbf{h}}_{t_d}^{(ij)} = \mathbf{h}_{t_d}^{i} - \mathbf{h}_{t_d}^{j}$ and the corresponding matrix form is $\hat{\mathbf{H}}_{t_d} \in \mathbb{R}^{N_{t_d} \times d'}$, where N_{t_d} is the number of samples of edge granularity. Graph-granularity representations can be derived using $\hat{\mathbf{h}}_{t_g} = \text{Mn}(\mathbf{H}_{t_g})$ and the corresponding matrix form is $\hat{\mathbf{H}}_{t_g} \in \mathbb{R}^{N_{t_g} \times d'}$, where N_{t_g} is the number of samples of graph granularity and Mn is the MeanPool function.

Although granularity-specific representations should exhibit distinctive characteristics, they are derived from identical original datasets. To maintain intrinsic graph topological consistencies for different encoders, CgGFAD leverages contrastive learning to achieve node-level alignment. Without loss of generality, denote training granularity as t_{tr}, testing granularity as t_{te}, and their corresponding node-level representations in the training dataset are denoted as $\mathbf{H}_{t_{\text{tr}}}$ and $\mathbf{H}_{t_{\text{te}}}$. SimSiam (Zhang & Luo, 2025b) is deployed to implement node-level alignment:

$$\mathcal{L}_c = -\frac{1}{2}[\cos(\text{sg}(\mathbf{H}_{t_{\text{tr}}}), \mathbf{H}_{t_{\text{te}}}) + \cos(\mathbf{H}_{t_{\text{tr}}}, \text{sg}(\mathbf{H}_{t_{\text{te}}}))], \tag{2}$$

where $\cos(\cdot, \cdot)$ is instance-level cosine similarity function and $\text{sg}(\cdot)$ is stop gradient. SimSiam is an effective and efficient contrastive learning paradigm, which improves the sampling strategy and saves the computational cost for negative sampling. Besides, $\text{sg}(\cdot)$ enables iterative optimization, preventing representation collapse (Zhang et al., 2022).

4.2 SEMANTIC-AWARE GRANULARITY DISCRIMINATION

Semantic overlap across granularities and lack of discriminability can compromise the ability to identify granularity-specific anomaly characteristics. To effectively bridge granularity-unique characteristics, it is critically important to discriminate granularity distributions and corresponding behavioral patterns. Hence, CgGFAD introduces semantic-aware granularity discrimination to isolate the latent components, enabling clear discrepancies and distinct features.

To explore the discrepancies of patterns under different granularities, we denote $p_{t_k}^{(i)} = \mathcal{N}(\mu_{t_k}^{(i)}, \sigma_{t_k}^{(i)})$ as the distribution of the i-th instance in granularity $t_k (k = 1, 2)$. $\mu_{t_k}^{(i)}, \sigma_{t_k}^{(i)} \in \mathbb{R}$ are the empirical mean and standard deviation of the i-th instance across d' dimension. Instead of directly using vanilla Euclidean distance, which disregards the geometric support structure of the distribution, we employ the optimal transport distance (Zhang & Luo, 2026) to provide a geometry-aware similarity metric between probability distributions. Since the covariance matrices are diagonal, we simplify the 2-Wasserstein distance ϕ between two instances:

$$\phi(p_{t_1}^{(i)}, p_{t_2}^{(j)}) = (\mu_{t_1}^{(i)} - \mu_{t_2}^{(j)})^2 + (\sigma_{t_1}^{(i)} - \sigma_{t_2}^{(j)})^2. \tag{3}$$

Note that $t_1 = t_2$ indicates the similarity between samples within the same granularity, and $t_1 \neq t_2$ represents the similarity between samples across different granularities.

To optimize the encoders, we aim to maximize the discrimination between different granularities. MMD is a widely applied paradigm, such as in transfer learning (Yang et al., 2025), to evaluate the dissimilarity between distributions by comparing their means in high-dimensional space. However, traditional MMD usually relies on cosine similarity, failing to capture distribution characteristics

between two granularities. Instead, we design an MMD-based loss function to bridge the discrimination between training and testing granularities, t_{tr} and t_{te}. The loss function is formulated as:

$$\begin{aligned}\mathcal{L}_g = &\frac{1}{\sum_{i=1}^{N_{t_{\text{tr}}}}\sum_{i'=1}^{N_{t_{\text{tr}}}}\delta_{i,i'}}\sum_{i=1}^{N_{t_{\text{tr}}}}\sum_{i'=1}^{N_{t_{\text{tr}}}}\delta_{i,i'}\phi(p_{t_{\text{tr}}}^{(i)}, p_{t_{\text{tr}}}^{(i')}) \\ &+ \frac{1}{\sum_{j=1}^{N_{t_{\text{te}}}}\sum_{j'=1}^{N_{t_{\text{te}}}}\delta_{j,j'}}\sum_{j=1}^{N_{t_{\text{te}}}}\sum_{j'=1}^{N_{t_{\text{te}}}}\delta_{j,j'}\phi(p_{t_{\text{te}}}^{(j)}, p_{t_{\text{te}}}^{(j')}) \\ &- \frac{2}{N_{t_{\text{tr}}}N_{t_{\text{te}}}}\sum_{i=1}^{N_{t_{\text{tr}}}}\sum_{j=1}^{N_{t_{\text{te}}}}\phi(p_{t_{\text{tr}}}^{(i)}, p_{t_{\text{te}}}^{(j)}),\end{aligned} \tag{4}$$

where $\delta_{i,i'}, \delta_{j,j'}$ are semantic-aware scores. The first and second terms of $\mathcal{L}_g$ minimize the distance between samples within the same granularity, while the third term maximizes the distance between samples in granularity t_1 and t_2. With this optimization objective, encoders can be trained to identify distinctive semantics for different granularities, avoiding the overlaps of semantic patterns across granularities. Note that this loss function fully exploits representations of training and testing granularities in the training graph, and maintains inaccessibility to testing datasets.

However, samples in specific granularities are not exclusively and equivalently distributed. To better evaluate the semantic contributions of samples within the granularity, CgGFAD introduces prototypes to generate semantic-aware scores for intra-granularity evaluation, the first two items in Eq. (4). Specifically, for the representations $\hat{\mathbf{H}}_{t_k}$ in granularity $t_k \in \{t_{\text{tr}}, t_{\text{te}}\}$, we calculate the granularity prototype $\mathbf{c}_{t_k} = \text{Mn}(\hat{\mathbf{H}}_{t_k}) \in \mathbb{R}^{d'}$ and the corresponding distribution characteristics $p_{t_k}^{\mathbf{c}} = \mathcal{N}(\mu_{t_k}^{\mathbf{c}}, \sigma_{t_k}^{\mathbf{c}})$. Then, for i-th representation $\mathbf{h}_{t_k}^{(i)}$, we compute the similarity using 2-Wasserstein distance $\phi(\mathbf{h}_{t_k}^{(i)}, \mathbf{c}_{t_k}) \in \mathbb{R}$. Given two representations $\mathbf{h}_{t_k}^{(i)}$ and $\mathbf{h}_{t_k}^{(i')}$, the semantic-aware score $\delta_{i,i'}$ can be obtained using $\delta_{i,i'} = \log(1 + \phi(\mathbf{h}_{t_k}^{(i)}, \mathbf{c}_{t_k})/\phi(\mathbf{h}_{t_k}^{(i')}, \mathbf{c}_{t_k}))$.

4.3 Cluster-based Self-distillation and Prototype Generalization

To address the second challenge, on the one hand, CgGFAD captures comprehensive normal patterns of samples within the granularity via knowledge distillation, which is an effective way to capture normal patterns of latent representations. On the other hand, CgGFAD introduces prototype generalization to achieve cross-domain anomaly detection.

Traditional distillation requires well-trained teacher networks, which consume massive computing resources while lacking flexibility and scalability. Therefore, CgGFAD introduces a cluster-based self-distillation strategy, which generates sharpened target and prediction distributions from prototype-guided probabilities and learns their consistency through KL divergence. With the teacher-free architecture, this strategy achieves efficient pattern modeling without additional training and inference overhead.

Specifically, for the training granularity t_{tr}, we introduce a lightweight perceptron to transform the embeddings:

$$\mathbf{Z} = \mathbf{E}_d(\hat{\mathbf{H}}_{\text{tr}}), \hat{\mathbf{Z}} = \mathbf{E}_s(\mathbf{Z}), \tag{5}$$

where $\mathbf{E}_d(\cdot), \mathbf{E}_S(\cdot)$ are lightweight learnable perceptrons, $\mathbf{Z}, \hat{\mathbf{Z}} \in \mathbb{R}^{N_{t_{tr}} \times d'}$ are respectively denotes target and prediction representations for self-distillation. Then, we utilize target representations to generate student distributions. Specifically, we first utilize KMeans on $\mathbf{Z}$ to generate prototypes, denoted as $\mathbf{P} = \{\mathbf{p}_1, \cdots, \mathbf{p}_M\}$, where M is the number of prototypes. $M \ll N_{t_{\text{tr}}}$ guarantees most prototypes has evident representativeness for samples. The prototype assignment distribution is computed as:

$$r_{ij} = \frac{(1 + \|\mathbf{z}^{(i)} - \mathbf{p}_j\|^2)^{-1}}{\sum_{k=1}^{M}(1 + \|\mathbf{z}^{(i)} - \mathbf{p}_k\|^2)^{-1}}. \tag{6}$$

To generate more discriminative supervisory signals, we sharpen the assignment distribution:

$$p_{ij} = \frac{(r_{ij})^2}{\sum_{k=1}^{M}(r_{ik})^2}. \tag{7}$$

Next, we generate prediction distribution based on $\hat{\mathbf{Z}}$:

$$q_{ij} = \frac{(1 + ||\hat{\mathbf{z}}^{(i)} - \mathbf{p}_j||^2)^{-1}}{\sum_{k=1}^{M}(1 + ||\hat{\mathbf{z}}^{(i)} - \mathbf{p}_k||^2)^{-1}}. \tag{8}$$

Finally, the prediction branch is optimized by matching the sharpened target distribution through KL divergence:

$$\mathcal{L}_{sd} = \frac{1}{N_{t_{\text{tr}}}} \sum_{i=1}^{N_{t_{\text{tr}}}} \mathcal{D}_{\text{KL}}(\mathbf{p}_i \| \mathbf{q}_i). \tag{9}$$

where $\mathcal{D}_{\text{KL}}(\cdot||\cdot)$ is KL divergence. With KL loss $\mathcal{L}_{sd}$, clustering can generate prototypes that represent high-level abstract concepts, which guide the prediction distribution to capture more universal behavior patterns of representations.

Despite the effectiveness of cluster-based self-distillation, data across different datasets may still exhibit considerable distribution discrepancies due to variations in graph topology, semantic characteristics, and feature statistics. Such cross-dataset heterogeneity hinders the direct transferability of the trained models, limiting the cross-domain detection performance. To address this issue, we introduce a prototype generalization strategy that minimizes the distance between samples and their corresponding prototypes to enhance the cross-dataset transferability of knowledge. Compared to the traditional prompt learning-based generalization paradigm (Qiao et al., 2025; Niu et al., 2025; Chen et al., 2025), prototype-based methods avoid complex training pipelines and additionally injected learnable parameters.

Specifically, we first obtain the assigned prototypes $\mathbf{p}_{m^*}^{(i)}$ for sample $\mathbf{z}^{(i)}$, where $m^* = \arg\min_{m=1,\cdots,M} ||\mathbf{z}^{(i)} - \mathbf{p}_m||^2$. The loss function can be formulated as:

$$\mathcal{L}_p = \sum_{i=1}^{N_{t_{\text{tr}}}} w_i [||\mathbf{z}^{(i)} - \text{sg}(\mathbf{p}_{m^*}^{(i)})||^2 + ||\text{sg}(\mathbf{z}^{(i)}) - \mathbf{p}_{m^*}^{(i)}||^2]. \tag{10}$$

Here, we introduce adaptive weights $\mathbf{w} = [w_1, \cdots, w_{N_{t_{\text{tr}}}}]$ for samples. The intuition is that anomalous samples are often intermixed with normal ones during representation learning. Consequently, the resulting prototype distributions may be shifted toward abnormal patterns, compromising the prototype reliability and making cross-domain generalization more difficult. The adaptive weights are calculated as follows:

$$w_i = \frac{\cos^2(\mathbf{z}^{(i)}, \mathbf{p}_{m^*}^{(i)})}{\sum_{m=1}^{M} \cos(\mathbf{z}^{(i)}, \mathbf{p}_m^{(i)}) \sum_{j=1}^{N_{t_{\text{tr}}}} \cos(\mathbf{z}^{(j)}, \mathbf{p}_{m^*}^{(i)})}. \tag{11}$$

This score evaluates each sample's affinity toward its assigned prototype in two relative senses: relative to other prototypes, and relative to other samples within the same prototype. Based on adaptive weights, the loss function can dynamically adjust the contributions of samples to the generalization according to their relative affinity, alleviating the biased impact of anomalies.

4.4 Training and Inference

The total training objective is formulated as:

$$\mathcal{L} = \mathcal{L}_g + \alpha \mathcal{L}_c + \beta \mathcal{L}_{sd} + \kappa \mathcal{L}_p, \tag{12}$$

where α, β, κ are trade-off hyperparameters.

During inference, anomaly scores are calculated without retraining on the testing dataset. We integrate twofold information: (1) the distillation deviation, which represents the semantic discrepancy between the target distribution and the prediction, and (2) the prototypes deviation, which indicates the generalization of the sample towards the corresponding prototypes. The anomaly scores can be calculated as:

$$s_i = \mathcal{D}_{\text{KL}}(\mathbf{p}_i \| \mathbf{q}_i) + \lambda ||\mathbf{z}^{(i)} - \mathbf{p}_{m^*}^{(i)}||^2, \tag{13}$$

where λ is a hyperparameter. The complexity analysis and pseudo-code are in Section I.

Table 1: The performance of cross-granularity anomaly detection in zero-shot settings on single-graph datasets. RC indicates AUROC, and PR indicates AUPRC. N→E indicates training with the node-level task and testing on the edge-level task. E→N indicates training with the edge-level task and testing on the node-level task.

Task	Training	Reddit						Weibo					
	Test	Tolokers		Questions		T-Finance		Tolokers		Questions		T-Finance	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	UNPrompt	45.4±0.7	1.5±0.2	58.3±0.5	4.1±0.8	56.4±1.4	5.9±0.8	49.3±1.0	4.6±1.0	59.8±0.4	5.0±0.7	54.3±1.5	6.1±1.3
	AnomalyGFM	51.2±0.5	4.6±0.6	55.0±1.3	2.6±0.7	52.7±0.7	4.6±1.3	46.1±1.2	2.4±1.0	56.3±1.2	4.9±1.1	51.1±1.3	4.6±0.5
	Zero-GAD	50.6±1.5	3.4±0.2	58.9±0.4	4.9±1.5	57.4±1.2	5.8±1.3	50.3±0.8	5.5±0.6	60.7±0.6	7.8±1.4	55.6±1.2	7.0±0.8
	AffinityTune	47.5±0.6	2.9±0.3	53.2±0.7	2.8±1.3	49.6±0.9	1.0±0.1	44.0±1.1	1.1±1.4	54.2±0.9	4.2±1.3	47.8±1.4	3.5±0.9
	Ia-GGAD	46.4±1.5	2.2±0.6	57.9±0.5	4.6±0.9	55.6±0.9	4.2±0.5	48.7±1.3	3.9±0.6	58.7±1.0	6.3±0.8	53.4±0.7	5.9±1.3
	AutoGAD	42.7±1.2	1.3±0.8	61.5±0.8	7.3±0.8	63.1±0.8	10.1±0.5	53.4±0.8	7.6±0.5	64.5±0.5	10.4±0.9	59.0±0.8	8.9±1.0
	UniFORM	48.5±0.8	1.9±1.1	60.3±1.1	6.6±1.0	58.7±1.0	6.3±0.6	51.9±1.2	6.6±1.1	62.1±0.3	9.1±1.2	57.2±1.1	7.9±0.4
	GFM-UAD	51.6±1.0	4.3±0.9	57.0±0.6	3.6±1.0	54.2±1.4	4.4±1.2	47.3±1.3	3.1±0.6	57.8±1.0	5.8±0.9	52.2±0.8	5.0±1.1
	SGAD-DCR	49.7±0.7	2.3±1.4	52.5±1.5	1.3±0.2	45.1±0.8	0.8±0.2	36.4±0.3	0.3±0.0	44.9±1.0	0.8±0.1	49.6±0.8	0.9±0.1
	AD-GCL	41.5±1.0	0.8±0.1	55.3±0.9	2.6±1.3	48.7±0.6	0.4±0.2	44.9±1.2	1.2±0.1	52.1±0.9	1.9±0.3	37.2±0.6	0.5±0.2
	CgGFAD	**56.2**$_{\pm 0.4}$	**11.1**$_{\pm 0.5}$	**68.5**$_{\pm 0.6}$	**10.5**$_{\pm 0.5}$	**70.2**$_{\pm 0.7}$	**15.7**$_{\pm 0.8}$	**60.5**$_{\pm 0.6}$	**10.7**$_{\pm 0.4}$	**74.6**$_{\pm 0.8}$	**16.1**$_{\pm 0.9}$	**66.9**$_{\pm 0.3}$	**12.8**$_{\pm 0.4}$
E→N	UNPrompt	54.9±0.6	3.2±0.7	53.4±0.8	3.0±0.3	55.7±1.3	4.7±0.4	62.3±1.3	15.1±0.6	54.4±1.0	5.9±1.3	58.3±1.3	8.5±0.6
	AnomalyGFM	51.2±0.7	1.6±0.2	53.8±0.6	3.1±0.5	62.4±0.8	13.2±1.1	57.5±1.4	8.2±0.7	61.6±0.7	13.6±0.7	54.5±1.1	6.8±1.0
	Zero-GAD	55.5±0.8	4.7±0.5	44.3±0.8	0.5±0.1	56.7±0.6	5.6±0.5	63.7±1.2	16.0±0.8	65.9±1.0	18.8±1.3	60.3±1.1	9.8±1.2
	AffinityTune	48.3±1.3	0.7±0.1	47.8±0.7	0.8±0.2	60.5±0.7	12.7±0.8	54.6±0.4	3.9±1.0	59.6±0.9	10.6±0.6	51.8±1.3	6.0±1.5
	Ia-GGAD	54.6±0.9	4.8±0.3	52.3±1.0	1.2±0.4	65.1±0.6	14.1±1.0	61.2±0.8	14.7±1.1	64.1±0.8	16.6±1.3	58.0±0.6	7.9±1.3
	AutoGAD	57.7±1.1	5.5±0.2	57.0±0.6	4.3±1.3	59.1±0.5	6.6±0.5	66.6±0.8	17.5±0.5	68.9±1.2	22.9±1.4	62.4±1.4	11.3±1.6
	UniFORM	56.6±0.7	5.8±1.3	55.5±1.2	3.4±1.4	53.7±0.7	2.1±0.3	64.7±1.1	16.1±1.3	67.6±1.2	21.5±1.4	61.1±0.7	10.1±0.8
	GFM-UAD	53.2±0.9	2.2±0.6	51.4±1.3	1.9±0.2	63.7±1.1	13.8±0.6	59.4±0.9	10.0±0.8	62.7±0.5	17.2±1.4	56.2±1.0	7.2±1.2
	SGAD-DCR	47.7±1.3	0.5±0.1	47.0±0.9	0.7±0.4	53.1±0.3	2.6±0.7	49.6±1.2	1.9±0.3	52.9±1.2	3.9±0.4	44.4±1.3	1.3±0.2
	AD-GCL	46.6±0.7	0.8±0.2	35.5±1.1	0.2±0.0	57.7±0.6	4.1±0.8	54.7±0.9	2.9±0.4	57.6±0.6	8.5±1.4	48.1±0.6	1.8±0.3
	CgGFAD	**61.0**$_{\pm 0.6}$	**11.6**$_{\pm 0.5}$	**62.5**$_{\pm 0.7}$	**12.8**$_{\pm 0.4}$	**77.2**$_{\pm 0.8}$	**21.9**$_{\pm 0.4}$	**74.6**$_{\pm 0.8}$	**21.3**$_{\pm 0.7}$	**77.1**$_{\pm 0.2}$	**28.2**$_{\pm 0.6}$	**67.5**$_{\pm 0.7}$	**14.9**$_{\pm 0.8}$

Task	Training	Amazon						Yelp					
	Test	Tolokers		Questions		T-Finance		Tolokers		Questions		T-Finance	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	UNPrompt	54.5±1.4	8.6±0.7	57.8±0.7	7.6±1.3	49.9±0.2	0.8±0.1	51.3±1.3	2.4±0.3	62.3±1.5	9.8±0.8	62.8±0.7	9.8±0.7
	AnomalyGFM	51.0±1.0	6.6±1.2	53.0±0.8	5.4±1.4	46.0±0.2	0.5±0.1	47.5±0.3	0.8±0.1	59.0±0.6	8.2±0.6	58.9±1.5	8.2±1.4
	Zero-GAD	55.9±1.3	9.4±1.2	58.6±1.2	8.4±0.9	51.2±0.8	1.1±0.5	52.0±1.4	3.3±0.8	63.3±1.2	10.8±0.5	63.6±0.9	10.3±1.0
	AffinityTune	48.8±0.8	5.4±1.1	50.2±1.0	4.9±0.6	43.3±0.6	0.4±1.1	44.9±0.2	0.7±0.2	57.1±0.8	7.7±1.5	55.4±0.1	6.8±0.9
	Ia-GGAD	53.3±0.7	8.1±1.0	56.5±0.9	7.2±0.6	49.5±1.4	1.1±0.2	50.5±1.0	1.7±1.0	61.9±1.4	9.5±1.2	61.7±1.0	9.3±1.3
	AutoGAD	58.2±1.4	11.3±1.5	62.1±0.8	10.2±1.4	54.6±1.1	1.9±0.3	54.8±0.6	3.2±1.3	67.1±1.4	12.6±0.7	66.4±1.2	12.2±1.2
	UniFORM	56.8±0.6	10.6±1.0	59.9±0.6	9.3±1.2	53.2±0.8	1.1±0.2	53.1±0.7	3.4±1.2	64.6±1.5	11.7±0.5	64.9±0.6	11.2±0.7
	GFM-UAD	52.0±1.1	7.5±0.6	54.8±1.2	6.4±1.2	48.1±1.4	0.9±0.3	49.2±0.8	1.8±1.5	60.5±1.1	9.0±1.4	60.2±1.1	8.7±0.7
	SGAD-DCR	38.2±1.4	0.3±0.1	42.2±1.3	1.2±1.4	40.6±1.1	0.5±0.1	44.8±1.2	0.6±0.2	56.1±1.1	3.6±0.5	48.4±1.0	2.2±0.3
	AD-GCL	36.8±0.7	0.2±0.0	48.9±0.9	2.3±0.3	38.2±1.4	0.2±0.0	50.1±1.0	1.4±0.3	51.6±1.8	1.7±0.3	42.9±1.5	1.1±0.3
	CgGFAD	**66.4**$_{\pm 0.8}$	**15.8**$_{\pm 0.5}$	**71.7**$_{\pm 0.3}$	**14.2**$_{\pm 0.7}$	**61.9**$_{\pm 0.6}$	**9.1**$_{\pm 0.8}$	**63.3**$_{\pm 0.2}$	**11.1**$_{\pm 0.3}$	**76.5**$_{\pm 0.7}$	**18.0**$_{\pm 0.6}$	**75.2**$_{\pm 0.6}$	**17.6**$_{\pm 0.8}$
E→N	UNPrompt	61.8±1.4	13.1±0.9	59.6±1.4	4.0±0.7	57.5±1.3	5.3±1.0	51.3±1.0	1.0±0.2	60.8±1.0	11.2±1.3	56.5±1.1	7.5±1.4
	AnomalyGFM	58.1±0.6	7.0±0.9	55.7±1.2	3.0±0.7	53.5±1.0	3.4±1.3	47.4±1.5	0.8±0.1	56.7±0.7	7.7±0.5	52.0±0.5	6.2±0.9
	Zero-GAD	63.1±0.5	15.9±1.5	60.1±1.4	4.6±0.9	58.9±1.4	6.1±1.2	52.3±0.6	1.4±0.3	61.8±0.8	11.9±0.7	57.0±1.2	7.9±0.8
	AffinityTune	56.0±1.2	4.1±0.7	53.8±0.8	1.9±1.4	51.9±1.4	2.5±0.9	45.9±1.1	0.6±0.1	54.3±0.6	3.1±0.8	49.9±0.7	4.9±1.0
	Ia-GGAD	61.4±1.0	11.5±1.1	58.6±0.6	3.7±1.2	56.5±1.1	4.7±1.0	50.1±1.3	0.9±0.1	59.7±0.5	6.7±0.5	55.5±0.9	7.1±0.7
	AutoGAD	65.9±1.0	14.4±1.2	63.0±1.0	6.2±0.6	61.8±0.8	8.2±0.6	54.6±0.8	2.6±0.2	64.6±1.3	13.5±0.7	59.9±0.9	9.2±0.6
	UniFORM	64.1±0.8	12.7±0.8	61.2±1.4	5.5±0.9	59.8±1.0	7.1±1.0	53.3±0.9	1.2±0.3	62.7±1.4	11.5±1.0	58.1±0.8	8.6±1.1
	GFM-UAD	60.1±0.8	9.8±1.1	57.1±1.1	3.3±0.6	54.7±1.2	4.3±0.5	49.1±0.7	0.7±0.1	58.0±0.4	6.7±1.1	53.7±0.7	6.6±1.2
	SGAD-DCR	55.9±0.6	3.4±1.2	48.0±0.7	0.8±0.2	45.8±0.7	0.5±0.2	48.6±0.7	0.4±0.2	53.6±0.9	1.8±0.5	49.9±1.0	3.2±0.3
	AD-GCL	50.1±0.6	1.7±0.3	51.2±1.3	1.1±0.3	49.8±0.9	0.9±0.1	50.3±1.3	0.9±0.2	44.7±1.1	0.5±0.1	48.1±0.5	2.6±0.2
	CgGFAD	**74.1**$_{\pm 0.3}$	**20.6**$_{\pm 0.4}$	**67.8**$_{\pm 0.9}$	**9.8**$_{\pm 0.4}$	**66.3**$_{\pm 0.3}$	**11.2**$_{\pm 0.7}$	**58.2**$_{\pm 0.7}$	**5.1**$_{\pm 0.2}$	**72.5**$_{\pm 0.7}$	**20.1**$_{\pm 0.3}$	**65.8**$_{\pm 0.4}$	**11.8**$_{\pm 0.3}$

Table 2: The performance of cross-granularity anomaly detection in zero-shot settings on multi-graph datasets. RC indicates AUROC, and PR indicates AUPRC. N→G indicates training with the node-level task and testing on the graph-level task. G→N indicates training with the graph-level task and testing on the node-level task.

Task	Training	BM-MN						MUTAG					
	Test	BM-MS		BM-MT		T-Group		BM-MS		BM-MT		T-Group	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→G	UNPrompt	64.8±1.6	25.5±1.0	65.6±1.5	15.0±1.1	72.0±1.5	32.2±0.8	58.2±1.5	7.4±1.1	62.8±1.6	11.5±1.3	70.1±1.4	18.5±1.0
	AnomalyGFM	58.7±1.9	6.2±0.7	62.1±1.7	12.2±1.1	65.9±1.6	11.2±1.3	54.2±1.8	3.5±1.2	55.4±1.9	4.1±1.0	68.5±1.7	10.4±1.5
	Zero-GAD	67.6±2.0	26.8±1.8	57.9±1.8	7.5±0.5	74.0±1.6	35.2±1.5	57.8±1.9	5.2±1.4	59.2±2.0	9.4±1.5	70.6±1.8	17.3±1.6
	AffinityTune	59.8±1.8	7.3±0.4	53.2±1.9	4.4±0.4	73.1±1.5	32.8±1.4	55.0±1.7	4.2±1.1	54.1±1.8	6.8±0.9	73.4±1.6	22.8±1.4
	Ia-GGAD	63.1±1.7	19.0±1.0	61.4±1.5	8.4±1.2	69.6±1.4	17.2±2.1	58.5±1.6	8.8±1.1	64.8±1.7	15.5±1.3	72.0±1.5	21.8±1.7
	AutoGAD	61.9±1.6	15.2±0.7	67.8±1.7	16.2±1.7	66.4±1.3	14.0±1.4	60.5±1.5	10.8±0.8	61.4±1.6	12.3±1.4	74.2±1.4	24.5±1.3
	UniFORM	62.9±1.6	14.2±1.8	55.2±1.5	3.7±1.7	71.4±1.6	28.2±2.8	56.4±1.5	7.1±1.2	63.2±1.7	13.8±1.5	70.8±1.5	20.2±2.0
	GFM-UAD	66.1±1.3	27.0±1.0	68.3±1.2	19.2±0.9	75.7±1.1	36.2±1.7	55.8±1.4	6.5±1.1	60.5±1.3	11.8±1.0	77.8±1.2	28.4±1.5
	G-GLAD	49.2±1.5	5.4±0.8	51.0±1.6	6.2±0.5	60.5±1.4	11.5±1.3	45.1±1.4	1.2±0.2	46.8±1.5	3.8±0.3	55.2±1.3	7.4±1.2
	MoEGAD	51.5±1.4	6.8±0.9	53.2±1.5	7.5±1.1	55.4±1.3	6.6±1.5	47.3±1.3	2.1±0.7	48.9±1.4	4.8±0.8	57.8±1.2	9.0±1.1
	CgGFAD	**72.0**$_{\pm 0.7}$	**31.2**$_{\pm 1.1}$	**77.1**$_{\pm 0.6}$	**38.2**$_{\pm 0.8}$	**80.3**$_{\pm 0.6}$	**42.2**$_{\pm 1.1}$	**66.1**$_{\pm 0.7}$	**15.8**$_{\pm 0.5}$	**70.1**$_{\pm 0.6}$	**22.2**$_{\pm 0.6}$	**75.3**$_{\pm 0.6}$	**26.5**$_{\pm 1.1}$
G→N	UNPrompt	64.2±1.6	11.4±1.4	72.3±1.5	31.7±1.7	61.8±1.5	13.7±1.1	63.2±1.5	15.4±1.2	58.5±1.6	9.5±1.4	61.2±1.4	12.8±1.1
	AnomalyGFM	75.4±1.9	32.8±1.5	64.9±1.7	11.5±1.8	60.5±1.6	13.3±1.1	65.1±1.8	16.5±1.4	54.2±1.8	7.4±1.1	59.4±1.5	9.2±1.3
	Zero-GAD	66.0±2.0	17.3±1.2	70.0±1.8	28.4±1.4	54.2±1.6	2.9±0.4	58.3±1.9	10.8±1.3	60.5±1.9	12.1±1.4	53.2±1.7	5.5±1.5
	AffinityTune	64.9±1.8	12.4±2.1	69.4±1.9	14.7±1.1	52.6±1.5	2.1±0.2	57.2±1.7	9.2±1.3	52.0±1.8	5.4±0.9	54.1±1.6	5.9±1.1
	Ia-GGAD	68.5±1.7	21.2±1.0	65.7±1.5	16.3±1.5	58.2±1.4	4.0±0.3	66.0±1.6	17.5±1.1	56.4±1.7	8.4±1.0	47.5±1.5	1.2±0.2
	AutoGAD	55.6±1.6	10.8±1.8	61.2±1.7	9.9±1.1	48.9±1.3	0.9±0.1	57.4±1.5	9.5±1.2	59.8±1.6	10.4±1.3	52.0±1.4	2.5±1.5
	UniFORM	59.2±1.6	15.2±0.5	75.6±1.5	35.8±1.0	60.7±1.6	6.4±0.3	52.1±1.5	6.2±0.9	54.8±1.7	6.2±1.3	46.4±1.5	0.8±0.1
	GFM-UAD	61.5±1.3	14.9±1.1	59.0±1.2	10.1±1.3	51.8±1.1	1.1±0.1	64.8±1.4	16.2±1.2	64.4±1.3	18.9±1.1	54.5±1.2	5.4±1.0
	G-GLAD	49.2±0.6	2.4±0.8	51.0±1.6	3.2±1.0	45.1±1.4	0.8±0.1	48.0±1.6	2.8±0.5	40.5±1.4	1.1±0.2	42.1±1.3	0.5±0.0
	MoEGAD	51.5±1.4	4.8±0.9	53.2±1.5	5.5±1.1	39.3±1.3	0.3±0.0	46.2±1.5	1.5±0.6	46.8±1.3	2.9±0.3	44.6±1.2	0.7±0.1
	CgGFAD	**84.2**$_{\pm 0.1}$	**46.0**$_{\pm 0.9}$	**85.4**$_{\pm 0.6}$	**50.2**$_{\pm 1.1}$	**72.8**$_{\pm 0.6}$	**36.0**$_{\pm 1.4}$	**74.2**$_{\pm 0.7}$	**34.0**$_{\pm 0.9}$	**71.4**$_{\pm 0.9}$	**28.2**$_{\pm 1.1}$	**68.8**$_{\pm 0.6}$	**18.0**$_{\pm 1.4}$

Task	Training	MNIST0						MNIST1					
	Test	BM-MS		BM-MT		T-Group		BM-MS		BM-MT		T-Group	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→G	UNPrompt	60.5±1.6	10.3±1.2	58.2±1.4	7.8±1.1	60.4±1.5	9.2±1.2	60.8±1.5	9.5±1.1	53.2±1.4	2.8±0.4	47.9±1.5	0.5±0.1
	AnomalyGFM	52.1±1.7	3.4±0.9	53.8±1.8	3.9±1.0	55.2±1.6	4.7±1.2	59.4±1.9	8.8±1.3	54.8±1.7	3.6±1.1	52.0±1.8	1.4±1.0
	Zero-GAD	56.4±2.1	9.1±1.3	57.0±1.9	6.5±1.1	63.8±1.7	14.0±1.3	62.5±2.0	11.2±1.6	47.3±1.8	1.1±0.3	58.1±1.9	5.3±0.9
	AffinityTune	53.8±1.9	6.5±1.0	52.9±1.7	3.0±0.8	61.8±1.5	10.4±1.3	60.2±1.8	10.5±1.2	53.5±1.9	4.2±0.9	54.3±1.7	2.8±0.7
	Ia-GGAD	57.2±1.8	8.4±1.0	60.5±1.6	11.2±1.2	62.5±1.4	12.3±1.5	54.4±1.7	5.8±1.4	59.0±1.6	7.1±1.1	60.2±1.5	9.2±1.0
	AutoGAD	59.2±1.7	10.5±1.2	58.2±1.5	8.8±1.1	57.5±1.5	6.2±1.2	55.0±1.6	4.4±1.1	61.1±1.5	11.9±1.0	54.6±1.6	2.6±0.9
	UniFORM	54.8±1.7	7.0±1.0	58.1±1.6	6.5±1.3	53.2±1.6	4.1±1.7	52.5±1.8	3.9±0.5	57.4±1.7	6.9±1.2	58.2±1.6	6.1±1.1
	GFM-UAD	63.2±1.5	12.8±1.0	62.4±1.4	12.1±0.9	55.5±1.3	3.5±1.4	59.4±1.5	8.2±1.2	59.2±1.3	8.5±1.0	61.8±1.4	10.7±1.1
	G-GLAD	44.0±1.6	2.8±0.5	44.5±1.4	3.1±0.4	49.8±1.5	1.0±0.1	48.5±1.7	0.6±0.1	45.2±1.5	0.8±0.1	50.9±1.6	1.9±0.2
	MoEGAD	46.2±1.5	3.5±0.6	46.8±1.3	3.9±0.8	52.1±1.4	2.2±1.0	51.0±1.6	1.8±0.1	47.5±1.4	1.2±0.2	48.2±1.5	0.8±0.1
	CgGFAD	**65.0**$_{\pm 0.7}$	**14.6**$_{\pm 0.5}$	**66.1**$_{\pm 0.6}$	**16.7**$_{\pm 0.6}$	**70.5**$_{\pm 0.6}$	**23.6**$_{\pm 1.1}$	**71.8**$_{\pm 0.7}$	**24.1**$_{\pm 0.5}$	**65.1**$_{\pm 0.6}$	**16.9**$_{\pm 0.6}$	**66.9**$_{\pm 0.6}$	**14.4**$_{\pm 1.1}$
G→N	UNPrompt	61.5±1.4	11.8±1.1	65.2±1.6	16.4±1.5	55.4±1.5	5.1±0.9	57.2±1.6	6.4±1.0	59.4±1.5	7.6±1.1	53.0±1.4	4.2±0.8
	AnomalyGFM	65.0±1.7	16.8±1.5	55.3±1.8	8.1±1.2	50.1±1.6	3.5±1.1	60.8±1.7	10.1±1.2	51.0±1.9	4.8±0.9	54.8±1.8	5.6±1.0
	Zero-GAD	57.1±1.4	9.4±1.1	61.4±2.0	13.2±1.4	58.0±1.8	8.4±1.2	53.5±2.0	5.1±0.9	55.2±1.8	5.8±1.0	51.2±1.9	3.6±0.7
	AffinityTune	55.8±1.9	8.1±1.2	50.5±1.7	4.6±0.9	51.2±1.5	4.8±1.0	52.1±1.8	4.8±1.0	48.3±1.3	3.1±0.7	49.0±1.7	3.3±0.8
	Ia-GGAD	63.1±1.8	14.5±2.2	57.0±1.6	8.9±1.1	60.8±1.4	11.0±1.2	59.5±1.7	9.2±1.1	52.4±1.6	5.5±0.8	55.3±1.5	6.7±1.0
	AutoGAD	53.0±1.6	5.0±0.8	54.8±1.5	5.5±0.9	50.8±1.6	3.4±0.6						
	UniFORM	48.5±1.8	3.2±0.6	58.1±1.7	9.1±1.1	47.3±1.6	2.8±0.5						
	GFM-UAD	58.6±1.5	9.4±1.0	62.1±1.4	11.5±1.1	54.2±1.3	6.1±0.8						
	G-GLAD	41.2±1.6	2.0±0.5	37.5±1.5	1.1±0.2								
	MoEGAD	43.0±1.5	2.7±0.6	39.8±1.4	1.4±0.3								
	CgGFAD	**67.2**$_{\pm 0.7}$	**15.0**$_{\pm 0.2}$	**66.4**$_{\pm 0.8}$	**12.2**$_{\pm 1.2}$	**60.8**$_{\pm 0.6}$	**10.9**$_{\pm 0.4}$						

5 EXPERIMENTS

5.1 EXPERIMENTAL SETTINGS

Datasets. We introduce 14 datasets (Lin et al., 2024), including 7 single-graph datasets for node- and edge-level anomaly detection and 7 multi-graph datasets for node- and graph-level anomaly detection. The single-graph datasets contain Reddit, Weibo, Amazon, Yelp, Tolokers, Questions, and T-finance. The multi-graph datasets contain BM-MN, BM-MS, BM-MT, MUTAG, MNIST0, MNIST1, and T-Group. More details about the datasets are in Section B.

Baselines. For node- and edge-level detection on single-graph datasets, since there are no GFMs for cross-granularity GAD, we mainly introduce three types of state-of-the-art baselines, including GSCD, CGDS, and traditional GAD methods. (1) GSCD baselines: UNPrompt (Niu et al., 2025), AnomalyGFM (Qiao et al., 2025), Zero-GAD (Zheng et al., 2025), AffinityTune Chen et al. (2025), and Ia-GGAD (Zhang et al., 2026c). (2) GSCD baselines: AutoGAD (Yu et al., 2025), UniFORM (Song et al., 2025), and GFM-UAD (Han et al., 2026). (3) Traditional baselines: SGAD-DCR (Chen et al., 2026b) and AD-GCL (Xu et al., 2025). For node- and graph-level detection on multi-graph datasets, we introduce two types of baselines, including GSCD, CGDS, and traditional graph-level anomaly detection methods. GSCD and CGDS baselines follow the single-graph tasks, while traditional methods include G-GLAD (Zhang et al., 2026a) and MoEGAD (Cai et al., 2026). More illustrations are in Section C.

Implementations. The hidden dimension d' is set to 64. For unified experiments, α, β, κ, λ, and M are respectively set to 1, 1, 1, 0.8, and 30. The training epoch is set to 200. The learning rate is 1e-4.

Table 3: The ablation performance of cross-granularity anomaly detection in zero-shot settings on single-graph datasets.

Task	Training	Reddit						Weibo					
	Test	Tolokers		Questions		T-Finance		Tolokers		Questions		T-Finance	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	CgGFAD$_{\mathcal{L}_c}$	44.8±1.1	1.5±0.3	52.3±1.2	2.8±0.3	54.1±1.4	3.2±0.4	48.2±0.8	1.3±0.1	58.1±1.1	4.8±0.9	50.4±1.5	1.9±0.2
	CgGFAD$_{\mathcal{L}_g}$	41.2±1.4	1.2±0.6	49.8±0.7	1.2±0.4	51.6±1.7	2.4±0.5	45.1±1.6	1.0±0.2	55.3±1.5	2.9±0.5	48.9±0.7	1.0±0.1
	CgGFAD$_{\mathcal{L}_{sd}}$	46.5±1.3	1.6±0.4	54.1±1.4	3.3±0.6	56.3±0.6	4.7±0.8	39.8±1.2	0.8±0.2	60.4±1.4	8.5±1.0	53.2±1.5	2.9±0.4
	CgGFAD$_{\mathrm{KL}}$	40.1±1.2	0.9±0.3	48.5±0.9	1.1±0.3	50.9±1.6	2.0±0.4	42.4±0.5	1.1±0.1	57.2±1.3	6.4±0.9	50.1±1.4	1.4±0.8
	CgGFAD$_{\mathcal{L}_c}$	47.2±0.6	1.4±0.4	55.6±1.6	3.5±0.9	57.1±1.8	8.1±0.8	51.2±1.2	2.8±0.4	61.9±1.4	7.1±1.3	44.8±0.9	0.6±0.0
	CgGFAD$_{\mathrm{Rel}}$	38.9±1.2	0.7±0.2	46.9±1.5	0.9±0.1	49.4±0.7	1.2±0.2	47.0±1.4	1.8±0.5	56.8±1.3	4.1±1.1	50.7±1.1	2.1±0.3
	CgGFAD$_{\mathrm{ft}}$	53.1±1.0	7.8±0.6	58.4±1.1	4.1±0.7	58.2±1.3	8.6±0.7	53.3±1.2	4.4±0.9	69.5±1.0	9.5±0.8	61.8±0.6	7.5±0.5
	CgGFAD$_{\mathrm{fl}}$	51.4±1.5	3.2±0.7	61.2±1.3	5.6±0.8	63.1±1.9	11.3±0.9	52.1±1.3	3.9±0.8	67.2±1.6	9.8±1.2	59.5±1.1	5.9±0.2
	CgGFAD$_{\mathrm{w}}$	52.6±0.8	6.5±0.8	62.8±0.8	5.9±0.8	64.7±1.2	11.8±1.1	55.5±1.3	5.1±0.7	65.4±1.1	9.7±1.0	60.9±1.5	7.2±0.9
	CgGFAD	**56.2**±0.4	**11.1**±0.5	**68.5**±0.6	**10.5**±0.5	**70.2**±0.7	**15.7**±0.8	**60.5**±0.6	**10.7**±0.4	**74.6**±0.8	**16.1**±0.9	**66.9**±0.3	**12.8**±0.4
E→N	CgGFAD$_{\mathcal{L}_c}$	49.3±1.1	1.0±0.2	45.2±1.2	0.9±0.1	66.4±1.6	9.2±1.1	60.1±1.3	7.5±0.9	59.8±1.5	7.3±0.7	56.8±1.1	4.5±0.5
	CgGFAD$_{\mathcal{L}_g}$	50.1±1.3	2.1±0.5	52.0±1.4	3.5±0.8	54.1±1.8	3.1±0.8	61.8±1.5	8.6±1.0	58.4±1.6	5.2±0.5	54.5±1.7	2.7±0.4
	CgGFAD$_{\mathrm{Cos}}$	45.5±1.3	8.3±1.0	51.6±1.5	2.8±0.3	57.8±1.6	5.0±1.0	54.4±1.3	3.1±0.5	60.2±1.7	8.1±1.1	58.1±1.0	7.1±0.8
	CgGFAD$_{\mathrm{Bin}}$	48.2±1.4	1.5±0.3	50.1±1.3	2.1±0.5	65.2±1.7	8.6±1.1	55.9±1.4	4.0±0.6	62.6±1.5	11.8±1.5	55.6±1.5	5.1±0.3
	CgGFAD$_{\mathcal{L}_c}$	52.2±0.9	4.7±0.6	46.4±0.8	1.2±0.4	68.7±1.4	10.5±1.7	61.2±1.3	10.6±0.9	64.1±1.4	13.7±1.0	59.0±1.4	8.6±1.2
	CgGFAD$_{\mathrm{Rel}}$	57.0±1.4	7.7±1.1	46.9±1.7	1.9±0.3	56.0±0.8	5.0±0.6	52.7±1.4	2.3±0.4	62.4±0.9	10.1±1.0	56.4±1.5	5.4±0.7
	CgGFAD$_{\mathrm{ft}}$	58.9±0.9	9.4±0.9	54.2±1.0	3.5±0.7	74.6±1.1	14.8±1.6	65.9±1.0	14.2±0.8	69.5±0.7	17.8±0.8	65.1±1.0	10.5±1.1
	CgGFAD$_{\mathrm{fl}}$	55.5±1.1	8.8±0.7	54.9±1.1	3.7±0.6	72.8±1.3	13.6±0.8	69.1±1.1	16.0±0.9	72.8±0.9	20.2±1.9	63.2±1.2	10.6±0.9
	CgGFAD$_{\mathrm{w}}$	58.4±1.0	9.2±0.8	55.8±1.0	4.2±0.6	74.0±1.2	15.4±1.7	68.3±1.0	15.8±1.8	72.9±0.8	20.8±1.8	64.5±1.1	11.2±1.2
	CgGFAD	**61.0**±0.6	**11.6**±0.5	**62.5**±0.7	**12.8**±0.4	**77.2**±0.8	**21.9**±0.4	**74.6**±0.8	**21.3**±0.7	**77.1**±0.2	**28.2**±0.6	**67.5**±0.7	**14.9**±0.8

Task	Training	Amazon						Yelp					
	Test	Tolokers		Questions		T-Finance		Tolokers		Questions		T-Finance	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	CgGFAD$_{\mathcal{L}_c}$	51.2±0.8	2.3±0.7	55.4±1.3	6.2±1.0	47.5±1.6	0.8±0.1	48.7±1.3	1.5±0.3	59.8±1.0	6.1±1.0	58.7±1.4	5.8±0.7
	CgGFAD$_{\mathcal{L}_g}$	48.6±0.6	1.3±0.3	52.8±1.5	2.4±0.9	44.9±1.1	0.6±0.1	46.1±1.5	0.9±0.2	57.0±1.6	4.2±1.1	55.9±1.6	4.9±1.2
	CgGFAD$_{\mathcal{L}_{sd}}$	53.0±1.6	2.1±0.5	57.1±1.6	8.0±0.8	49.1±0.9	1.1±0.2	50.3±1.2	2.8±0.7	62.1±1.8	7.8±0.9	61.0±1.5	9.5±0.8
	CgGFAD$_{\mathrm{KL}}$	49.8±1.3	1.0±0.2	54.0±1.1	5.7±0.7	46.0±1.6	1.0±0.1	47.2±1.1	0.9±0.1	58.9±1.4	5.1±1.0	57.8±1.3	7.4±0.9
	CgGFAD$_{\mathcal{L}_c}$	54.3±1.1	4.6±0.7	58.6±1.8	7.4±0.8	50.5±1.7	2.7±0.4	51.8±1.2	2.1±0.5	63.6±1.3	7.4±0.9	62.4±1.5	10.1±0.8
	CgGFAD$_{\mathrm{Rel}}$	50.4±1.5	2.5±0.6	54.7±1.5	4.3±0.7	46.8±1.5	1.7±0.3	48.0±1.1	1.0±0.1	59.5±0.7	5.4±1.0	58.4±1.3	8.1±0.9
	CgGFAD$_{\mathrm{ft}}$	58.5±0.9	8.2±0.9	60.2±0.7	9.2±0.6	55.2±1.3	5.0±0.8	56.5±0.7	5.8±0.7	71.2±1.1	13.2±0.8	62.9±1.2	11.8±0.7
	CgGFAD$_{\mathrm{fl}}$	60.3±1.4	8.5±1.2	63.9±1.2	9.3±0.6	55.0±1.7	5.6±0.5	56.2±0.9	5.2±0.9	68.9±1.3	12.1±1.3	65.6±1.6	12.1±0.8
	CgGFAD$_{\mathrm{w}}$	61.8±1.7	9.9±0.8	65.1±1.4	9.7±0.9	56.1±1.8	5.8±1.0	53.4±1.5	4.5±0.3	70.1±0.8	12.9±1.2	66.8±1.1	12.5±1.4
	CgGFAD	**63.3**±0.2	**11.1**±0.3	**71.7**±0.3	**14.2**±0.7	**61.9**±0.6	**9.1**±0.8	**76.5**±0.7	**18.0**±0.6	**75.2**±0.6	**17.6**±0.8
E→N	CgGFAD$_{\mathcal{L}_c}$	59.2±1.2	8.8±1.0	55.4±1.4	3.1±0.8	55.9±1.7	4.0±0.3	49.5±1.3	1.1±0.2	59.4±1.3	7.5±1.0	55.3±1.5	4.2±0.8
	CgGFAD$_{\mathcal{L}_g}$	60.9±1.4	9.9±1.1	55.0±1.6	3.4±0.7	53.6±1.9	2.2±0.5	47.2±1.4	0.9±0.1	59.0±1.5	7.6±1.1	53.0±1.6	2.4±0.4
	CgGFAD$_{\mathrm{Cos}}$	64.6±1.1	12.4±1.0	54.8±1.3	2.6±0.5	51.2±1.6	1.6±0.2	50.8±1.2	1.5±0.3	55.9±1.3	4.1±0.7	56.7±1.4	5.8±0.6
	CgGFAD$_{\mathrm{Bin}}$	53.0±1.3	3.3±0.5	53.2±1.5	2.7±0.3	53.8±1.8	3.6±0.7	48.3±1.3	1.2±0.3	60.2±1.4	7.0±1.0	54.1±1.5	4.8±0.8
	CgGFAD$_{\mathcal{L}_c}$	60.5±1.2	10.9±1.0	54.6±0.8	3.0±0.5	52.1±1.6	2.0±0.6	51.5±1.2	2.8±0.5	58.8±1.3	5.6±0.9	57.5±1.4	6.3±0.5
	CgGFAD$_{\mathrm{Rel}}$	55.8±1.3	4.6±0.7	52.0±1.5	1.8±0.4	54.5±1.8	3.9±0.9	49.1±1.3	1.8±0.2	61.0±1.4	8.3±1.0	54.9±1.5	4.1±0.9
	CgGFAD$_{\mathrm{ft}}$	64.6±0.6	12.7±0.6	57.3±1.1	4.6±0.6	60.9±0.7	6.1±0.4	53.9±1.5	3.5±0.6	67.8±0.9	12.2±0.5	60.3±0.7	6.6±0.5
	CgGFAD$_{\mathrm{fl}}$	69.8±1.7	15.4±0.7	61.5±1.2	5.8±0.2	61.1±0.9	7.3±0.8	54.5±1.0	3.0±0.7	64.1±1.0	10.0±1.6	60.6±0.8	7.7±0.6
	CgGFAD$_{\mathrm{w}}$	67.0±1.6	13.3±1.6	60.7±0.9	5.3±0.3	60.3±0.8	6.8±0.5	54.4±0.9	3.3±0.6	65.2±0.9	10.8±1.5	62.8±0.7	8.3±0.5
	CgGFAD	**67.8**±0.9	**9.8**±0.4	**66.3**±0.3	**11.2**±0.7	**58.2**±0.7	**5.1**±0.2	**72.5**±0.7	**20.1**±0.3	**65.8**±0.4	**11.8**±0.3

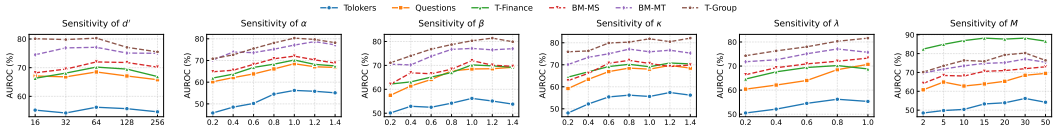

Figure 2: The sensitivity experiments. Solid line represents $N \rightarrow E$ task training on Reddit, and dashed line represents $N \rightarrow G$ task training on BM-MN.

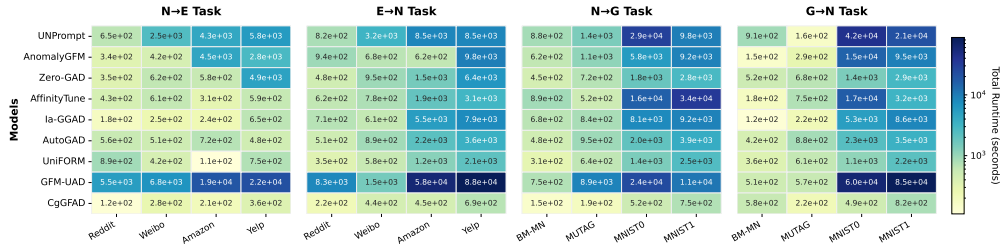

Figure 3: The efficiency experiment on training time.

All methods are executed in a Python 3.9.12 environment, utilizing a single NVIDIA A40 GPU, 40GB of RAM, and a 2.60GHz Xeon (R) Gold 6240 CPU. We utilize two metrics to evaluate the performance: AUROC and AUPRC. Each experiment is run 5 times with 5 different random seeds, and the average performance is recorded with standard deviation. More details are in Section D.

5.2 Experimental Analysis

Performance. Table 1 exhibits the cross-granularity performance on single-graph datasets. Table 2 shows the cross-granularity performance on multi-graph datasets. From their observations, CgGFAD achieves superior performance compared to all baselines on different cross-granularity detection tasks, demonstrating the advancement and effectiveness of CgGFAD. The advantages of CgGFAD lie in two aspects. First, CgGFAD discriminates the granularity-specific semantics, strengthening discrepancies between cross-granularity representations. Second, CgGFAD introduces cluster-based self-distillation and prototype generalization to capture discriminative normal patterns in an unsupervised manner and achieve cross-domain detection ability toward unseen testing datasets. For baselines, GSCD and CGDS outperform traditional methods, but they exhibit unstable performance because they fail to address two critical challenges simultaneously. More experimental results are in Section E.

Ablation. We implement ablation studies on critical components in CgGFAD. From the observations in Table 3, we can see that CgGFAD outperforms all variants significantly, demonstrating the utility of the components in CgGFAD. Specifically, the performance of CgGFAD$_{\mathcal{L}_c}$ and CgGFAD$_{\mathcal{L}_g}$ indicates that node-level alignment helps to maintain semantics for different granularities, and granularity discrimination is vital for identifying cross-granularity anomalies. CgGFAD$_{\mathcal{L}_{sd}}$ and CgGFAD$_{\mathrm{KL}}$ verify the importance of the self-distillation module in anomaly distillation and

anomaly inference. CgGFAD$_{\mathcal{L}_p}$ and CgGFAD$_{\text{Eud}}$ show the efficacy of the prototype generalization module in achieving zero-shot cross-domain anomaly inference. The performance of CgGFAD$_{\text{sg}}$ indicates the effectiveness of stop gradient strategies. CgGFAD$_{\delta}$ and CgGFAD$_{w}$ demonstrate the refinement influences of weights in loss functions. More details and experiments are in Section F.

Sensitivity. We implement sensitivity experiments about hyperparameters, including d', α, β, κ, λ, M. As shown in Fig. 2, changes in d' comprehensively maintain the stability of performance. However, α, β, κ, and λ control the training and inference process and influence the critical ability of CgGFAD to capture latent semantics, derive discriminative characteristics, and achieve generalization. Therefore, relatively low values lead to significant performance degradation. M influences the performance of clusters, and it can be seen that more cluster centers help to improve generalization ability. More sensitivity results are in Section G.

Efficiency and Scalability We first visualize the time cost of the training process to demonstrate the efficiency of CgGFAD. As shown in Fig. 3, we can see that CgGFAD has superior efficiency over most baselines, especially UNPrompt, AnomalyGFM, and GFM-UAD. This is because these baselines spend a significant amount of time on pre-training and tuning strategies. Then, we also exhibit the relationship between the number of edges and inference time to validate the scalability. As shown in Fig. 4, the relationship between the number of edges and inference time closely follows a linear trend rather than a quadratic or exponential growth, demonstrating the strong scalability of the proposed model. More experiments are shown in Section H.

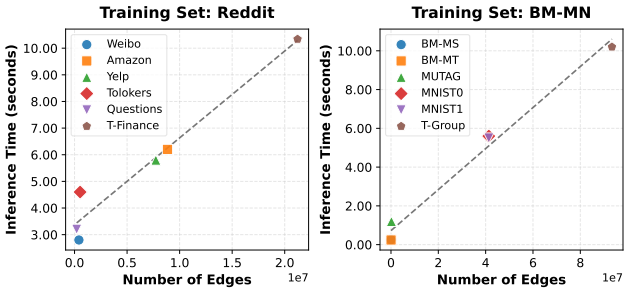

Figure 4: The efficiency and scalability experiment.

Granularity Discrepancy. To validate the effectiveness of the 2-Wasserstein distance, we introduce several variants, including cosine similarity, Euclidean distance, and KL divergence. As shown in Fig. 5, the observations indicate that using the 2-Wasserstein distance outperforms other distances. The reason is that vanilla metrics may fail to depict complex geometric semantic discrepancies between granularities. The variants exhibit unstable performance across datasets, demonstrating their limited ability to capture discriminative semantic patterns for different granularities.

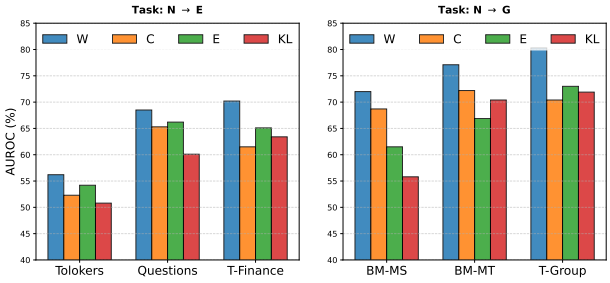

Figure 5: The granularity discrepancy experiment. For the $N \rightarrow E$ task, we leverage Reddit as the training dataset. For the $N \rightarrow G$ task, we leverage BM-MN as the training dataset.

6 Conclusion

This paper is the first attempt to extend GFMs for GAD into cross-granularity tasks and propose a novel framework, CgGFAD. CgGFAD tackles twofold challenges. On the one hand, CgGFAD introduces MMD based on the 2-Wasserstein distance to enhance the semantic discrimination across training and testing granularities. On the other hand, CgGFAD designs a cluster-based self-distillation module with a prototype generalization strategy, capturing comprehensive normal patterns and enhancing cross-domain detection. Extensive experiments on comprehensive public datasets exhibit the superiority of CgGFAD over state-of-the-art baselines. In the future, we will explore more semantic consistency from multi-view representations to boost this challenging task. The limitations are in Section J.

AI USE STATEMENT

During manuscript preparation, the authors used ChatGPT solely for grammatical correction and English-language polishing. The tool was not used for research design, methodological development, code generation, experimentation, data analysis, or the generation of scientific results or conclusions. The authors take full responsibility for the content, accuracy, and integrity of the manuscript.

ETHICS STATEMENT

This study does not involve human subjects, practices related to dataset releases, potentially harmful insights, methodologies and applications, potential conflicts of interest and sponsorship, discrimination/bias/fairness concerns, privacy and security issues, legal compliance, and research integrity issues (e.g., IRB, documentation, research ethics).

REPRODUCIBILITY STATEMENT

Sections 3 and 4 define the problem, the proposed framework, and the corresponding objective. The supplement specifies more details about experimental configuration and results. All evaluated datasets are public and cited; no private data are used.

REFERENCES

Jinyu Cai, Yunhe Zhang, Pengyang Wang, and See-Kiong Ng. Moegad: A mixture-of-experts framework with pseudo-anomaly generation for graph-level anomaly detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 48(4):4473–4489, 2026. doi: 10.1109/TPAMI.2025.3646069.

Jingyan Chen, Guanghui Zhu, Guansong Pang, Chunfeng Yuan, and Yihua Huang. Affinitytune: A prompt-tuning framework for few-shot anomaly detection on graphs. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 179–190, 2025. doi: 10.1145/3711896.3736836.

Qian Chen, Huiying Xu, Ruidong Wang, Yue Liu, and Xinzhong Zhu. Structure adversarial augmented graph anomaly detection via multi-view contrastive learning. *Knowl. Based Syst.*, 338: 115455, 2026a. doi: 10.1016/J.KNOSYS.2026.115455.

Yujia Chen, Jian Li, Linfei Sun, Tongcun Liu, and Guanjun Liu. Semi-supervised graph anomaly detection via dual-channel reconstruction. *Neurocomputing*, 668:132401, 2026b. doi: 10.1016/J.NEUCOM.2025.132401.

Renda Han, Xiaobao Wang, Luzhi Wang, Wenxin Zhang, Guangzhen Yao, and Hongxiang Liang. A graph foundation model for unified anomaly detection. In *Proceedings of the ACM Web Conference 2026*, pp. 523–534, 2026. doi: 10.1145/3774904.3792108.

Qian Huang, Hongyu Ren, Peng Chen, Gregor Krzmanc, Daniel Zeng, Percy Liang, and Jure Leskovec. PRODIGY: enabling in-context learning over graphs. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023*, 2023.

Yiqing Lin, Jianheng Tang, Chenyi Zi, H Vicky Zhao, Yuan Yao, and Jia Li. Unigad: Unifying multi-level graph anomaly detection. *Advances in neural information processing systems*, 37: 136120–136148, 2024.

Hao Liu, Jiarui Feng, Lecheng Kong, Ningyue Liang, Dacheng Tao, Yixin Chen, and Muhan Zhang. One for all: Towards training one graph model for all classification tasks. In *The Twelfth International Conference on Learning Representations*, 2024a.

Jie Liu, Mengting He, Xuequn Shang, Jieming Shi, Bin Cui, and Hongzhi Yin. Bourne: Bootstrapped self-supervised learning framework for unified graph anomaly detection. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pp. 2820–2833, 2024b.

Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, Philip S. Yu, and Shirui Pan. From few-shot to zero-shot: Towards generalist graph anomaly detection. *IEEE Trans. Knowl. Data Eng.*, 38(7):4357–4372, 2026a. doi: 10.1109/TKDE.2026.3691902.

Zhaowei Liu, Leilei Jiang, and Haitao Yang. Multi-dimensional adaptive mix-hop contextual learning framework for universal graph anomaly detection. In *Fortieth AAAI Conference on Artificial Intelligence*, pp. 15412–15420, 2026b. doi: 10.1609/AAAI.V40I18.38568.

Rami Mohawesh, Sumbal Maqsood, Haythem Bany Salameh, and Ghaleb A. El Refae. Leveraging graphsage and large language models with cross-attention for transaction fraud detection: Evidence from credit card and paysim datasets. *Inf. Process. Manag.*, 63(7):104860, 2026. doi: 10.1016/J.IPM.2026.104860.

Chaoxi Niu, Hezhe Qiao, Changlu Chen, Ling Chen, and Guansong Pang. Zero-shot generalist graph anomaly detection with unified neighborhood prompts. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3226–3234, 2025.

Hezhe Qiao, Chaoxi Niu, Ling Chen, and Guansong Pang. Anomalygfm: Graph foundation model for zero/few-shot anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2326–2337, 2025.

Haodan Ran, Yang Fang, Jiuyang Tang, Weiming Zhang, Jinzhi Liao, and Xiang Zhao. Dual-prompting based event anomaly detection in dynamic graphs. In *Database Systems for Advanced Applications*, pp. 20–36, 2025. doi: 10.1007/978-981-95-3906-2_2.

Junhua Shi, Qingyun Sun, Haonan Yuan, and Xingcheng Fu. Sa^{2}gfm: Enhancing robust graph foundation models with structure-aware semantic augmentation. In *Fortieth AAAI Conference on Artificial Intelligence*, pp. 15716–15724, 2026. doi: 10.1609/AAAI.V40I18.38602.

Chuancheng Song, Xixun Lin, Hanyang Shen, Yanmin Shang, and Yanan Cao. Uniform: Towards unified framework for anomaly detection on graphs. In *the Association for the Advancement of Artificial Intelligence*, pp. 12559–12567, 2025. URL https://doi.org/10.1609/aaai.v39i12.33369.

Maolin Wang, Beining Bao, Hongyu Chen, Zichun Liu, Lang Fu, Jun Chu, Langzhang Liang, and Zenglin Xu. GRAND: A robust diffusion framework for multi-granularity graph anomaly detection in web platforms. In *Proceedings of the ACM Web Conference 2026*, pp. 9572–9580, 2026. doi: 10.1145/3774904.3793012.

Chunyu Wei, Siyuan He, Yu Wang, Yueguo Chen, Yunhai Wang, Bing Bai, Yidong Zhang, Yong Xie, Shunming Zhang, and Fei Wang. Balanced anomaly-guided ego-graph diffusion model for inductive graph anomaly detection. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1542–1553, 2026. doi: 10.1145/3770854.3780240.

Chunjing Xiao, Yao Tong, Jiahui Lu, Yuxia Xue, Wenxin Tai, Zhangtao Cheng, and Fan Zhou. Graph-enhanced multi-scale contrastive learning for graph anomaly detection with adaptive diffusion models. *IEEE Trans. Big Data*, 11(6):2946–2957, 2025.

Yiming Xu, Zhen Peng, Bin Shi, Xu Hua, Bo Dong, Song Wang, and Chen Chen. Revisiting graph contrastive learning on anomaly detection: a structural imbalance perspective. In *the Association for the Advancement of Artificial Intelligence*, pp. 12972–12980, 2025. URL https://doi.org/10.1609/aaai.v39i12.33415.

Zhe-Rui Yang, Jindong Han, Chang-Dong Wang, and Hao Liu. Graphlora: Structure-aware contrastive low-rank adaptation for cross-graph transfer learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1785–1796, 2025. doi: 10.1145/3690624.3709186.

Jiaqi Yu, Yang Gao, Hong Yang, Zhihong Tian, Peng Zhang, and Xingquan Zhu. Automated graph anomaly detection with large language models. *Knowledge-based Systems*, 324:113809, 2025. URL https://doi.org/10.1016/j.knosys.2025.113809.

Haonan Yuan, Qingyun Sun, Jiacheng Tao, Xingcheng Fu, and Jianxin Li. RAG-GFM: overcoming in-memory bottlenecks in graph foundation models via retrieval-augmented generation. In *Proceedings of the ACM Web Conference 2026*, pp. 626–637, 2026. doi: 10.1145/3774904.3792139.

Chaoning Zhang, Kang Zhang, Chenshuang Zhang, Trung X. Pham, Chang D. Yoo, and In So Kweon. How does simsiam avoid collapse without negative samples? A unified understanding with self-supervised contrastive learning. In *The Tenth International Conference on Learning Representations*, 2022.

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 *Proceedings of the ACM Web Conference 2026*, pp. 1068–1079, 2026a. doi: 10.1145/3774904.3792429.

Muhan Zhang. Graph neural networks: link prediction. In *Graph Neural Networks: Foundations, Frontiers, and Applications*, pp. 195–223. Springer, 2022.

Wenxin Zhang and Cuicui Luo. GE-GNN: gated edge-augmented graph neural network for fraud detection. *IEEE Trans. Big Data*, 11(4):1664–1676, 2025a.

Wenxin Zhang and Cuicui Luo. Decomposition-based multi-scale transformer framework for time series anomaly detection. *Neural Networks*, 187:107399, 2025b. doi: 10.1016/J.NEUNET.2025.107399.

Wenxin Zhang and Cuicui Luo. Mclast: Multi-hierarchy contrastive learning graph anomaly detection with structure-awareness. *Neurocomputing*, 669:132480, 2026.

Wenxin Zhang, Yuhao Zhu, Cuicui Luo, Desheng Wu, Xi Xuan, Ljupco Kocarev, Renda Han, and Xiangxiang Lang. Feature and structural uncertainty modeling for risk identification in supply chain networks. *Chaos, Solitons & Fractals*, 201:117238, 2025.

Wenxin Zhang, Xi Xuan, Renda Han, Zonghao Ying, Cuicui Luo, Desheng Wu, and Ljupco Kocarev. Robust rumor detection against noise. *Neurocomputing*, 671:132741, 2026b. doi: 10.1016/J.NEUCOM.2026.132741.

Xiong Zhang, Zhenli He, Changlong Fu, and Cheng Xie. Ia-ggad: Zero-shot generalist graph anomaly detection via invariant and affinity learning. *Advances in Neural Information Processing Systems*, 38:83728–83760, 2026c.

Jialun Zheng, Jie Liu, Jiannong Cao, Xiao Wang, Hanchen Yang, and Yankai Chen. DP-DGAD: A generalist dynamic graph anomaly detector with dynamic prototypes. In *Proceedings of the ACM Web Conference 2026*, pp. 857–868, 2026. doi: 10.1145/3774904.3792268.

Xiangping Zheng, Xuan Feng, Bo Wu, Bin Ren, Wei Li, Xiuxin Hao, Xun Liang, Bin Tang, and Zhiwen Yu. Breaking semantic barriers: A zero-shot generalized framework for graph anomaly detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1423–1432, 2025. doi: 10.1145/3746027.3755173.

A Related Work

Some studies in general graph learning explore similar problems, and we manage to clarify the differences between them with our study. One-For-All (Liu et al., 2024a) explicitly unifies node-, link-, and graph-level tasks across domains. However, it focuses on the traditional classification problem, instead of graph anomaly detection. Although graph anomaly detection can also be regarded as a classification task, GAD tasks are more challenging because of the complexity of feature characteristics and data distribution and the rarity of anomalies. PRODIGY (Huang et al., 2023) explores a graph pretraining method, which utilizes large language models to achieve node- and edge-level representations. However, it lacks an elaborate design for different downstream tasks, and it also relies on labels. Moreover, both of them unify multi-granularity tasks in a supervised manner, instead of achieving cross-granularity GAD without reliance on labels. Cross-granularity GAD requires the model to capture transferable knowledge between different granularities, and they are unable to handle this task.

B DATASETS

The statistics of the datasets are shown in Table 4.

Table 4: Detailed statistics of the datasets used in our experiments.

Dataset	Train%	# Graphs	# Edges	# Nodes	# Dims	Nodes$_{ab}$	Edges$_{ab}$	Graphs$_{ab}$
Reddit	40%	1	168,016	10,984	64	3.33%	2.72%	/
Weibo	40%	1	416,368	8,405	400	10.33%	5.71%	/
Amazon	70%	1	8,847,096	11,944	25	6.87%	2.49%	/
Yelp	70%	1	7,739,912	45,954	32	14.53%	13.89%	/
Tolokers	50%	1	530,758	11,758	10	21.82%	33.44%	/
Questions	50%	1	202,461	48,921	301	2.98%	7.50%	/
T-Finance	40%	1	21,222,543	39,357	10	4.58%	2.77%	/
BM-MN	40%	700	40,032	12,911	1	48.91%	/	14.29%
BM-MS	40%	700	30,238	9,829	1	31.99%	/	14.29%
BM-MT	40%	700	32,042	10,147	1	34.49%	/	14.29%
MUTAG	40%	2,951	179,732	88,926	14	4.81%	/	34.40%
MNIST0	10%	70,000	41,334,380	4,939,668	5	35.46%	/	9.86%
MNIST1	10%	70,000	41,334,380	4,939,668	5	35.46%	/	11.25%
T-Group	40%	37,402	93,367,082	11,015,616	10	0.64%	/	4.26%

C BASELINES

We first state the imperative explanation for baseline fairness. On the one hand, some related methods, such as UniGAD (Lin et al., 2024) and DP-DGAD (Zheng et al., 2026), are supervised methods, and they are not applicable in this study because of the baseline fairness. On the other hand, since there are no active cross-granularity GAD studies, we mainly select three groups of baselines to maintain experimental fairness, including traditional GAD methods, graph foundation models for GAD, and multi-granularity GAD methods. Therefore, the number of baselines in each group may be limited, resulting in overlooking some related works.

Then, we introduce the details of the baselines:

- UNPrompt introduces a universal feature prompt on pretrained graph encoders and utilizes neighboring prediction to detect anomalies.
- AnomalyGFM introduces prototype-based prompts and one-class prompts to achieve generalized alignments.
- Zero-GAD leverages spectral reconstruction to detect anomalies.
- AffinityTune introduces different prompts for multi-level downstream learning tasks to capture the global affinity of nodes.
- Ia-GGAD develops an invariant feature discovery module integrated with affinity exploration to enhance cross-dataset generalization of GAD.
- BOURNE extracts a subgraph centered on each target node as node context and transforms it into a dual hypergraph as edge context, and enables the mutual detection of node and edge anomalies by swapping the context embeddings between nodes and edges and measuring the agreement in the embedding space.
- AutoGAD first designs a novel search space tailored for GAD tasks based on the characteristics of the graph autoencoder framework and then leverages a large language model as the controller of GNAS, guiding the LLM to rapidly generate architectures suitable for GAD within the search space through well-designed prompts.
- UniFORM unifies node-, edge-, and graph-level tasks from a subgraph perspective, leveraging an energy-based GNN for iterative multi-granular anomaly detection, and enhances meta-learning through contrastive learning and employs Langevin dynamics to generate phantom samples as substitutes for anomalous data, reducing reliance on labeled data.
- GFM-UAD utilizes adversarial learning to generate negative samples for pretraining and introduces KL divergence to achieve alignment.

Table 5: The performance of additional cross-granularity anomaly detection in zero-shot settings on single-graph datasets. RC indicates AUROC, and PR indicates AUPRC. N→E indicates training with the node-level task and testingtesting on the edge-level task. E→N indicates training with the edge-level task and test on the node-level task.

	Training	Tolokers								Questions								T-Finance							
Task	Test	Reddit		Weibo		Amazon		Yelp		Reddit		Weibo		Amazon		Yelp		Reddit		Weibo		Amazon		Yelp	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	UNPrompt	$53.1_{\pm1.1}$	$2.8_{\pm0.5}$	$48.2_{\pm0.9}$	$1.8_{\pm0.3}$	$64.5_{\pm1.2}$	$13.5_{\pm0.7}$	$58.0_{\pm1.0}$	$3.7_{\pm0.6}$	$61.2_{\pm1.3}$	$6.8_{\pm0.8}$	$47.5_{\pm1.1}$	$2.1_{\pm0.4}$	$67.2_{\pm0.9}$	$14.1_{\pm0.8}$	$49.4_{\pm1.2}$	$2.8_{\pm0.5}$	$64.0_{\pm1.4}$	$9.8_{\pm0.6}$	$59.3_{\pm1.0}$	$5.8_{\pm0.5}$	$57.1_{\pm1.2}$	$3.9_{\pm0.6}$	$48.2_{\pm1.1}$	$2.3_{\pm0.4}$
	AnomalyGFM	$50.4_{\pm1.3}$	$2.2_{\pm0.6}$	$45.1_{\pm1.1}$	$1.4_{\pm0.4}$	$61.8_{\pm1.4}$	$12.0_{\pm0.8}$	$54.2_{\pm1.2}$	$3.0_{\pm0.5}$	$57.6_{\pm1.5}$	$5.9_{\pm0.7}$	$44.0_{\pm1.3}$	$1.6_{\pm0.5}$	$63.8_{\pm1.1}$	$12.6_{\pm0.9}$	$46.8_{\pm1.4}$	$2.2_{\pm0.6}$	$61.2_{\pm1.6}$	$8.5_{\pm0.8}$	$56.5_{\pm1.2}$	$5.0_{\pm0.6}$	$53.8_{\pm1.5}$	$3.1_{\pm0.7}$	$45.1_{\pm1.3}$	$1.8_{\pm0.5}$
	Zero-GAD	$51.8_{\pm0.8}$	$2.4_{\pm0.2}$	$52.0_{\pm0.8}$	$2.2_{\pm0.3}$	$68.1_{\pm0.9}$	$15.2_{\pm0.5}$	$61.3_{\pm0.9}$	$4.5_{\pm0.4}$	$64.7_{\pm1.1}$	$7.8_{\pm0.5}$	$51.2_{\pm0.9}$	$2.6_{\pm0.3}$	$71.5_{\pm0.8}$	$16.0_{\pm0.6}$	$53.1_{\pm1.0}$	$3.4_{\pm0.4}$	$67.9_{\pm1.1}$	$11.8_{\pm0.5}$	$63.0_{\pm0.8}$	$6.9_{\pm0.4}$	$61.2_{\pm1.1}$	$4.8_{\pm0.5}$	$51.9_{\pm0.9}$	$2.9_{\pm0.3}$
	AffinityTune	$48.2_{\pm1.5}$	$1.9_{\pm0.7}$	$42.3_{\pm1.3}$	$1.1_{\pm0.5}$	$59.0_{\pm1.6}$	$10.8_{\pm0.9}$	$51.5_{\pm1.4}$	$2.5_{\pm0.6}$	$54.1_{\pm1.6}$	$4.8_{\pm0.9}$	$41.2_{\pm1.4}$	$1.3_{\pm0.6}$	$61.0_{\pm1.3}$	$11.2_{\pm1.0}$	$44.0_{\pm1.5}$	$1.8_{\pm0.7}$	$58.3_{\pm1.7}$	$7.2_{\pm0.9}$	$53.1_{\pm1.4}$	$4.1_{\pm0.7}$	$50.2_{\pm1.6}$	$2.5_{\pm0.8}$	$42.3_{\pm1.4}$	$1.5_{\pm0.5}$
	Ia-GGAD	$55.0_{\pm1.0}$	$3.1_{\pm0.5}$	$50.1_{\pm1.0}$	$2.0_{\pm0.4}$	$66.8_{\pm1.1}$	$14.3_{\pm0.6}$	$59.4_{\pm1.1}$	$4.1_{\pm0.5}$	$62.8_{\pm1.2}$	$7.3_{\pm0.6}$	$49.4_{\pm1.0}$	$2.3_{\pm0.4}$	$69.2_{\pm0.9}$	$15.1_{\pm0.7}$	$51.2_{\pm1.1}$	$3.0_{\pm0.5}$	$65.8_{\pm1.3}$	$10.6_{\pm0.6}$	$61.2_{\pm1.0}$	$6.3_{\pm0.5}$	$59.0_{\pm1.2}$	$4.4_{\pm0.6}$	$49.8_{\pm1.0}$	$2.6_{\pm0.4}$
	AutoGAD	$58.1_{\pm0.7}$	$3.7_{\pm0.3}$	$53.5_{\pm0.7}$	$2.4_{\pm0.3}$	$63.4_{\pm0.8}$	$10.1_{\pm0.4}$	$62.8_{\pm0.8}$	$4.9_{\pm0.3}$	$66.1_{\pm0.9}$	$8.2_{\pm0.4}$	$52.8_{\pm0.7}$	$2.8_{\pm0.3}$	$58.9_{\pm0.6}$	$6.0_{\pm0.5}$	$54.8_{\pm0.8}$	$3.7_{\pm0.3}$	$69.6_{\pm0.9}$	$12.6_{\pm0.4}$	$60.5_{\pm0.7}$	$5.3_{\pm0.4}$	$62.6_{\pm0.9}$	$5.1_{\pm0.4}$	$53.4_{\pm0.7}$	$3.2_{\pm0.3}$
	UniFORM	$52.3_{\pm0.9}$	$2.6_{\pm0.5}$	$47.0_{\pm0.8}$	$1.6_{\pm0.3}$	$63.2_{\pm1.0}$	$12.8_{\pm0.6}$	$56.8_{\pm0.9}$	$3.4_{\pm0.5}$	$59.4_{\pm1.1}$	$6.3_{\pm0.6}$	$46.1_{\pm0.8}$	$1.9_{\pm0.4}$	$65.8_{\pm0.8}$	$13.3_{\pm0.7}$	$48.0_{\pm1.0}$	$2.5_{\pm0.5}$	$62.9_{\pm1.2}$	$9.0_{\pm0.6}$	$58.0_{\pm0.9}$	$5.4_{\pm0.5}$	$55.6_{\pm1.1}$	$3.6_{\pm0.5}$	$46.8_{\pm0.9}$	$2.0_{\pm0.4}$
	GFM-UAD	$49.8_{\pm1.2}$	$2.1_{\pm0.6}$	$44.2_{\pm1.1}$	$1.3_{\pm0.4}$	$60.5_{\pm1.3}$	$11.2_{\pm0.8}$	$53.1_{\pm1.3}$	$2.8_{\pm0.5}$	$56.2_{\pm1.4}$	$5.2_{\pm0.8}$	$42.8_{\pm1.2}$	$1.5_{\pm0.5}$	$62.4_{\pm1.0}$	$12.0_{\pm0.8}$	$45.2_{\pm1.3}$	$2.0_{\pm0.6}$	$59.8_{\pm1.5}$	$8.0_{\pm0.8}$	$55.0_{\pm1.1}$	$4.5_{\pm0.6}$	$52.1_{\pm1.3}$	$2.8_{\pm0.7}$	$43.9_{\pm1.2}$	$1.6_{\pm0.5}$
	SGAD-DCR	$47.5_{\pm0.8}$	$1.6_{\pm0.4}$	$48.8_{\pm0.8}$	$1.3_{\pm0.3}$	$60.8_{\pm0.8}$	$10.7_{\pm0.5}$	$55.0_{\pm0.8}$	$3.7_{\pm0.4}$	$60.3_{\pm1.0}$	$5.0_{\pm0.5}$	$47.0_{\pm0.8}$	$1.2_{\pm0.3}$	$65.1_{\pm0.7}$	$11.6_{\pm0.6}$	$52.0_{\pm0.5}$	$2.5_{\pm0.4}$	$61.8_{\pm1.0}$	$8.2_{\pm0.5}$	$60.8_{\pm0.8}$	$6.1_{\pm0.4}$	$56.9_{\pm0.9}$	$3.9_{\pm0.5}$	$50.7_{\pm0.3}$	$2.1_{\pm0.3}$
	AD-GCL	$51.2_{\pm1.0}$	$2.4_{\pm0.5}$	$45.5_{\pm0.9}$	$1.5_{\pm0.4}$	$62.0_{\pm1.1}$	$12.0_{\pm0.7}$	$55.3_{\pm1.0}$	$3.2_{\pm0.5}$	$58.1_{\pm1.2}$	$5.8_{\pm0.7}$	$44.9_{\pm0.9}$	$1.7_{\pm0.4}$	$64.2_{\pm0.9}$	$12.7_{\pm0.8}$	$46.8_{\pm1.1}$	$2.3_{\pm0.5}$	$61.2_{\pm1.3}$	$8.5_{\pm0.7}$	$56.3_{\pm1.0}$	$4.8_{\pm0.5}$	$53.8_{\pm1.2}$	$3.2_{\pm0.6}$	$45.1_{\pm1.0}$	$1.8_{\pm0.4}$
	CgGFAD	$\mathbf{63.2_{\pm0.5}}$	$\mathbf{5.0_{\pm0.4}}$	$\mathbf{58.4_{\pm0.6}}$	$\mathbf{3.5_{\pm0.5}}$	$\mathbf{74.1_{\pm0.7}}$	$\mathbf{19.2_{\pm0.3}}$	$\mathbf{67.7_{\pm0.6}}$	$\mathbf{6.7_{\pm0.4}}$	$\mathbf{71.0_{\pm0.5}}$	$\mathbf{10.3_{\pm0.5}}$	$\mathbf{58.5_{\pm0.6}}$	$\mathbf{4.0_{\pm0.4}}$	$\mathbf{78.0_{\pm0.5}}$	$\mathbf{20.4_{\pm0.5}}$	$\mathbf{60.8_{\pm0.6}}$	$\mathbf{5.2_{\pm0.4}}$	$\mathbf{75.6_{\pm0.7}}$	$\mathbf{15.8_{\pm0.3}}$	$\mathbf{70.3_{\pm0.5}}$	$\mathbf{9.8_{\pm0.4}}$	$\mathbf{68.5_{\pm0.7}}$	$\mathbf{7.0_{\pm0.5}}$	$\mathbf{59.2_{\pm0.6}}$	$\mathbf{4.7_{\pm0.4}}$
E→N	UNPrompt	$68.5_{\pm1.2}$	$15.2_{\pm0.7}$	$54.8_{\pm1.0}$	$6.5_{\pm0.5}$	$66.1_{\pm1.3}$	$13.0_{\pm0.8}$	$52.1_{\pm1.1}$	$3.8_{\pm0.6}$	$50.3_{\pm1.2}$	$2.6_{\pm0.5}$	$56.0_{\pm1.0}$	$3.9_{\pm0.6}$	$73.8_{\pm0.9}$	$24.8_{\pm0.8}$	$69.0_{\pm1.2}$	$13.5_{\pm0.7}$	$48.5_{\pm1.3}$	$1.8_{\pm0.5}$	$53.2_{\pm1.1}$	$4.1_{\pm0.6}$	$63.5_{\pm1.0}$	$9.8_{\pm0.7}$	$45.0_{\pm1.2}$	$1.2_{\pm0.5}$
	AnomalyGFM	$66.2_{\pm1.0}$	$13.8_{\pm0.8}$	$52.0_{\pm1.2}$	$5.1_{\pm0.6}$	$63.8_{\pm1.1}$	$11.5_{\pm0.9}$	$50.0_{\pm1.3}$	$2.9_{\pm0.5}$	$48.0_{\pm1.0}$	$2.0_{\pm0.4}$	$53.5_{\pm1.3}$	$3.1_{\pm0.5}$	$71.5_{\pm1.1}$	$22.0_{\pm0.9}$	$67.2_{\pm1.0}$	$12.0_{\pm0.8}$	$46.0_{\pm1.1}$	$1.4_{\pm0.4}$	$51.0_{\pm0.9}$	$3.2_{\pm0.5}$	$61.0_{\pm1.2}$	$8.2_{\pm0.6}$	$43.2_{\pm1.0}$	$1.0_{\pm0.4}$
	Zero-GAD	$69.8_{\pm0.9}$	$16.0_{\pm0.6}$	$53.2_{\pm0.9}$	$5.8_{\pm0.4}$	$67.5_{\pm0.8}$	$14.2_{\pm0.6}$	$53.5_{\pm0.8}$	$4.5_{\pm0.4}$	$51.8_{\pm0.9}$	$3.1_{\pm0.4}$	$57.2_{\pm0.8}$	$4.5_{\pm0.4}$	$74.5_{\pm0.8}$	$25.5_{\pm0.7}$	$70.5_{\pm0.8}$	$14.8_{\pm0.6}$	$49.8_{\pm0.9}$	$2.1_{\pm0.4}$	$54.8_{\pm0.8}$	$4.9_{\pm0.4}$	$64.8_{\pm0.9}$	$10.6_{\pm0.5}$	$46.5_{\pm0.9}$	$1.4_{\pm0.3}$
	AffinityTune	$64.0_{\pm1.4}$	$12.1_{\pm0.9}$	$51.0_{\pm1.1}$	$4.6_{\pm0.5}$	$61.2_{\pm1.5}$	$10.0_{\pm1.0}$	$48.2_{\pm1.2}$	$2.2_{\pm0.6}$	$46.2_{\pm1.3}$	$1.6_{\pm0.5}$	$51.0_{\pm1.2}$	$2.4_{\pm0.6}$	$69.0_{\pm1.3}$	$19.8_{\pm1.0}$	$65.0_{\pm1.4}$	$10.5_{\pm0.9}$	$44.1_{\pm1.2}$	$1.1_{\pm0.5}$	$49.0_{\pm1.2}$	$2.5_{\pm0.6}$	$58.5_{\pm1.3}$	$6.9_{\pm0.8}$	$41.0_{\pm1.3}$	$0.8_{\pm0.4}$
	Ia-GGAD	$67.2_{\pm1.1}$	$14.5_{\pm0.7}$	$55.3_{\pm0.8}$	$6.9_{\pm0.4}$	$64.9_{\pm1.0}$	$12.1_{\pm0.7}$	$51.0_{\pm1.0}$	$3.2_{\pm0.5}$	$49.2_{\pm1.1}$	$2.3_{\pm0.5}$	$54.8_{\pm0.9}$	$3.5_{\pm0.5}$	$72.2_{\pm1.0}$	$23.2_{\pm0.8}$	$68.1_{\pm0.9}$	$12.8_{\pm0.7}$	$47.2_{\pm1.0}$	$1.6_{\pm0.4}$	$52.1_{\pm1.0}$	$3.6_{\pm0.5}$	$62.2_{\pm1.1}$	$8.9_{\pm0.6}$	$44.1_{\pm1.1}$	$1.1_{\pm0.4}$
	AutoGAD	$70.5_{\pm0.8}$	$16.8_{\pm0.5}$	$56.0_{\pm0.7}$	$7.3_{\pm0.4}$	$68.1_{\pm0.9}$	$14.8_{\pm0.5}$	$54.2_{\pm0.9}$	$4.9_{\pm0.4}$	$52.5_{\pm0.8}$	$3.3_{\pm0.4}$	$58.1_{\pm0.7}$	$4.8_{\pm0.4}$	$75.2_{\pm0.7}$	$26.1_{\pm0.6}$	$71.8_{\pm0.7}$	$15.6_{\pm0.5}$	$51.0_{\pm0.8}$	$2.3_{\pm0.3}$	$55.6_{\pm0.7}$	$5.3_{\pm0.4}$	$65.6_{\pm0.8}$	$11.2_{\pm0.5}$	$47.2_{\pm0.8}$	$1.5_{\pm0.3}$
	UniFORM	$65.5_{\pm1.1}$	$13.0_{\pm0.8}$	$49.8_{\pm1.0}$	$4.0_{\pm0.5}$	$62.5_{\pm1.2}$	$10.8_{\pm0.8}$	$49.1_{\pm1.1}$	$2.5_{\pm0.5}$	$47.1_{\pm1.2}$	$1.8_{\pm0.4}$	$52.2_{\pm1.1}$	$2.8_{\pm0.5}$	$70.2_{\pm1.0}$	$21.0_{\pm0.9}$	$66.0_{\pm1.1}$	$11.2_{\pm0.8}$	$45.0_{\pm1.1}$	$1.2_{\pm0.4}$	$50.1_{\pm1.0}$	$2.8_{\pm0.5}$	$59.8_{\pm1.1}$	$7.5_{\pm0.7}$	$42.0_{\pm1.1}$	$0.9_{\pm0.4}$
	GFM-UAD	$63.2_{\pm1.3}$	$11.5_{\pm0.9}$	$48.5_{\pm1.2}$	$3.4_{\pm0.6}$	$60.1_{\pm1.4}$	$9.2_{\pm0.9}$	$47.0_{\pm1.3}$	$1.9_{\pm0.6}$	$45.0_{\pm1.4}$	$1.4_{\pm0.5}$	$49.8_{\pm1.3}$	$2.0_{\pm0.6}$	$68.0_{\pm1.2}$	$18.5_{\pm1.1}$	$63.8_{\pm1.3}$	$9.8_{\pm0.9}$	$43.0_{\pm1.3}$	$0.9_{\pm0.5}$	$48.0_{\pm1.2}$	$2.1_{\pm0.6}$	$57.2_{\pm1.4}$	$6.1_{\pm0.8}$	$39.8_{\pm1.2}$	$0.7_{\pm0.4}$
	SGAD-DCR	$58.1_{\pm1.0}$	$8.2_{\pm0.7}$	$42.5_{\pm1.1}$	$1.8_{\pm0.4}$	$54.2_{\pm1.2}$	$6.0_{\pm0.8}$	$41.0_{\pm1.0}$	$0.9_{\pm0.3}$	$39.5_{\pm1.1}$	$0.6_{\pm0.3}$	$43.0_{\pm1.0}$	$0.8_{\pm0.3}$	$62.0_{\pm1.1}$	$14.2_{\pm0.9}$	$58.1_{\pm1.0}$	$6.1_{\pm0.6}$	$37.2_{\pm1.1}$	$0.4_{\pm0.2}$	$42.0_{\pm0.9}$	$0.9_{\pm0.3}$	$51.0_{\pm1.0}$	$3.2_{\pm0.5}$	$35.0_{\pm1.0}$	$0.3_{\pm0.2}$
	AD-GCL	$56.5_{\pm1.2}$	$7.5_{\pm0.8}$	$41.0_{\pm1.1}$	$1.2_{\pm0.4}$	$52.8_{\pm1.3}$	$5.2_{\pm0.9}$	$39.8_{\pm1.2}$	$0.6_{\pm0.3}$	$38.0_{\pm1.2}$	$0.4_{\pm0.2}$	$41.5_{\pm1.1}$	$0.6_{\pm0.3}$	$60.5_{\pm1.2}$	$13.0_{\pm0.6}$	$56.5_{\pm1.2}$	$5.2_{\pm0.7}$	$35.8_{\pm1.2}$	$0.3_{\pm0.2}$	$40.5_{\pm1.1}$	$0.6_{\pm0.3}$	$49.2_{\pm1.2}$	$2.5_{\pm0.6}$	$33.5_{\pm1.1}$	$0.2_{\pm0.1}$
	CgGFAD	$\mathbf{75.0_{\pm0.6}}$	$\mathbf{20.5_{\pm0.5}}$	$\mathbf{60.5_{\pm0.7}}$	$\mathbf{10.8_{\pm0.4}}$	$\mathbf{73.2_{\pm0.8}}$	$\mathbf{18.5_{\pm0.4}}$	$\mathbf{59.8_{\pm0.6}}$	$\mathbf{6.8_{\pm0.5}}$	$\mathbf{57.0_{\pm0.6}}$	$\mathbf{4.5_{\pm0.5}}$	$\mathbf{62.5_{\pm0.7}}$	$\mathbf{6.2_{\pm0.5}}$	$\mathbf{80.0_{\pm0.5}}$	$\mathbf{30.5_{\pm0.6}}$	$\mathbf{76.8_{\pm0.7}}$	$\mathbf{18.5_{\pm0.5}}$	$\mathbf{55.8_{\pm0.8}}$	$\mathbf{3.0_{\pm0.4}}$	$\mathbf{60.2_{\pm0.6}}$	$\mathbf{7.0_{\pm0.5}}$	$\mathbf{70.5_{\pm0.7}}$	$\mathbf{14.2_{\pm0.5}}$	$\mathbf{52.2_{\pm0.7}}$	$\mathbf{2.0_{\pm0.5}}$

- SGAD-DCR employs a dual-filter attention encoder to construct dual aggregation channels for low-frequency and high-frequency signals and mitigates the homophily trap via semantic-level alignment.
- AD-GCL devises the neighbor pruning strategy to filter noisy edges for head nodes and facilitate the detection of genuine tail nodes by aligning from head nodes to forged tail nodes. Moreover, AD-GCL actively explores potential neighbors to enlarge the receptive field of tail nodes through anomaly-guided neighbor completion.
- G-GLAD introduces prompt-based anomaly space expansion to learn a richer and more generalizable anomaly space; and leverages subgraph-based normality extraction to isolate essential substructures for normality prediction while filtering out spurious motifs.
- MoEGAD introduces an iterative anomalous graph generation module to produce pseudo-anomalous graphs, and proposes a latent MoE to promote cross-task adaptability for diverse GLAD problems.

D Implementation

Considering the limited studies for edge-level detection, we integrate a concatenated strategy (Zhang, 2022) based on link prediction into the above baselines, and we generate abnormal edges based on the probability of nodes $p(e_{ij}) = \mathrm{Avg}(p_i, p_j)$ (Lin et al., 2024). For edge-level anomaly detection, we ignore edges between the training nodes and the testing nodes to prevent data leakage; for graph-level anomaly detection,

E Performance

Table 5 shows the performance on Reddit, Weibo, Amazon, and Yelp datasets using Tolokers, Questions, and T-Finance as training datasets. Table 6 shows the performance on BM-MN, MUTAG, MNIST0, and MNIST1 datasets using BM-MS, BM-MT, and T-Group as training datasets. From these observations, we can see that CgGFAD achieves the best performance with respect to AUROC and AUPRC compared with comprehensive state-of-the-art baselines, including GSCD, CGDS, and traditional GAD baselines, which further demonstrates the effectiveness and advancement of CgGFAD.

F Ablation

Table 7 exhibits the ablation performance on Reddit, Weibo, Amazon, and Yelp datasets using Tolokers, Questions, and T-Finance as training datasets. Table 8 shows the ablation performance on BM-MN, MUTAG, MNIST0, and MNIST1 datasets using BM-MS, BM-MT, and T-Group as training datasets. From the observations, we can see that CgGFAD achieves the best performance

Table 6: The performance of cross-granularity anomaly detection in zero-shot settings on multi-graph datasets. RC indicates AUROC, and PR indicates AUPRC. N→G indicates training with the node-level task and test on the graph-level task. G→N indicates training with the edge-level task and test on the graph-level task.

	Training	BM-MS								BM-MT								T-Group							
Task	Test	BM-MN		MUTAG		MNIST0		MNIST1		BM-MN		MUTAG		MNIST0		MNIST1		BM-MN		MUTAG		MNIST0		MNIST1	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→G	UNPrompt	77.2±1.4	51.1±0.9	74.5±1.2	38.2±1.0	83.1±1.1	54.2±0.8	78.5±1.3	38.0±1.1	76.2±1.0	32.5±0.8	68.4±1.2	22.1±0.7	63.2±1.1	15.8±0.6	74.0±1.3	24.1±0.5	62.5±1.2	18.2±0.8	62.1±1.0	10.5±0.6	63.8±0.8	16.5±0.6	70.2±0.9	22.1±1.0
	AnomalyGFM	71.5±1.5	44.2±1.1	78.0±1.1	41.5±0.9	76.8±1.4	48.0±1.2	84.2±0.9	45.1±0.8	71.0±1.2	28.4±0.9	71.5±1.0	25.8±0.6	67.5±1.0	18.2±0.7	69.1±1.2	21.0±0.6	67.8±1.1	21.5±0.7	57.2±1.1	9.2±0.5	66.5±0.9	18.2±0.5	66.4±1.1	19.5±0.8
	Zero-GAD	74.8±1.0	47.5±0.8	80.2±0.9	43.8±0.8	81.5±0.8	52.0±0.7	81.4±1.0	42.5±0.9	77.5±0.8	35.3±0.6	73.8±0.8	27.2±0.5	61.5±0.9	14.1±0.5	71.8±0.9	23.2±0.4	64.4±1.0	19.5±0.6	62.5±0.8	11.8±0.5	61.2±1.0	14.7±0.5	72.1±0.7	24.8±0.6
	AffinityTune	68.0±1.3	39.1±1.2	72.1±1.4	35.0±1.1	79.2±1.2	49.5±1.0	86.0±0.8	46.8±0.7	68.5±1.1	26.5±0.8	65.2±1.3	18.9±0.8	65.8±1.2	16.9±0.6	66.2±1.1	18.5±0.5	69.1±0.9	22.8±0.8	54.0±1.2	7.5±0.6	64.1±1.1	16.0±0.7	63.5±1.1	16.2±0.7
	Ia-GGAD	76.0±0.9	48.8±0.7	76.5±1.0	39.2±0.8	82.0±0.7	53.1±0.9	82.5±1.1	43.2±0.8	74.2±0.9	31.8±0.7	70.8±0.7	24.5±0.6	64.1±0.8	15.2±0.5	73.1±0.8	24.8±0.4	65.8±0.8	20.1±0.5	61.0±0.9	11.0±0.4	65.2±0.8	17.5±0.4	69.0±0.8	21.5±0.6
	AutoGAD	73.2±0.8	45.0±0.6	79.1±0.7	42.1±0.6	80.5±0.7	51.2±0.6	87.1±0.6	48.2±0.5	78.9±0.6	36.1±0.5	74.2±0.6	28.5±0.5	68.2±0.7	19.0±0.4	70.5±0.7	22.5±0.3	71.2±0.7	24.2±0.4	63.8±0.7	12.5±0.4	67.8±0.6	19.8±0.4	73.0±0.5	25.6±0.5
	UniFORM	70.8±1.2	42.1±1.0	71.0±1.3	33.5±0.9	77.1±1.1	46.8±1.1	83.1±1.0	41.5±0.9	72.5±1.0	29.7±0.8	67.0±1.1	21.4±0.7	62.6±1.1	13.5±0.6	68.0±1.0	19.8±0.6	63.1±1.1	17.5±0.7	56.4±1.1	8.2±0.5	62.0±1.0	13.8±0.6	65.8±1.0	18.2±0.8
	GFM-UAD	75.5±0.8	49.2±0.7	77.8±0.9	40.5±0.7	82.4±0.8	53.8±0.8	80.2±0.9	39.8±0.8	75.8±0.8	33.2±0.6	72.1±0.8	26.0±0.6	66.0±0.9	17.1±0.5	72.5±0.9	23.9±0.4	68.0±0.8	21.8±0.5	61.8±0.8	11.5±0.5	66.2±0.8	18.1±0.5	71.0±0.8	23.5±0.6
	G-GLAD	62.4±1.3	10.5±1.1	68.5±1.2	12.3±1.0	71.7±1.3	16.8±1.1	72.0±1.2	28.5±1.0	63.5±1.2	8.2±0.9	58.0±1.2	6.5±0.8	55.5±1.1	4.5±0.6	61.1±1.2	12.7±0.6	56.9±1.2	7.2±0.7	48.5±1.2	1.5±0.4	52.5±1.2	1.8±0.2	58.9±1.2	5.7±0.8
	MoEGAD	58.5±1.4	9.0±1.2	65.1±1.4	16.2±1.1	68.2±1.4	24.5±1.2	74.5±1.1	31.0±1.0	60.1±1.3	15.0±0.8	54.2±1.3	4.8±0.7	52.1±1.2	3.2±0.5	58.2±1.3	9.5±0.5	52.8±1.3	3.8±0.6	45.9±1.3	1.1±0.3	48.5±1.3	1.5±0.5	54.2±1.3	8.1±0.7
	CgGFAD	**80.0**±0.7	**53.8**±0.8	**83.2**±1.2	**46.7**±1.1	**85.3**±0.6	**58.6**±1.1	**89.8**±0.7	**50.4**±1.3	**80.1**±0.9	**38.6**±0.7	**76.7**±0.5	**30.8**±1.8	**70.5**±1.1	**21.4**±0.8	**77.3**±1.5	**28.5**±0.3	**73.3**±1.2	**26.3**±1.1	**66.4**±1.0	**14.2**±0.9	**70.1**±0.7	**22.4**±0.5	**75.6**±0.6	**28.0**±1.4
G→N	UNPrompt	68.5±1.6	18.4±1.2	72.5±1.5	35.7±1.4	66.8±1.3	23.0±1.1	74.2±1.4	28.1±1.0	62.3±1.2	19.6±0.9	66.8±1.3	28.0±1.2	61.1±1.0	14.5±0.8	66.4±1.1	24.2±0.9	61.0±1.0	8.2±0.6	55.2±0.9	5.1±0.4	51.0±1.0	4.2±0.3	60.4±0.8	9.5±0.7
	AnomalyGFM	71.4±1.8	20.2±1.3	68.0±1.4	31.5±1.0	70.5±1.5	27.0±1.2	72.0±1.5	26.5±1.1	59.1±1.1	17.0±0.8	64.5±1.2	26.8±1.1	62.0±1.2	15.8±0.7	64.2±1.0	22.5±0.8	62.2±0.9	9.0±0.5	53.0±1.1	4.8±0.3	49.5±0.9	3.8±0.3	58.1±0.9	8.0±0.6
	Zero-GAD	66.0±1.9	16.7±1.0	70.2±1.6	33.0±0.9	68.2±1.4	24.8±1.3	76.5±1.3	30.0±1.2	63.0±0.9	21.2±0.7	69.1±1.0	32.0±1.0	59.5±1.1	12.2±0.5	68.0±0.9	26.1±0.7	60.2±0.8	7.1±0.4	56.8±0.8	5.8±0.4	52.8±0.8	4.5±0.4	62.2±0.7	10.8±0.5
	AffinityTune	62.9±1.7	14.4±1.5	62.4±1.3	22.7±0.8	61.6±1.2	16.0±1.0	69.1±1.4	21.4±1.1	54.5±1.0	13.2±0.8	60.2±1.1	21.5±0.9	57.2±1.0	11.0±0.6	60.1±0.9	18.0±0.8	58.5±0.8	6.5±0.4	51.1±1.0	3.8±0.3	47.2±0.8	2.8±0.2	54.0±0.8	6.8±0.5
	Ia-GGAD	69.5±1.5	19.8±0.9	66.7±1.2	29.3±0.8	69.2±1.3	25.0±1.1	73.8±1.2	27.0±1.0	61.0±1.0	18.8±0.7	67.0±0.9	29.2±0.9	60.0±0.9	13.8±0.5	67.2±0.8	25.5±0.7	61.1±0.7	7.8±0.4	54.2±0.7	4.9±0.3	50.5±0.7	3.6±0.3	60.8±0.6	9.0±0.5
	AutoGAD	64.6±1.4	15.8±1.2	71.2±1.3	34.9±0.8	67.9±1.1	22.0±1.2	75.1±1.1	29.1±0.9	62.8±0.8	20.5±0.6	68.2±0.8	30.8±0.8	62.1±0.8	14.8±0.5	69.1±0.7	27.8±0.6	62.8±0.6	8.5±0.4	55.1±0.6	5.2±0.3	51.0±0.6	4.0±0.3	62.9±0.5	10.1±0.4
	UniFORM	59.2±1.5	12.2±0.4	73.6±1.2	36.8±0.8	63.7±1.4	18.4±1.3	68.0±1.3	19.8±0.9	52.2±1.1	11.5±0.7	59.8±1.0	20.2±0.8	55.1±0.9	9.2±0.4	58.0±0.9	15.5±0.7	56.8±0.8	5.2±0.3	49.0±0.9	2.9±0.2	45.2±0.8	2.1±0.2	52.1±0.8	5.5±0.4
	GFM-UAD	67.2±1.1	17.0±0.8	74.0±1.0	38.1±0.9	68.8±0.9	24.1±0.6	74.8±1.0	28.2±0.8	61.8±0.8	19.2±0.5	67.5±0.7	29.8±0.7	61.8±0.7	14.1±0.4	66.8±0.7	25.0±0.5	61.8±0.5	8.0±0.3	54.8±0.5	5.0±0.3	50.2±0.5	3.7±0.2	61.0±0.5	9.2±0.4
	G-GLAD	55.2±1.4	4.2±0.4	58.6±1.2	11.8±0.7	56.7±1.3	6.4±1.2	62.0±1.2	14.2±0.8	48.2±1.0	1.5±0.2	52.8±0.9	1.8±0.1	49.0±0.8	1.1±0.1	52.0±0.8	3.2±0.2	50.0±0.7	1.2±0.2	44.1±0.7	0.8±0.0	54.0±0.7	3.3±0.4	46.5±0.7	1.1±0.1
	MoEGAD	52.0±1.2	3.0±0.3	54.0±1.1	5.1±0.6	53.8±1.1	4.1±0.6	59.1±1.1	12.0±0.7	45.5±0.9	0.8±0.1	49.2±0.8	1.2±0.2	46.2±0.7	0.9±0.1	48.8±0.7	0.5±0.0	47.1±0.6	1.1±0.1	55.5±0.6	5.6±0.1	50.1±0.6	2.4±0.1	53.8±0.6	2.7±0.2
	CgGFAD	**76.2**±1.3	**24.8**±0.9	**80.4**±0.6	**47.2**±1.3	**75.8**±0.6	**33.0**±1.4	**84.8**±0.7	**38.7**±1.5	**70.1**±1.3	**27.6**±1.1	**77.7**±0.9	**40.3**±1.4	**68.3**±1.2	**20.6**±0.6	**75.5**±0.8	**34.1**±1.5	**67.5**±0.6	**12.5**±1.4	**60.2**±1.1	**8.3**±1.2	**55.9**±1.1	**6.8**±1.5	**68.5**±0.7	**14.1**±1.4

Table 7: The additional ablation performance of cross-granularity anomaly detection in zero-shot settings on single-graph datasets.

	Training	Tolokers								Questions								T-Finance							
Task	Test	Reddit		Weibo		Amazon		Yelp		Reddit		Weibo		Amazon		Yelp		Reddit		Weibo		Amazon		Yelp	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→E	$\text{CgGFAD}_{\mathcal{L}_e}$	57.4±0.9	3.8±0.5	51.2±0.8	2.1±0.4	68.5±0.9	15.1±0.4	60.2±0.8	4.8±0.5	64.8±0.7	7.6±0.6	52.1±0.8	2.8±0.3	70.4±0.7	15.8±0.6	53.5±0.8	3.6±0.5	68.2±0.9	11.4±0.4	63.1±0.8	6.7±0.5	61.5±0.9	4.8±0.6	52.7±0.8	3.1±0.5
	$\text{CgGFAD}_{\mathcal{L}_a}$	55.1±1.1	3.1±0.6	48.6±1.0	1.7±0.3	65.8±1.2	13.2±0.5	58.5±1.0	4.1±0.6	61.9±0.9	6.2±0.7	49.8±0.9	2.1±0.4	67.2±1.0	13.9±0.7	51.0±1.1	2.8±0.4	65.9±1.1	9.8±0.5	60.5±0.9	5.3±0.6	58.9±1.1	3.7±0.5	50.1±0.9	2.3±0.4
	$\text{CgGFAD}_{\mathcal{L}_{sg}}$	58.9±0.8	4.2±0.4	53.7±0.7	2.8±0.4	69.8±0.8	16.3±0.3	62.8±0.7	5.5±0.4	66.2±0.6	8.5±0.5	54.3±0.7	3.3±0.3	72.1±0.6	17.0±0.5	55.8±0.7	4.2±0.4	70.5±0.8	12.9±0.3	65.4±0.7	7.8±0.4	63.8±0.8	5.6±0.5	54.8±0.7	3.8±0.4
	$\text{CgGFAD}_{\text{KL}}$	54.2±1.2	2.8±0.5	47.3±1.1	1.4±0.3	64.1±1.3	12.0±0.6	57.0±1.1	3.6±0.5	60.5±0.9	5.6±0.6	48.2±1.0	1.8±0.3	65.8±1.1	12.8±0.8	49.6±1.2	2.3±0.4	64.0±1.2	8.7±0.5	58.8±1.0	4.6±0.5	57.1±1.2	3.0±0.4	48.5±1.0	1.9±0.3
	$\text{CgGFAD}_{\mathcal{L}_x}$	56.8±0.9	3.5±0.4	52.0±0.8	2.3±0.4	67.2±0.9	14.3±0.4	61.3±0.8	5.0±0.4	63.5±0.7	7.0±0.5	51.0±0.8	2.5±0.3	69.1±0.7	14.9±0.6	52.4±0.8	3.2±0.4	67.0±0.9	10.6±0.4	62.0±0.8	6.1±0.5	60.2±0.9	4.2±0.5	51.5±0.8	2.7±0.4
	$\text{CgGFAD}_{\mathcal{L}_{\text{con}}}$	53.5±1.1	2.5±0.5	46.0±1.0	1.1±0.3	63.0±1.2	11.2±0.5	56.1±1.0	3.2±0.5	59.2±0.9	5.0±0.6	47.0±0.9	1.5±0.3	64.5±1.3	11.9±0.7	48.2±1.1	1.9±0.4	62.8±1.1	7.9±0.5	57.5±0.9	4.0±0.5	55.8±1.1	2.4±0.4	47.2±0.9	1.5±0.3
	$\text{CgGFAD}_{\mathcal{L}_{\text{rec}}}$	61.8±0.6	4.7±0.4	52.1±0.8	2.2±0.5	71.5±0.8	17.0±0.4	63.2±0.7	5.1±0.5	68.9±0.6	9.5±0.4	53.9±0.7	2.6±0.4	73.2±0.6	16.5±0.6	57.1±0.7	3.8±0.5	73.5±0.6	14.8±0.3	64.5±0.9	7.1±0.5	62.1±0.8	4.3±0.6	56.8±0.8	3.9±0.4
	CgGFAD_{δ}	56.7±0.8	3.1±0.5	56.3±0.6	3.1±0.4	67.0±0.9	13.5±0.5	65.8±0.6	6.2±0.4	64.0±0.8	6.8±0.6	57.1±0.6	3.8±0.3	76.2±0.5	19.1±0.5	54.2±0.4	2.9±0.6	69.0±0.8	11.5±0.4	68.1±0.6	8.9±0.4	66.5±0.7	6.2±0.4	52.1±0.8	2.8±0.5
	CgGFAD_{w}	62.0±0.5	4.4±0.4	53.8±0.7	2.7±0.4	72.8±0.7	18.1±0.3	61.5±0.8	4.3±0.6	69.8±0.5	8.7±0.5	50.5±0.8	1.9±0.5	72.0±0.7	15.2±0.7	58.9±0.6	4.8±0.4	71.2±0.7	12.8±0.4	63.0±0.8	6.2±0.6	63.8±0.8	4.9±0.5	57.9±0.6	4.2±0.4
	CgGFAD	**63.2**±0.5	**5.0**±0.4	**58.4**±0.6	**3.5**±0.5	**74.1**±0.7	**19.2**±0.3	**67.7**±0.6	**6.7**±0.4	**71.0**±0.5	**10.3**±0.5	**58.5**±0.6	**4.0**±0.4	**78.0**±0.5	**20.4**±0.5	**60.8**±0.6	**5.2**±0.4	**75.6**±0.7	**15.8**±0.3	**70.3**±0.5	**9.8**±0.4	**68.5**±0.7	**7.0**±0.5	**59.2**±0.6	**4.7**±0.4
E→N	$\text{CgGFAD}_{\mathcal{L}_e}$	67.1±1.1	14.2±0.8	51.8±0.9	5.1±0.5	65.2±1.1	12.3±0.6	51.0±1.0	2.8±0.4	48.5±0.9	1.8±0.3	54.2±1.0	2.1±0.4	71.5±0.9	21.0±0.8	68.4±1.0	11.5±0.7	47.3±1.1	1.2±0.3	52.1±0.9	2.9±0.4	62.8±1.0	8.1±0.6	43.5±1.1	0.7±0.2
	$\text{CgGFAD}_{\mathcal{L}_a}$	64.8±1.3	12.5±0.9	49.2±1.1	3.8±0.6	62.9±1.3	10.8±0.7	48.3±1.2	1.9±0.5	45.8±1.1	1.1±0.4	51.6±1.2	1.4±0.5	68.9±1.1	18.5±0.9	65.8±1.2	9.7±0.8	44.8±1.3	0.7±0.3	49.5±1.1	1.9±0.5	60.1±1.2	6.7±0.7	40.8±1.3	0.4±0.2
	$\text{CgGFAD}_{\mathcal{L}_{sg}}$	68.9±0.9	15.8±0.7	53.5±0.8	6.3±0.4	67.0±0.9	13.9±0.5	53.2±0.8	3.6±0.4	50.3±0.8	2.3±0.3	56.0±0.8	2.9±0.4	73.8±0.7	23.1±0.7	70.5±0.8	13.2±0.6	49.2±0.9	1.6±0.3	54.0±0.8	3.7±0.4	64.9±0.8	9.6±0.5	45.2±0.9	1.0±0.3
	$\text{CgGFAD}_{\text{KL}}$	63.2±1.4	11.0±1.0	47.5±1.2	2.9±0.7	61.0±1.4	9.5±0.8	46.5±1.3	1.2±0.5	44.0±1.2	0.8±0.4	49.8±1.3	1.0±0.5	67.0±1.2	16.8±1.0	64.0±1.3	8.2±0.9	43.0±1.4	0.4±0.2	47.8±1.2	1.2±0.5	58.2±1.3	5.5±0.8	39.0±1.4	0.3±0.2
	$\text{CgGFAD}_{\mathcal{L}_x}$	66.7±1.0	13.8±0.8	51.0±0.9	4.8±0.5	64.5±1.0	11.8±0.6	50.2±0.9	2.5±0.4	47.8±0.9	1.5±0.3	53.5±0.9	1.8±0.4	70.8±0.8	20.1±0.8	67.5±0.9	10.8±0.7	46.5±1.0	1.0±0.3	51.2±0.9	2.5±0.4	62.0±0.9	7.6±0.6	42.8±1.0	0.6±0.2
	$\text{CgGFAD}_{\text{Dist}}$	62.0±1.5	10.1±1.1	46.0±1.3	2.1±0.8	59.5±1.5	8.8±0.9	45.0±1.4	0.9±0.6	42.5±1.3	0.5±0.4	48.2±1.4	0.7±0.5	65.5±1.3	15.5±1.1	62.5±1.4	7.5±1.0	41.5±1.5	0.3±0.2	46.2±1.3	0.9±0.6	56.8±1.4	4.8±0.9	37.5±1.5	0.2±0.1
	$\text{CgGFAD}_{\text{sg}}$	73.5±0.6	18.8±0.5	58.5±0.9	9.3±0.6	71.2±0.7	16.8±0.4	54.0±0.8	3.8±0.5	52.0±0.8	2.1±0.4	60.5±0.7	5.1±0.4	74.8±0.7	24.2±0.7	71.5±0.8	13.5±0.6	52.8±0.8	2.1±0.3	54.2±0.8	3.8±0.5	68.5±0.7	12.3±0.5	46.0±0.9	0.9±0.3
	CgGFAD_{δ}	69.1±0.8	15.2±0.6	57.5±0.7	8.5±0.4	66.8±0.9	12.8±0.5	57.8±0.6	5.5±0.4	54.8±0.7	3.4±0.4	55.0±0.9	2.8±0.5	78.2±0.5	28.5±0.6	74.8±0.7	16.8±0.5	48.5±0.9	1.1±0.4	58.5±0.6	5.9±0.4	64.1±0.8	9.0±0.6	50.5±0.7	1.8±0.3
	CgGFAD_{w}	71.8±0.7	17.0±0.5	55.0±0.8	6.8±0.5	69.5±0.8	15.1±0.4	51.5±0.5	2.6±0.6	50.2±0.9	1.5±0.5	58.1±0.8	4.1±0.4	72.0±0.8	20.8±0.8	69.8±0.9	12.1±0.7	50.1±0.9	1.6±0.3	56.0±0.7	4.6±0.4	66.2±0.8	10.5±0.5	48.2±0.8	1.2±0.3
	CgGFAD	**75.0**±0.6	**20.5**±0.5	**60.5**±0.7	**10.8**±0.4	**73.2**±0.8	**18.5**±0.4	**59.8**±0.6	**6.8**±0.5	**57.0**±0.6	**4.5**±0.5	**62.5**±0.7	**6.2**±0.5	**80.0**±0.5	**30.5**±0.6	**76.8**±0.7	**18.5**±0.5	**55.8**±0.8	**3.0**±0.4	**60.2**±0.6	**7.0**±0.5	**70.5**±0.7	**14.2**±0.5	**52.2**±0.7	**2.0**±0.5

compared with all variants, validating the effectiveness of each component in CgGFAD. In addition, $\text{CgGFAD}_{\mathcal{L}_{sg}}$, CgGFAD_{δ}, and CgGFAD_{w} exhibit minor adverse effects compared with other variants.

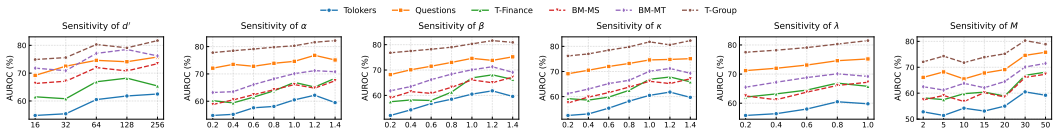

Figure 6: The additional sensitivity experiments. Solid line represents $N \rightarrow E$ task training on Weibo, and dashed line represents $N \rightarrow G$ task training on MUTAG.

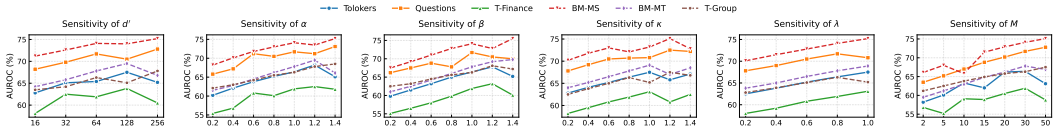

Figure 7: The additional sensitivity experiments. Solid line represents $N \rightarrow E$ task training on Amazon, and dashed line represents $N \rightarrow G$ task training on MNIST0.

Table 8: The additional ablation performance of cross-granularity anomaly detection in zero-shot settings on multi-graph datasets.

Task	Training	BM-MN						MUTAG						MNIST0						MNIST1					
	Test	BM-MS		BM-MT		T-Group		BM-MS		BM-MT		T-Group		BM-MS		BM-MT		T-Group		BM-MS		BM-MT		T-Group	
	Metrics	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR	RC	PR
N→G	CgGFAD$_{\mathcal{L}_r}$	70.2$_{\pm1.0}$	16.5$_{\pm0.8}$	73.8$_{\pm0.9}$	20.1$_{\pm0.7}$	67.9$_{\pm1.3}$	14.0$_{\pm0.6}$	76.3$_{\pm1.1}$	22.4$_{\pm0.9}$	72.5$_{\pm1.0}$	18.9$_{\pm0.8}$	71.2$_{\pm1.2}$	16.3$_{\pm0.7}$	64.8$_{\pm0.8}$	15.0$_{\pm1.0}$	71.6$_{\pm1.1}$	18.2$_{\pm0.6}$	69.5$_{\pm1.4}$	15.7$_{\pm0.5}$	58.9$_{\pm0.9}$	9.3$_{\pm0.8}$	62.4$_{\pm1.3}$	12.1$_{\pm0.9}$	69.0$_{\pm1.2}$	15.8$_{\pm0.6}$
	CgGFAD$_{\mathcal{L}_s}$	69.1$_{\pm1.3}$	15.8$_{\pm1.1}$	72.0$_{\pm1.2}$	19.0$_{\pm0.8}$	66.2$_{\pm1.5}$	13.1$_{\pm0.9}$	75.0$_{\pm1.4}$	21.5$_{\pm0.6}$	71.3$_{\pm1.6}$	18.1$_{\pm0.7}$	69.8$_{\pm1.1}$	15.2$_{\pm0.8}$	63.2$_{\pm1.0}$	13.2$_{\pm0.9}$	70.1$_{\pm1.3}$	17.5$_{\pm0.8}$	67.8$_{\pm1.6}$	14.3$_{\pm0.7}$	57.6$_{\pm1.2}$	8.5$_{\pm0.9}$	60.3$_{\pm1.4}$	11.4$_{\pm0.6}$	68.1$_{\pm1.3}$	14.9$_{\pm1.0}$
	CgGFAD$_{\mathcal{L}_{rs}}$	71.8$_{\pm1.1}$	18.9$_{\pm0.9}$	75.1$_{\pm1.3}$	22.7$_{\pm0.8}$	69.4$_{\pm1.6}$	15.8$_{\pm0.7}$	78.6$_{\pm1.0}$	24.8$_{\pm0.7}$	73.0$_{\pm1.2}$	20.5$_{\pm0.9}$	72.5$_{\pm1.3}$	17.9$_{\pm0.6}$	66.1$_{\pm0.9}$	16.6$_{\pm0.8}$	73.0$_{\pm1.4}$	19.9$_{\pm0.6}$	70.4$_{\pm1.5}$	16.6$_{\pm0.8}$	60.3$_{\pm1.0}$	10.8$_{\pm0.7}$	63.9$_{\pm1.1}$	13.9$_{\pm0.9}$	71.7$_{\pm1.4}$	17.3$_{\pm0.6}$
	CgGFAD$_{\mathrm{KL}}$	67.8$_{\pm1.2}$	14.2$_{\pm1.0}$	70.5$_{\pm1.4}$	17.3$_{\pm0.9}$	64.1$_{\pm1.7}$	11.6$_{\pm0.8}$	72.4$_{\pm1.3}$	19.1$_{\pm0.8}$	68.7$_{\pm1.5}$	15.8$_{\pm1.0}$	67.0$_{\pm1.4}$	13.1$_{\pm0.7}$	61.5$_{\pm1.1}$	11.8$_{\pm0.9}$	68.2$_{\pm1.5}$	15.1$_{\pm0.7}$	65.3$_{\pm1.6}$	12.4$_{\pm0.9}$	55.1$_{\pm1.3}$	7.2$_{\pm0.8}$	58.0$_{\pm1.2}$	9.5$_{\pm0.7}$	65.4$_{\pm1.5}$	12.8$_{\pm0.8}$
	CgGFAD$_{\mathcal{L}_r}$	72.5$_{\pm1.0}$	20.1$_{\pm0.8}$	76.2$_{\pm1.1}$	24.0$_{\pm0.7}$	71.0$_{\pm1.4}$	17.2$_{\pm0.8}$	79.8$_{\pm0.9}$	26.5$_{\pm0.8}$	74.2$_{\pm1.1}$	22.1$_{\pm0.8}$	73.8$_{\pm1.2}$	19.0$_{\pm0.7}$	67.0$_{\pm0.8}$	17.8$_{\pm0.7}$	74.2$_{\pm1.2}$	21.5$_{\pm0.5}$	71.5$_{\pm1.3}$	18.1$_{\pm0.7}$	61.5$_{\pm0.9}$	11.9$_{\pm0.6}$	65.2$_{\pm1.0}$	15.1$_{\pm0.8}$	72.8$_{\pm1.2}$	18.5$_{\pm0.7}$
	CgGFAD$_{\mathrm{Ind}}$	66.0$_{\pm1.4}$	12.5$_{\pm1.1}$	68.9$_{\pm1.5}$	15.2$_{\pm0.9}$	62.3$_{\pm1.8}$	10.0$_{\pm0.9}$	70.2$_{\pm1.5}$	17.0$_{\pm0.9}$	66.5$_{\pm1.6}$	13.6$_{\pm1.1}$	65.1$_{\pm1.5}$	11.2$_{\pm0.8}$	59.8$_{\pm1.3}$	10.1$_{\pm1.0}$	66.0$_{\pm1.6}$	13.0$_{\pm0.8}$	63.2$_{\pm1.7}$	10.5$_{\pm1.0}$	53.2$_{\pm1.4}$	5.8$_{\pm0.9}$	56.1$_{\pm1.4}$	7.8$_{\pm0.8}$	63.0$_{\pm1.6}$	10.7$_{\pm0.9}$
	CgGFAD$_{\mathrm{tg}}$	76.8$_{\pm0.8}$	38.5$_{\pm0.7}$	79.4$_{\pm0.9}$	35.0$_{\pm0.8}$	81.2$_{\pm0.7}$	42.0$_{\pm0.9}$	85.5$_{\pm0.8}$	39.2$_{\pm0.9}$	76.8$_{\pm0.8}$	29.5$_{\pm0.7}$	73.5$_{\pm0.6}$	24.8$_{\pm0.9}$	68.2$_{\pm0.9}$	19.1$_{\pm0.7}$	75.1$_{\pm0.9}$	24.6$_{\pm0.5}$	71.8$_{\pm0.8}$	21.5$_{\pm0.8}$	64.1$_{\pm0.8}$	12.5$_{\pm0.7}$	67.8$_{\pm0.8}$	18.2$_{\pm0.6}$	73.5$_{\pm0.7}$	22.1$_{\pm0.8}$
	CgGFAD$_{\mathrm{fj}}$	75.1$_{\pm0.9}$	35.0$_{\pm0.8}$	80.5$_{\pm0.7}$	39.2$_{\pm0.9}$	79.0$_{\pm0.8}$	38.5$_{\pm1.0}$	86.8$_{\pm0.6}$	43.0$_{\pm0.8}$	75.2$_{\pm1.0}$	27.0$_{\pm0.8}$	74.8$_{\pm0.7}$	27.1$_{\pm0.8}$	67.5$_{\pm1.0}$	18.2$_{\pm0.8}$	76.0$_{\pm0.8}$	26.1$_{\pm0.4}$	72.4$_{\pm0.9}$	23.0$_{\pm0.7}$	63.0$_{\pm0.9}$	11.2$_{\pm0.8}$	68.5$_{\pm0.7}$	19.5$_{\pm0.5}$	74.2$_{\pm0.6}$	24.5$_{\pm0.9}$
	CgGFAD$_{\mathrm{ae}}$	77.9$_{\pm0.7}$	42.1$_{\pm0.8}$	78.0$_{\pm1.0}$	32.8$_{\pm1.0}$	82.5$_{\pm0.6}$	46.3$_{\pm0.8}$	83.9$_{\pm0.9}$	36.0$_{\pm1.1}$	78.0$_{\pm0.7}$	32.4$_{\pm0.6}$	72.0$_{\pm0.8}$	22.5$_{\pm1.0}$	69.1$_{\pm0.8}$	20.2$_{\pm0.6}$	74.0$_{\pm1.1}$	22.8$_{\pm0.6}$	70.8$_{\pm0.9}$	19.8$_{\pm0.9}$	65.0$_{\pm0.7}$	13.4$_{\pm0.6}$	66.5$_{\pm0.5}$	16.1$_{\pm0.7}$	72.0$_{\pm0.8}$	20.0$_{\pm0.8}$
	CgGFAD	**80.0**$_{\pm0.7}$	**53.8**$_{\pm0.8}$	**83.2**$_{\pm1.2}$	**46.7**$_{\pm1.1}$	**85.3**$_{\pm0.6}$	**58.6**$_{\pm1.1}$	**89.8**$_{\pm0.7}$	**50.4**$_{\pm1.3}$	**80.1**$_{\pm0.9}$	**38.6**$_{\pm0.7}$	**76.7**$_{\pm0.5}$	**30.8**$_{\pm1.6}$	**70.5**$_{\pm1.1}$	**21.4**$_{\pm0.8}$	**77.3**$_{\pm1.5}$	**28.5**$_{\pm0.3}$	**73.3**$_{\pm1.2}$	**26.3**$_{\pm1.1}$	**66.4**$_{\pm1.0}$	**14.2**$_{\pm0.9}$	**70.1**$_{\pm0.7}$	**22.4**$_{\pm0.5}$	**75.6**$_{\pm0.6}$	**28.0**$_{\pm1.4}$
G→N	CgGFAD$_{\mathcal{L}_r}$	66.8$_{\pm1.5}$	14.0$_{\pm1.0}$	70.2$_{\pm1.3}$	17.5$_{\pm0.8}$	64.0$_{\pm1.7}$	11.5$_{\pm1.2}$	72.5$_{\pm1.4}$	19.2$_{\pm0.9}$	63.0$_{\pm1.6}$	16.0$_{\pm1.0}$	67.4$_{\pm1.5}$	13.2$_{\pm0.9}$	61.2$_{\pm1.2}$	11.0$_{\pm1.1}$	68.0$_{\pm1.5}$	15.5$_{\pm0.7}$	60.5$_{\pm1.8}$	8.0$_{\pm1.0}$	52.3$_{\pm1.3}$	4.8$_{\pm0.9}$	48.0$_{\pm1.4}$	3.7$_{\pm1.1}$	60.0$_{\pm1.6}$	8.0$_{\pm0.8}$
	CgGFAD$_{\mathcal{L}_s}$	68.9$_{\pm1.2}$	15.5$_{\pm0.9}$	72.5$_{\pm1.1}$	19.0$_{\pm0.7}$	66.5$_{\pm1.4}$	13.0$_{\pm0.8}$	75.0$_{\pm1.0}$	21.4$_{\pm0.8}$	64.3$_{\pm1.2}$	17.8$_{\pm0.9}$	70.0$_{\pm1.3}$	15.2$_{\pm0.8}$	62.5$_{\pm0.9}$	12.0$_{\pm1.0}$	70.3$_{\pm1.3}$	17.0$_{\pm0.7}$	61.3$_{\pm1.5}$	9.5$_{\pm0.6}$	53.8$_{\pm1.0}$	5.5$_{\pm0.8}$	49.4$_{\pm1.2}$	4.2$_{\pm0.9}$	61.0$_{\pm1.3}$	9.8$_{\pm0.7}$
	CgGFAD$_{\mathcal{L}_{rs}}$	65.1$_{\pm1.3}$	12.8$_{\pm1.1}$	68.4$_{\pm1.4}$	15.2$_{\pm0.9}$	62.0$_{\pm1.6}$	10.1$_{\pm1.0}$	70.8$_{\pm1.2}$	16.5$_{\pm0.9}$	61.0$_{\pm1.5}$	13.8$_{\pm1.1}$	65.2$_{\pm1.4}$	11.0$_{\pm0.9}$	59.0$_{\pm1.1}$	9.2$_{\pm0.9}$	66.0$_{\pm1.4}$	13.2$_{\pm0.8}$	58.1$_{\pm1.6}$	6.8$_{\pm0.8}$	50.1$_{\pm1.2}$	3.5$_{\pm0.8}$	46.2$_{\pm1.3}$	2.8$_{\pm0.7}$	58.0$_{\pm1.4}$	6.5$_{\pm0.8}$
	CgGFAD$_{\mathrm{Ind}}$	69.5$_{\pm1.1}$	16.0$_{\pm0.9}$	72.0$_{\pm1.2}$	19.8$_{\pm0.7}$	66.2$_{\pm1.5}$	12.8$_{\pm0.8}$	75.5$_{\pm1.1}$	21.8$_{\pm0.8}$	65.0$_{\pm1.3}$	18.5$_{\pm1.0}$	69.5$_{\pm1.4}$	14.8$_{\pm0.7}$	63.0$_{\pm1.0}$	12.8$_{\pm0.9}$	70.0$_{\pm1.4}$	17.0$_{\pm0.7}$	62.5$_{\pm1.6}$	10.1$_{\pm0.9}$	54.2$_{\pm1.1}$	5.9$_{\pm0.8}$	50.0$_{\pm1.2}$	4.8$_{\pm0.9}$	62.8$_{\pm1.4}$	10.5$_{\pm0.7}$
	CgGFAD$_{\mathcal{L}_r}$	71.0$_{\pm1.0}$	17.8$_{\pm0.8}$	74.2$_{\pm1.1}$	21.5$_{\pm0.6}$	68.1$_{\pm1.3}$	14.5$_{\pm0.7}$	77.2$_{\pm0.9}$	24.0$_{\pm0.7}$	66.8$_{\pm1.1}$	20.2$_{\pm0.8}$	72.0$_{\pm1.2}$	17.2$_{\pm0.6}$	64.5$_{\pm0.8}$	14.2$_{\pm0.8}$	72.1$_{\pm1.2}$	19.5$_{\pm0.6}$	63.8$_{\pm1.4}$	11.0$_{\pm0.7}$	55.8$_{\pm0.9}$	6.8$_{\pm0.7}$	51.5$_{\pm1.0}$	5.5$_{\pm0.8}$	64.2$_{\pm1.2}$	11.8$_{\pm0.6}$
	CgGFAD$_{\mathrm{Ind}}$	63.5$_{\pm1.6}$	11.2$_{\pm1.2}$	66.8$_{\pm1.5}$	13.5$_{\pm1.0}$	60.2$_{\pm1.8}$	8.9$_{\pm1.1}$	69.0$_{\pm1.4}$	14.8$_{\pm1.0}$	59.2$_{\pm1.7}$	12.0$_{\pm1.2}$	63.0$_{\pm1.6}$	9.5$_{\pm1.0}$	57.2$_{\pm1.3}$	7.8$_{\pm1.0}$	64.1$_{\pm1.5}$	11.5$_{\pm0.9}$	56.0$_{\pm1.7}$	5.2$_{\pm0.9}$	48.5$_{\pm1.4}$	2.8$_{\pm0.8}$	44.1$_{\pm1.5}$	1.9$_{\pm0.8}$	56.0$_{\pm1.5}$	5.0$_{\pm0.9}$
	CgGFAD$_{\mathrm{tg}}$	73.8$_{\pm0.9}$	21.0$_{\pm0.7}$	77.2$_{\pm0.8}$	38.5$_{\pm0.9}$	72.5$_{\pm0.8}$	25.0$_{\pm0.9}$	81.0$_{\pm0.8}$	32.5$_{\pm1.1}$	67.5$_{\pm1.0}$	22.8$_{\pm0.8}$	75.2$_{\pm0.8}$	31.0$_{\pm0.9}$	66.0$_{\pm1.0}$	16.5$_{\pm0.5}$	73.5$_{\pm0.9}$	26.8$_{\pm0.8}$	65.2$_{\pm0.7}$	10.8$_{\pm0.8}$	58.0$_{\pm0.8}$	7.2$_{\pm0.8}$	53.8$_{\pm0.9}$	5.9$_{\pm0.7}$	66.5$_{\pm0.8}$	12.8$_{\pm0.8}$
	CgGFAD$_{\mathrm{fj}}$	72.5$_{\pm1.0}$	19.2$_{\pm0.5}$	78.6$_{\pm0.7}$	42.0$_{\pm0.8}$	71.0$_{\pm0.9}$	22.8$_{\pm1.0}$	82.4$_{\pm0.7}$	35.0$_{\pm0.9}$	68.8$_{\pm0.9}$	25.0$_{\pm0.7}$	73.8$_{\pm0.9}$	28.2$_{\pm0.8}$	66.8$_{\pm0.8}$	18.0$_{\pm0.7}$	72.0$_{\pm1.0}$	24.2$_{\pm0.9}$	66.1$_{\pm0.6}$	11.0$_{\pm0.7}$	56.8$_{\pm0.9}$	6.5$_{\pm0.7}$	54.5$_{\pm0.8}$	6.2$_{\pm0.6}$	67.2$_{\pm0.6}$	13.5$_{\pm0.7}$
	CgGFAD$_{\mathrm{ae}}$	74.9$_{\pm0.8}$	22.8$_{\pm0.7}$	75.8$_{\pm1.0}$	34.0$_{\pm1.1}$	74.0$_{\pm0.7}$	28.5$_{\pm0.8}$	79.2$_{\pm0.9}$	29.8$_{\pm1.2}$	68.2$_{\pm0.8}$	24.1$_{\pm0.9}$	76.1$_{\pm0.7}$	34.5$_{\pm0.8}$	65.2$_{\pm0.9}$	15.2$_{\pm0.8}$	74.2$_{\pm0.7}$	29.5$_{\pm0.7}$	64.5$_{\pm0.8}$	9.8$_{\pm0.9}$	59.1$_{\pm0.7}$	7.8$_{\pm0.6}$	52.0$_{\pm1.0}$	5.1$_{\pm0.8}$	65.8$_{\pm0.8}$	12.0$_{\pm0.9}$
	CgGFAD	**76.2**$_{\pm1.3}$	**24.8**$_{\pm0.9}$	**80.4**$_{\pm0.6}$	**47.2**$_{\pm1.3}$	**75.8**$_{\pm0.6}$	**33.0**$_{\pm1.4}$	**84.8**$_{\pm0.7}$	**38.7**$_{\pm1.5}$	**70.1**$_{\pm1.3}$	**27.6**$_{\pm1.1}$	**77.7**$_{\pm0.9}$	**40.3**$_{\pm1.4}$	**68.3**$_{\pm1.2}$	**20.6**$_{\pm0.6}$	**75.5**$_{\pm0.8}$	**34.1**$_{\pm1.5}$	**67.5**$_{\pm0.6}$	**12.5**$_{\pm1.4}$	**60.2**$_{\pm1.1}$	**8.3**$_{\pm1.2}$	**55.9**$_{\pm1.1}$	**6.8**$_{\pm1.5}$	**68.5**$_{\pm0.7}$	**14.1**$_{\pm1.4}$

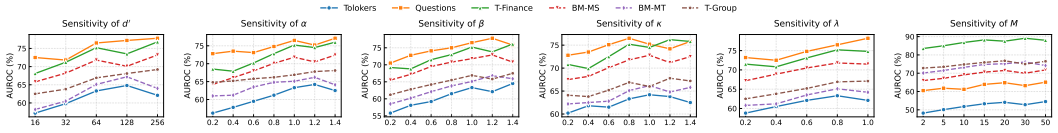

Figure 8: The additional sensitivity experiments. Solid line represents $N \rightarrow E$ task training on Yelp, and dashed line represents $N \rightarrow G$ task training on MNIST1.

G SENSITIVITY

We conduct additional sensitivity experiments on other datasets. Fig. 6, Fig. 7, and Fig. 8 exhibit the sensitivity performance on single-graph datasets (Tolokers, Questions, T-Finance) and multi-graph datasets (BM-MS, BM-MT, T-Group) with different training datasets. These observations help identify more suitable configurations of hyperparameters when deploying different training and testing datasets. In addition, we also exhibit the sensitivity of different tasks, as shown in Fig. 9.

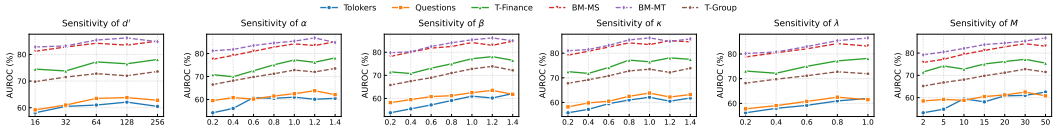

Figure 9: The additional sensitivity experiments. Solid line represents $E \rightarrow N$ task training on Reddit, and dashed line represents $G \rightarrow N$ task training on BM-MN.

H EFFICIENCY AND SCALABILITY

Fig. 10 presents the efficiency experiment regarding training time, using the rest of the datasets as training datasets. From the observations, we can see that CgGFAD maintains superior efficiency over other baselines. In addition, we implement extensive scalability experiments, as shown in Fig. 11. We validate the scalability on different tasks and on different testing datasets. The performance indicates that CgGFAD has powerful scalability as the number of edges increases.

I COMPLEXITY AND PSEUDO-CODE

We analyze the computational complexity. For SimSiam loss, the complexity is $\mathcal{O}(N_{t_{tr}} d')$. For MMD loss, the complexity is $\mathcal{O}((N_{t_{tr}}^2 + N_{t_{te}}^2 + N_{t_{tr}} N_{t_{te}}) d')$. For the KL divergence loss, the complexity is $\mathcal{O}(N_{t_{tr}} d')$. The complexity of prototype loss is $\mathcal{O}(N_{t_{tr}} d')$. The algorithm is shown in Algorithm 1.

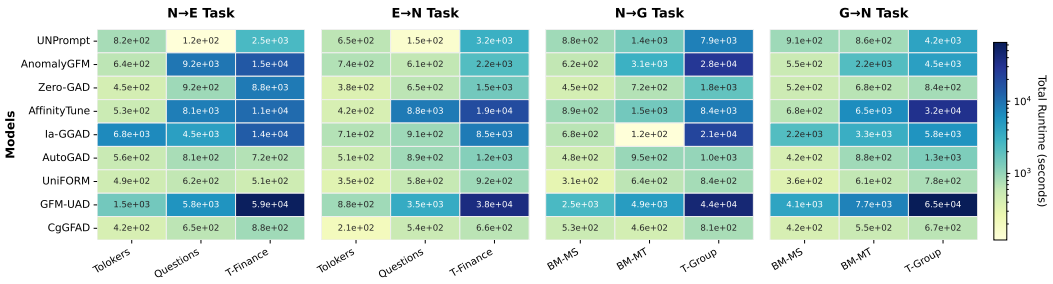

Figure 10: The efficiency experiment on training time.

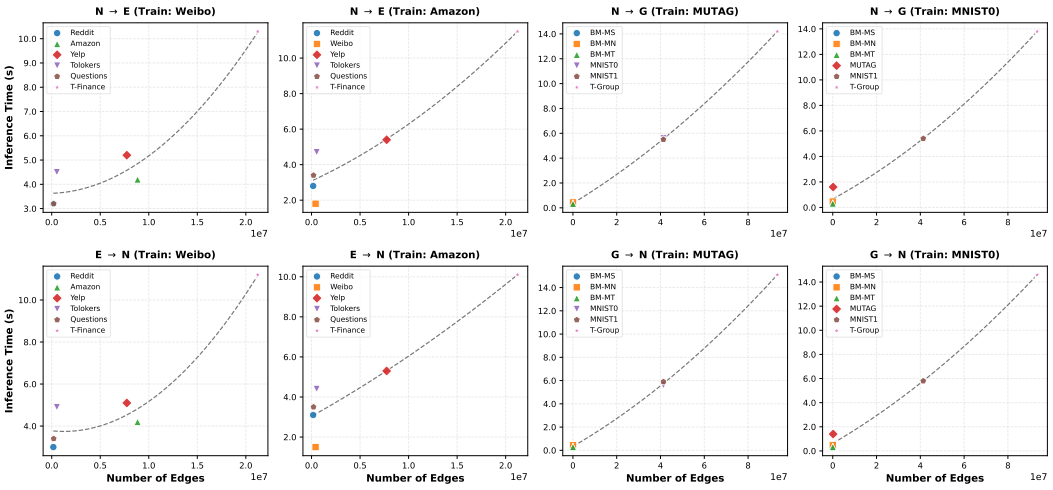

Figure 11: The scalability experiment on inference time.

Algorithm 1 Algorithm of CgGFAD.

Input: Graph $\mathcal{G} = (\mathcal{V}, \mathbf{X}, \mathbf{A})$; Epochs E.

Initialize the parameters of CgGFAD.
for $e \in (1, 2, \cdots, E)$ **do**
  Obtain node-level representations via Eq. (1) and calculate the node-level alignment via Eq. (2);
  Obtain granularity-specific representations and calculate the MMD between training and testing granularities via Eq. (4);
  Obtain $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ for cluster-based self-distillation via Eq. (5);
  Calculate target and prediction probabilities via Eq. Eq. (6)-Eq. (8);
  Capture self-distillation loss via Eq. (9);
  Capture the weighted generalization loss via Eq. (10);
  Calculate the total loss functions via Eq. (12);
  Update the model's parameters.
end for
return Scores

J LIMITATION

This study mainly focus on existing multi-granularity GAD paradigm, that is node-edge unification and node-graph unification. Due to the limited resources of datasets on graph-edge unification, we leave graph-edge unification task unexplored.