Revisiting Graph Contrastive Learning on Anomaly Detection: A Structural Imbalance Perspective

Yiming Xu[1,2], Zhen Peng[1,2*], Bin Shi[1,2], Xu Hua[1,2], Bo Dong[2,3], Song Wang[4], Chen Chen[4]

[1]School of Computer Science and Technology, Xi'an Jiaotong University
[2]Shaanxi Provincial Key Laboratory of Big Data Knowledge Engineering, Xi'an Jiaotong University
[3]School of Distance Education, Xi'an Jiaotong University
[4]University of Virginia, Charlottesville, Virginia, USA
{xym0924, huaxu}@stu.xjtu.edu.cn, zhenpeng27@outlook.com, {shibin, dong.bo}@xjtu.edu.cn, {sw3wv, zrh6du}@virginia.edu

Abstract

The superiority of graph contrastive learning (GCL) has prompted its application to anomaly detection tasks for more powerful risk warning systems. Unfortunately, existing GCL-based models tend to excessively prioritize overall detection performance while neglecting robustness to structural imbalance, which can be problematic for many real-world networks following power-law degree distributions. Particularly, GCL-based methods may fail to capture tail anomalies (abnormal nodes with low degrees). This raises concerns about the security and robustness of current anomaly detection algorithms and therefore hinders their applicability in a variety of realistic high-risk scenarios. To the best of our knowledge, research on the robustness of graph anomaly detection to structural imbalance has received little scrutiny. To address the above issues, this paper presents a novel GCL-based framework named AD-GCL. It 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. We further introduce intra- and inter-view consistency loss of the original and augmentation graph for enhanced representation. The performance evaluation of the whole, head, and tail nodes on multiple datasets validates the comprehensive superiority of the proposed AD-GCL in detecting both head anomalies and tail anomalies.

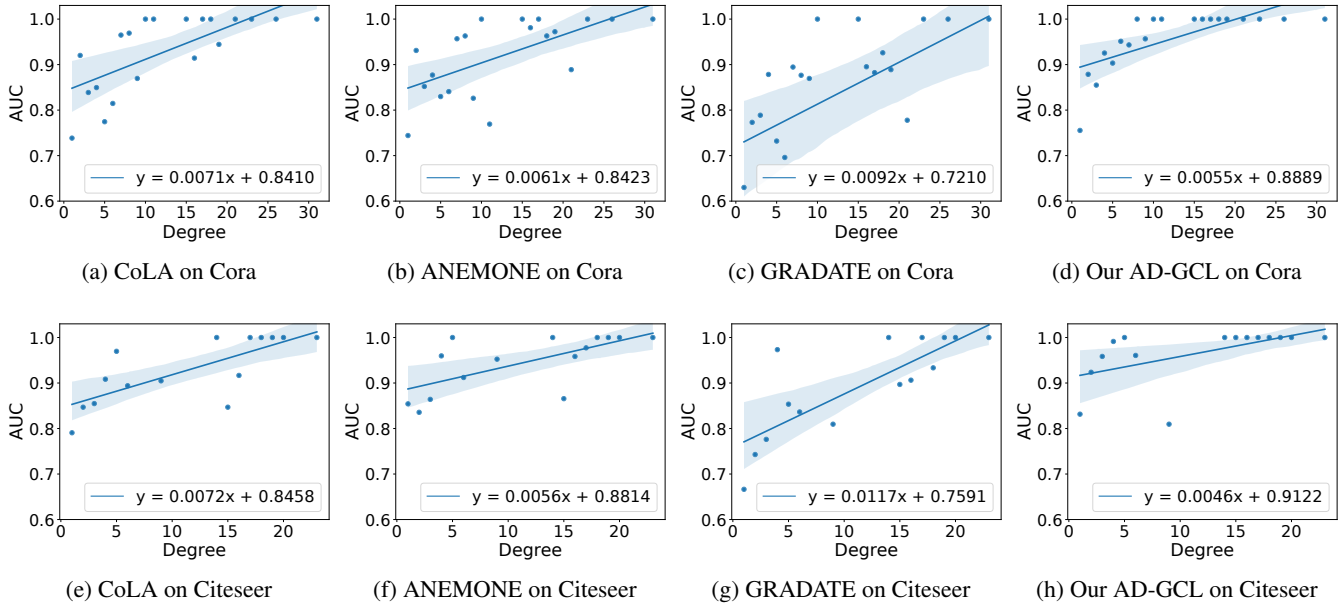
Code — https://github.com/yimingxu24/AD-GCL

Introduction

Graph anomaly detection plays a crucial role in various domains including cybersecurity (Ye et al. 2021), finance (Zheng et al. 2023), and distributed systems (Pazho et al. 2023). Its primary objective is to identify rare, unexpected, and suspicious observations that significantly deviate from the patterns of the majority in datasets, which helps minimize the considerable damage caused by detrimental events (Ma et al. 2021). Traditional anomaly detection methods are mainly based on feature space to uncover anomalies, neglecting the relational information inherent in real-world data. On the other hand, labeling anomalous samples usually requires domain experts and is undoubtedly time-consuming and expensive. Fortunately, graph contrastive learning (GCL) has demonstrated its superiority in mining graph structural information from unlabeled data in the wild, which has led to extensive research on contrastive learning-based unsupervised graph anomaly detection (GAD) in recent years (Liu et al. 2021b; Duan et al. 2023).

Existing GAD techniques are predominantly designed to optimize overall detection performance, without paying much attention to the algorithm's robustness over structural imbalance (Duan et al. 2023; Chen et al. 2024). However, such structural imbalance commonly exists in most real-world graphs that follow the power-law degree distribution, where a substantial fraction of nodes have low degrees, termed as tail nodes while the rest are referred to as head nodes (see Appendix for statistical analysis on the degree distribution). So it naturally raises a concern: can GCL-based anomaly detection models that combine the advantages of both graph neural networks (GNNs) and contrastive learning meet the challenge of structural imbalance? To figure this out, we choose three representative models for empirical study. As illustrated in Figure 1, the detection performance notably drops as the node degree decreases. It reveals a nonnegligible finding that existing GCL-based anomaly detection methods struggle to effectively capture tail anomalies as we expected. This disparity has raised concerns regarding the security and robustness of current anomaly detection methods in identifying tail anomalies, which severely impair their application in several high-impact detection scenarios. For instance, in social networks, most identity impersonation attacks are not targeted at celebrities (usually with very high node degrees) but involve cloning ordinary individuals (with low degrees) within the network to engage in malicious activities (Goga, Venkatadri, and Gummadi 2015; Hooi et al. 2017). Similarly, in transaction networks, fraudsters often establish multiple shell companies hidden within complex networks to commit one-time financial fraud and then shut down the business (Pacini et al. 2018), resulting in

*Corresponding author.



(a) CoLA on Cora (b) ANEMONE on Cora (c) GRADATE on Cora (d) Our AD-GCL on Cora

(e) CoLA on Citeseer (f) ANEMONE on Citeseer (g) GRADATE on Citeseer (h) Our AD-GCL on Citeseer

Figure 1: An illustrative showcasing performance disparity between head and tail anomalies. Blue scatters represent the AUC at a specific degree. The blue line represents the regression line, with a steeper slope indicating a more pronounced structural bias.

massive low-degree anomalies. Motivated by this observation, a natural question arises: *Can we improve the detection capability of tail anomalies without compromising the performance of head anomalies?*

Through in-depth investigation, we find that the successful identification of anomalies within the contrastive learning paradigm is based on local inconsistency mining (Chen et al. 2024). It begins with learning the matching patterns between normal nodes and their neighbors (i.e., satisfying the homophily assumption), which are dominant in the graph. In contrast, anomalies are consistently different from their local neighbors, and due to their irregularity and diversity, it is much more difficult to adapt to prevailing patterns in the data (Liu et al. 2021b; Jin et al. 2021; Duan et al. 2023). For example, CoLA (Liu et al. 2021b) enforces the discriminator to learn the neighbor matching patterns by treating nodes and their subgraphs sampled by a multi-round random walk with restart (RWR) as positive pairs, and subgraphs sampled based on other nodes as negative pairs. In the testing phase, the target node and its sampled neighboring structures are input into the optimized discriminator to derive a specific anomaly score. Notably, head anomalies possess more abundant structural links that ensure the diversity of constructed sample pairs, making it easier for the discriminator to model this difference. In contrast, the tail anomalies do not have enough structural relations to form sufficiently diverse contrast pairs, making the discriminator prone to overfit its sparse neighborhood and misinterpret these contrast pairs as normal graph patterns. Hence, the performance observed on tail anomalies declines. Inspired by the successful characteristics of head nodes, an intuitive solution is to enlarge the neighborhood of tail nodes by actively mining potential and underlying relational topologies to enhance the diversity of contrast samples and guide the discriminator to model the correct boundary between benign and anomaly. However, graph completion is not trivial due to the non-independently and identically distributed (i.i.d.) nature of the graph-structured data (Wu et al. 2020), naive completion may lead to suboptimal solutions by degrading the performance of neighboring nodes.

To improve the detection performance from the perspective of structural imbalance, especially the identification of tail anomalies, this paper presents a novel **A**nomaly **D**etection model based on **GCL**, named AD-GCL. Specifically, for tail nodes (classified based on topology only, regardless of whether the node is abnormal or not), we propose an anomaly-guided neighbor completion strategy to enlarge their limited neighboring receptive field, which employs a mixup between the ego network of the anchor tail node and the ego network of auxiliary nodes sampled based on anomaly scores and feature saliency information. As for head nodes, we introduce a neighbor pruning strategy based on the saliency information of features between nodes. The advantages of this strategy are two-fold: (1) it effectively filters out noise information associated with head nodes, such as problematic edge relationships; (2) it forges head nodes as tail nodes and regards the original view and the neighbor pruning view as contrastive views. Based on this, the knowledge learned from head nodes is used as guidance to provide supervisory signals to the corresponding forged tail nodes, facilitating the detection of genuine tail nodes. By enforcing the discriminator to maximize the agreement between the original and augmented graphs in terms of anomaly scores and features of nodes, anomalies that deviate from the majority of the data will be assigned higher anomaly scores. Our main contributions are summarized as follows:

• ***Significance***: We investigate the limitations of existing GCL-based graph anomaly detection models from a struc-

tural imbalance perspective through detailed statistical analysis. Our findings disclose the undesirable performance of existing works in spotting tail anomalies, which severely limits their application in various high-impact detection scenarios in the open world.

• ***Algorithm***: We propose a novel algorithm AD-GCL, which introduces neighbor pruning to filter out noisy edges for head nodes and enhances the model's performance in detecting genuine tail nodes by aligning head nodes with forged tail nodes. Moreover, we implement neighbor completion to enlarge the neighborhood for tail nodes and incorporate intra- and inter-view consistency loss as self-supervision to uncover abnormal patterns.

• ***Experiments***: Comprehensive performance evaluation of the whole, head, and tail nodes on six public datasets demonstrates the superiority of our proposed AD-GCL over the baselines. AD-GCL effectively enhances the detection quality for both tail and head anomalies.

Related Work

Anomaly Detection on Graphs. With the rise of graph-structured data, GAD has gained traction (Xu et al. 2024). DOMINANT (Ding et al. 2019) measures feature and structure reconstruction errors. GAD-NR (Roy et al. 2023) detects anomalies using reconstructed node neighborhoods. Contrastive learning, successful in various domains (He et al. 2020; Xu et al. 2023b), has been extended to GAD. CoLA (Liu et al. 2021b) first introduces GCL by measuring whether the target node matches its neighbor sample pairs. ANEMONE (Jin et al. 2021) adds patch-level (i.e., node versus node) consistency, and GRADATE (Duan et al. 2023), Sub-CR (Zhang and et al. 2022) and SAMCL (Hu et al. 2023) combine multi-view contrasts to estimate the anomaly score of nodes further. GADAM (Chen et al. 2024) designs adaptive message passing to avoid local anomaly signal loss. Despite their promise, GCL-based methods struggle with tail anomalies, reducing overall detection effectiveness.

Imbalanced Learning on Graphs. Imbalanced learning problems are widespread in real-world scenarios (Liu et al. 2023). Most research focuses on class imbalance (Park, Song, and Yang 2022; Zeng et al. 2023), with some addressing it in GAD (Liu et al. 2021a). Subsequent work noted graph structural bias. Early approaches introduce degree-specific graph convolution (Wu, He, and Xu 2019; Tang et al. 2020). Tail-GNN (Liu, Nguyen, and Fang 2021) transfers neighborhood translations from head to tail nodes. GRADE (Wang et al. 2022) shows the ability of GCL to handle structural imbalance. SAILOR (Liao et al. 2023) and GRACE (Xu et al. 2023a) enhance tail node representation by structural augmentation. Despite significant achievements, most methods are semi-supervised and focus on node classification. The ability of anomaly detection methods to handle structural imbalances has received less scrutiny.

Methodology

In this paper, we focus on the unsupervised anomaly detection problem for attributed graphs $\mathcal{G} = (\mathbf{A}, \mathbf{X})$, where $\mathcal{V} = \{v_1, \cdots, v_n\}$ is the set of nodes ($|\mathcal{V}| = n$), $\mathbf{A} \in \mathbb{R}^{n \times n}$ denotes the adjacency matrix and $\mathbf{X} \in \mathbb{R}^{n \times f}$ is the attribute matrix. For each node $v \in \mathcal{V}$, $\mathcal{N}_v$ is the neighborhood of v and $|\mathcal{N}_v|$ is the degree of v. Following the convention of prior works (Liu et al. 2020; Liu, Nguyen, and Fang 2021), K is predefined as a degree threshold, tail nodes are defined as $\mathcal{V}_{tail} = \{v : |\mathcal{N}_v| \leq K\}$, and other nodes are called head nodes $\mathcal{V}_{head} = \{u : |\mathcal{N}_u| > K\}$.

The goal of unsupervised anomaly detection is to learn an anomaly score function $f(\cdot) : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times d} \to \mathbb{R}^n$, which estimates the anomaly score $S_i = f(v_i)$ of each node. The anomaly score s_i measures the extent of abnormality for node v_i. A larger anomaly score means that it is more likely to be anomalous. Unlike previous anomaly detection work that only focused on the overall goal, in this paper, we pay attention to the robustness of structural imbalance, i.e., the detection performance of tail nodes $\{f(v) : v \in \mathcal{V}_{tail}\}$ and head nodes $\{f(u) : u \in \mathcal{V}_{head}\}$.

Neighbor Pruning Strategy

We propose a neighbor pruning strategy that leverages saliency information of inter-node features to effectively filter out noise information associated with head nodes, such as certain cunning abnormal objects that tend to establish relationships with these normal objects to camouflage themselves as conforming to the prevailing patterns (Ma et al. 2021). Note that models often perform better on high-resource groups (i.e., head nodes) with sufficient structural information. Therefore, the neighbor pruning strategy forges head nodes as tail nodes, constructing contrastive pairs between the original view and the neighbor pruned view. We expect to utilize knowledge learned from head nodes to supervise forged tail nodes through alignment, thereby enhancing the model's performance on genuine tail nodes. Specifically, for the head node $u \in \mathcal{V}_{head}$, we sample from the following multinomial distribution:

$$\widetilde{\mathcal{N}}_u \sim \text{Multinomial}\left(K; p(\cdot|u) \cdot p_{\text{sim}}(\cdot|u)\right), \tag{1}$$

where $\left|\widetilde{\mathcal{N}}_u\right| = K$, so that the forged tail node u retains K important edges while discarding potential problematic or redundant relationships in the head nodes. The neighbor distribution $p(v|u) = 1/|\mathcal{N}_u|$ if $v \in \mathcal{N}_u$ and $p(v|u) = 0$ otherwise. The feature similarity distribution $p_{sim}(v|u) = \text{sim}(x_v, x_u)$ for $v \neq u$ and $p_{sim}(u|u) = 0$, where all the features are ℓ_2 normalized and dot product (cosine) similarity is used to compare them $\text{sim}(q, k) = q^\top k / \|q\| \|k\|$.

Unlike existing strategies based on supervised label-aware (Dou et al. 2020) or degree-proportional pruning (Wang et al. 2022), we introduce a similarity-aware method to prune head nodes into forged tail nodes instead of discarding them proportionally. By employing self-supervised alignment between head and forged tail nodes in both features and anomaly scores, we enhance the discriminator's ability to model accurate decision boundaries for both node types (refer to the ablation study).

Anomaly-Guided Neighbor Completion

To improve the detection performance of tail nodes, a natural solution is to enlarge the tail node neighborhood by

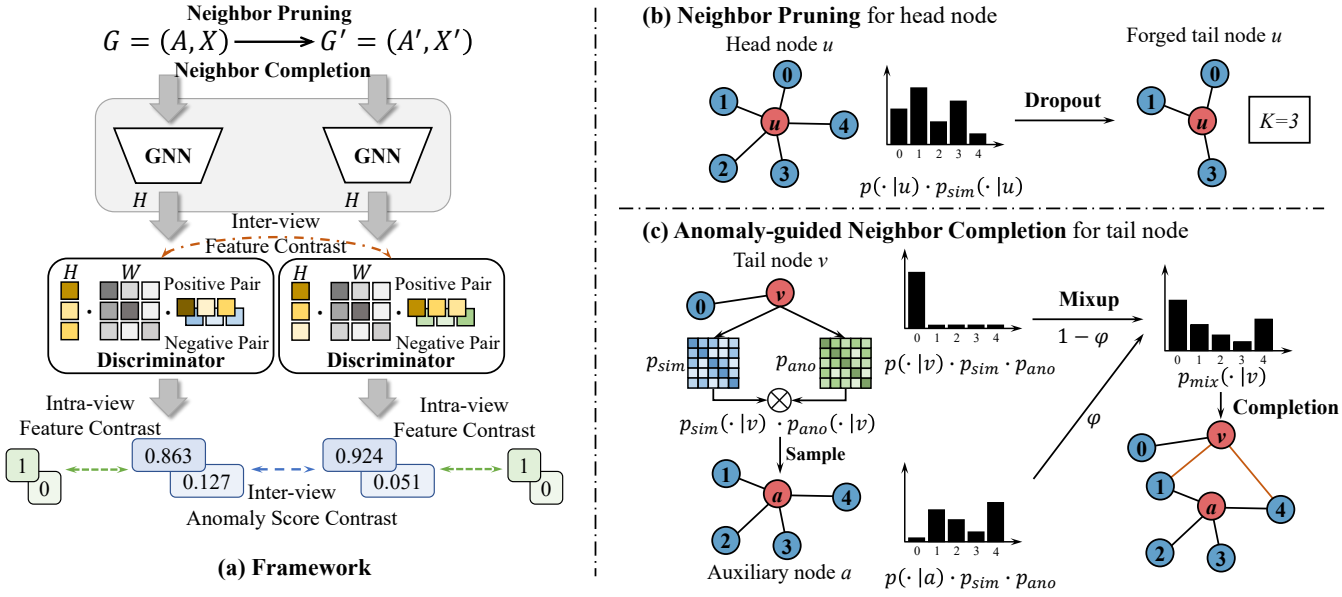


Figure 2: (a) The architecture of the AD-GCL; (b) Neighbor pruning to filter noise edges of head nodes; (c) Anomaly-guided neighbor completion to enlarge the receptive field of tail nodes. Intra- and inter-view contrast to enhance representation.

complementing potential or hidden relationships. However, due to the non-i.i.d. nature of the graph, biased graph completion may lead to a further decline in performance (as shown in Appendix). To address the aforementioned problem, we propose a novel anomaly-guided neighbor completion strategy for tail nodes. We first adopt a joint distribution of anomaly score similarity and feature saliency to sample auxiliary nodes similar to anchor tail nodes. Subsequently, we enlarge the neighborhood of tail nodes by mixing the ego network of anchor tail nodes and the ego network of auxiliary nodes. Different from PC-GNN (Liu et al. 2021a), which incorporates label distribution information to design a neighbor sampler for over-sampling the neighborhood of the minority class nodes, our method effectively enhances the receptive field of tail nodes in an unsupervised manner by integrating both anomaly and feature information.

In specific, we first derive a scoring matrix S to record the abnormality of nodes. Let t represent the current number of training epochs. In each epoch, the discriminator generates anomaly scores $s^t_{v,p}$ and $s^t_{v,n}$ for the anchor tail node with its positive and negative pairs, respectively (refer to Section Graph Contrastive Network for detailed explanation). Considering the fluctuation of anomaly scores during training, we use a sliding window to store the training results of the past w epochs. For a specific tail node v, its anomaly score is represented as $S_v = \left[s^{t-w+1}_{v,p}, s^{t-w+1}_{v,n}, \cdots, s^t_{v,p}, s^t_{v,n}\right]$, and $S \in \mathbb{R}^{n \times 2w}$. The abnormal similarity distribution is modeled as $p_{ano}(u, v) = S_u \cdot S_v^T$. For any tail node v, we jointly consider anomaly score similarity and feature similarity, and sample an auxiliary node a from the multimodal distribution Multinomial $(p_{nc}(\cdot|v))$, where $p_{nc}(\cdot|v) = p_{\text{sim}}(\cdot|v) \cdot p_{\text{ano}}(\cdot|v)$. Then, we form a larger neighborhood distribution by mixing up the ego network of nodes v and a. The neighborhood distribution of v after mixup is defined as:

$$p_{mix}(\cdot|v) = (1-\varphi)\left(p(\cdot|v) \cdot p_{nc}(\cdot|v)\right) + \varphi\left(p(\cdot|a) \cdot p_{nc}(\cdot|a)\right), \tag{2}$$

where the mixing ratio φ of auxiliary node a increases with the similarity to the tail node v, and φ is at most 0.5 to preserve the original neighborhood of the anchor tail node. Finally, we sample neighbors for the tail node v from the mixup neighborhood distribution $p_{mix}(\cdot|v)$. The number of neighbors is sampled from the degree distribution of the original graph to keep degree statistics.

Graph Contrastive Network

The key challenge in contrastive learning is selecting appropriate contrastive pairs. Most normal nodes tend to adhere to the homophily hypothesis, exhibiting potential matching patterns with their neighbors. On the contrary, anomalous nodes often demonstrate inconsistencies with their neighbors. Based on this, the graph contrastive network treats the anchor node and its neighbors as intra-view positive pairs, and the neighbors of other nodes as intra-view negative pairs. Furthermore, during the initial training stage, we regard the original graph (called $view1$) and the neighbor pruning graph (called $view2$) as contrastive views, and utilize well-trained head nodes from the original graph to supervise the corresponding forged tail nodes, i.e., the corresponding nodes are inter-view positive pairs. In the later stage of training, we use the neighbor completion strategy to enlarge the receptive field of tail nodes, generating two augmented graphs (referred to as $view1$ and $view2$) as contrastive views to further enhance node representations.

Specifically, given a graph $\mathcal{G}^{view} = \left(\mathbf{A}^{view}, \mathbf{X}^{view}\right)$, where $view$ denote contrastive view 1 or 2, representing the original graph or augmentation graph. The node representa-

tions are updated by a GNN-based encoder $f(\cdot)$:

$$\mathbf{H}^{view} = f\left(\mathbf{A}^{view}, \mathbf{X}^{view}\right), \tag{3}$$

where $\mathbf{H}^{view} \in \mathbb{R}^{n \times d}$ are the node representations of graph $\mathcal{G}^{view}$.

Then, we use the Readout function to calculate the neighbor representation of the node, which has been extensively used in prior works (Xu et al. 2023b). The formal expression is as follows:

$$\mathbf{h}_{\mathcal{N}_i}^{view} = \text{Readout}\left(\left\{\mathbf{h}_j^{view}, \forall j \in \hat{\mathcal{N}}_i^{view}\right\}\right), \tag{4}$$

where $\mathbf{h}_{\mathcal{N}_i}^{view} \in \mathbb{R}^d$ is the neighbor representation of node i. $\hat{\mathcal{N}}_i^{view}$ represents the neighbor set sampled for node i using random walk with restart (RWR) (Tong, Faloutsos, and Pan 2006) in $\mathcal{G}^{view}$. We adopt the average pooling function as our readout function. This approach is simple and effective, while not introducing additional model parameters.

Inspired by existing anomaly detection works based on GCL (Liu et al. 2021b), we devise a discriminator module to evaluate the potential edge relationships in the graph, leveraging both node representations and node neighbor representations. The discriminator is built by the bilinear scoring function. The predicted similarity score of the discriminator can be calculated as follows:

$$s_i^{view} = \text{Bilinear}\left(\mathbf{h}_{\mathcal{N}_j}^{view}, \mathbf{h}_i^{view}\right) = \sigma\left(\mathbf{h}_{\mathcal{N}_j}^{view} \mathbf{W} \mathbf{h}_i^{view\top}\right), \tag{5}$$

where $\mathbf{W}$ is a trainable matrix, and $\sigma(\cdot)$ is Sigmoid function. When the input to the discriminator is the potential neighbor relationship in the graph, i.e., $i = j$, the output score of the positive pair is $s_{i,p}^{view}$. When the input to the discriminator is a negative pair, i.e., $i \neq j$, the output score is $s_{i,n}^{view}$. $s_{v,p}^t$ are mentioned in the previous subsection is $s_{v,p}^t = s_{v,p}^{view1} + s_{v,p}^{view2}$ at training step t.

The representations of normal nodes, which are dominant in the graph, and their neighbors tend to be similar, aiming for $s_{i,p}^{view}$ to be close to 1. In contrast, normal nodes and neighbors of other nodes may be dissimilar in negative pairs. Therefore, we adopt binary cross-entropy (BCE) loss (Veličković et al. 2019) to train the intra-view contrastive pairs:

$$\mathcal{L}_{intra}^{view} = -\frac{1}{2n} \sum_{i=1}^{n} \left(\log\left(s_{i,p}^{view}\right) + \log\left(1 - s_{i,n}^{view}\right)\right), \tag{6}$$

where $\mathcal{L}_{intra}$ is the sum of losses from both views, denoted as $\mathcal{L}_{intra} = \mathcal{L}_{intra}^{view1} + \mathcal{L}_{intra}^{view2}$.

Then, we enforce the agreement between the two contrastive views by InfoNCE (Van den Oord, Li, and Vinyals 2018). We treat the features learned and scores output by the discriminator of the corresponding nodes as inter-view positive pairs. The inter-view contrastive loss is as follows:

$$\mathcal{L}_{inter} = \text{ctr}\left(\mathbf{H}^{view1}, \mathbf{H}^{view2}\right) + \text{ctr}\left(\mathbf{S}^{view1}, \mathbf{S}^{view2}\right) \tag{7}$$

where $\mathbf{H}$ represents the node feature and $\mathbf{S}$ is the score output by the discriminator. $\text{ctr}(\cdot, \cdot)$ is as follows:

$$\text{ctr}(\mathbf{Q}, \mathbf{K}) = -\sum_{i=1}^{N} \log \frac{\exp\left(\mathbf{q}_i \cdot \mathbf{k}_i / \tau\right)}{\sum_{j=1}^{N} \exp\left(\mathbf{q}_i \cdot \mathbf{k}_j / \tau\right)}, \tag{8}$$

where τ is a temperature parameter.

In the training phase, we jointly train the intra-view and inter-view contrastive losses. The overall objective is defined as follows:

$$\mathcal{L} = \mathcal{L}_{intra} + \alpha \mathcal{L}_{inter} \tag{9}$$

where α is a trade-off parameter to balance the importance between two objectives. Detailed complexity analysis is given in the Appendix.

Anomaly Score Calculation

After the graph contrastive network is well-trained, we use a statistical anomaly estimator (Liu et al. 2021b) during the inference stage to calculate anomaly scores for each node based on local inconsistency. For node i, the statistical anomaly estimator generates the final anomaly score S_i through multiple rounds of positive and negative sampling prediction scores:

$$S_i = \frac{\sum_{r=1}^{R} \left(s_{i,n}^r - s_{i,p}^r\right)}{R} \tag{10}$$

where R is the number of sampling rounds. The scores of the negative pair $s_{i,n}^r$ and positive pair $s_{i,p}^r$ are obtained from the discriminator in the r-th round. For a normal node, the positive pair score $s_{i,p}$ should be close to 1, while the negative pair score $s_{i,n}$ should be close to 0. However, the discriminator struggles to distinguish the matching patterns of positive and negative pairs for abnormal nodes. The higher the score of S_i, the stronger the indication of abnormality.

Experiments

Experiment Settings

Datasets. To conduct a comprehensive comparison, we evaluate AD-GCL on six widely used benchmark datasets for anomaly detection. Specifically, we choose two categories of datasets: 1) citation networks (Liu et al. 2021b; Jin et al. 2021) including Cora, Citeseer, and Pubmed, 2) bitcoin trading networks (Kumar et al. 2016, 2018) including Bitcoinotc, BITotc, and BITalpha.

Baselines. We compare AD-GCL with ten popular methods for anomaly detection: three classical non-deep methods, i.e., clustering-based SCAN (Xu et al. 2007), matrix factorization-based Radar (Li et al. 2017) and ONE (Bandyopadhyay, Lokesh, and Murty 2019); seven neural network based frameworks, i.e., AEGIS (Ding et al. 2021), GAAN (Chen et al. 2020), CoLA (Liu et al. 2021b), ANEMONE (Jin et al. 2021), AdONE (Bandyopadhyay et al. 2020), GRADATE (Duan et al. 2023), and GAD-NR (Roy et al. 2023).

Evaluation Metrics. For evaluating the performance of the proposed methods, we employ a widely-used anomaly detection metric ROC-AUC (Chen et al. 2020; Ding et al. 2021; Duan et al. 2023; Jin et al. 2021; Li et al. 2017; Liu et al. 2021b; Roy et al. 2023). The ROC curve plots the true positive rate against the false positive rate, while AUC is the area under the ROC curve with a higher value indicating better performance. We also report the AUPRC and AP results in the Appendix. To comprehensively analyze the

Method	Cora			Citeseer			Pubmed		
	AUC	TN	HN	AUC	TN	HN	AUC	TN	HN
SCAN	$67.42_{\pm 0.14}$	$46.74_{\pm 0.27}$	$74.22_{\pm 0.18}$	$69.80_{\pm 0.74}$	$48.70_{\pm 0.84}$	$76.81_{\pm 0.14}$	$74.16_{\pm 0.68}$	$49.77_{\pm 0.38}$	$89.93_{\pm 0.06}$
Radar	$52.62_{\pm 0.33}$	$48.89_{\pm 0.79}$	$54.39_{\pm 0.46}$	$62.30_{\pm 0.87}$	$49.32_{\pm 0.96}$	$58.86_{\pm 0.04}$	$49.34_{\pm 0.49}$	$49.75_{\pm 0.01}$	$50.31_{\pm 0.30}$
ONE	$49.20_{\pm 0.42}$	$48.87_{\pm 0.44}$	$49.52_{\pm 0.78}$	$49.79_{\pm 0.10}$	$49.66_{\pm 0.13}$	$50.09_{\pm 0.91}$	$50.28_{\pm 0.82}$	$50.02_{\pm 0.78}$	$50.34_{\pm 0.52}$
AEGIS	$48.55_{\pm 1.34}$	$46.71_{\pm 1.40}$	$50.29_{\pm 1.41}$	$47.67_{\pm 1.05}$	$47.66_{\pm 1.26}$	$48.57_{\pm 1.70}$	$57.02_{\pm 1.24}$	$64.51_{\pm 1.37}$	$50.93_{\pm 1.85}$
GAAN	$53.54_{\pm 1.29}$	$50.69_{\pm 1.49}$	$53.20_{\pm 1.80}$	$51.48_{\pm 1.44}$	$49.92_{\pm 1.27}$	$50.24_{\pm 1.08}$	$52.66_{\pm 1.49}$	$48.94_{\pm 1.03}$	$51.45_{\pm 1.14}$
CoLA	$86.45_{\pm 0.62}$	$82.59_{\pm 0.36}$	$87.88_{\pm 0.43}$	$89.04_{\pm 0.56}$	$82.31_{\pm 0.77}$	$97.71_{\pm 0.35}$	$93.54_{\pm 0.37}$	$92.38_{\pm 0.50}$	$97.08_{\pm 0.29}$
ANEMONE	$90.70_{\pm 0.91}$	$85.09_{\pm 0.29}$	$94.19_{\pm 0.61}$	$92.04_{\pm 0.38}$	$87.83_{\pm 0.91}$	$96.38_{\pm 0.55}$	$95.29_{\pm 0.13}$	$93.96_{\pm 0.19}$	$\mathbf{98.20}_{\pm 0.11}$
AdONE	$67.68_{\pm 0.48}$	$33.73_{\pm 1.03}$	$80.00_{\pm 1.00}$	$69.68_{\pm 0.22}$	$43.61_{\pm 0.34}$	$91.44_{\pm 0.36}$	$86.54_{\pm 0.16}$	$85.37_{\pm 1.13}$	$73.49_{\pm 1.03}$
GRADATE	$86.99_{\pm 0.74}$	$77.87_{\pm 0.79}$	$91.56_{\pm 0.18}$	$81.71_{\pm 0.81}$	$72.17_{\pm 1.08}$	$93.54_{\pm 1.50}$	$87.44_{\pm 1.08}$	$79.46_{\pm 0.67}$	$97.33_{\pm 0.35}$
GAD-NR	$70.83_{\pm 0.56}$	$46.06_{\pm 0.30}$	$72.24_{\pm 0.36}$	$73.28_{\pm 0.15}$	$49.91_{\pm 0.30}$	$86.04_{\pm 0.39}$	$71.25_{\pm 0.16}$	$49.52_{\pm 0.77}$	$68.81_{\pm 0.82}$
AD-GCL	$\mathbf{92.83}_{\pm 0.43}$	$\mathbf{85.70}_{\pm 1.12}$	$\mathbf{98.67}_{\pm 0.12}$	$\mathbf{94.88}_{\pm 0.27}$	$\mathbf{90.51}_{\pm 0.56}$	$\mathbf{99.71}_{\pm 0.56}$	$\mathbf{95.74}_{\pm 0.10}$	$\mathbf{95.12}_{\pm 0.25}$	$97.88_{\pm 0.05}$

Method	Bitcoinotc			BITotc			BITalpha		
	AUC	TN	HN	AUC	TN	HN	AUC	TN	HN
SCAN	$66.37_{\pm 0.64}$	$51.64_{\pm 0.96}$	$69.64_{\pm 0.02}$	$64.92_{\pm 0.76}$	$47.59_{\pm 0.10}$	$70.02_{\pm 0.91}$	$66.28_{\pm 0.93}$	$49.39_{\pm 0.23}$	$72.30_{\pm 0.02}$
Radar	$47.34_{\pm 0.27}$	$44.61_{\pm 0.17}$	$48.64_{\pm 0.02}$	$49.96_{\pm 0.09}$	$48.44_{\pm 0.11}$	$52.48_{\pm 0.07}$	$53.09_{\pm 0.42}$	$52.91_{\pm 0.92}$	$53.73_{\pm 0.85}$
ONE	$49.71_{\pm 0.72}$	$48.95_{\pm 0.10}$	$50.11_{\pm 0.05}$	$49.55_{\pm 0.04}$	$48.88_{\pm 0.10}$	$50.13_{\pm 0.32}$	$49.64_{\pm 0.47}$	$49.07_{\pm 0.50}$	$50.13_{\pm 0.05}$
AEGIS	$52.79_{\pm 1.06}$	$57.29_{\pm 1.18}$	$50.46_{\pm 1.60}$	$53.64_{\pm 1.24}$	$56.41_{\pm 1.92}$	$52.62_{\pm 1.72}$	$54.21_{\pm 1.11}$	$57.02_{\pm 1.36}$	$51.14_{\pm 1.30}$
GAAN	$58.32_{\pm 1.32}$	$49.80_{\pm 1.88}$	$53.34_{\pm 1.72}$	$54.52_{\pm 1.95}$	$49.75_{\pm 1.37}$	$50.71_{\pm 1.82}$	$55.62_{\pm 1.82}$	$50.39_{\pm 1.85}$	$51.69_{\pm 1.91}$
CoLA	$79.30_{\pm 0.42}$	$65.00_{\pm 0.23}$	$70.13_{\pm 0.99}$	$80.11_{\pm 0.74}$	$66.42_{\pm 0.92}$	$85.41_{\pm 0.83}$	$76.91_{\pm 0.29}$	$66.62_{\pm 0.66}$	$74.28_{\pm 0.99}$
ANEMONE	$79.19_{\pm 1.13}$	$65.00_{\pm 0.95}$	$84.92_{\pm 0.68}$	$80.90_{\pm 0.42}$	$68.37_{\pm 0.25}$	$85.67_{\pm 0.84}$	$75.44_{\pm 0.88}$	$64.44_{\pm 0.27}$	$72.73_{\pm 1.09}$
AdONE	$79.24_{\pm 0.68}$	$68.64_{\pm 0.19}$	$75.72_{\pm 0.41}$	$80.57_{\pm 0.30}$	$\mathbf{72.06}_{\pm 0.86}$	$76.88_{\pm 0.39}$	$76.34_{\pm 0.29}$	$60.52_{\pm 0.30}$	$79.34_{\pm 1.06}$
GRADATE	$73.70_{\pm 0.98}$	$49.38_{\pm 0.21}$	$84.46_{\pm 0.32}$	$74.01_{\pm 0.22}$	$50.93_{\pm 0.88}$	$82.55_{\pm 0.42}$	$69.34_{\pm 0.97}$	$51.89_{\pm 0.65}$	$68.09_{\pm 0.63}$
GAD-NR	$69.82_{\pm 0.29}$	$48.53_{\pm 0.88}$	$61.55_{\pm 0.78}$	$69.78_{\pm 0.85}$	$52.77_{\pm 0.77}$	$60.08_{\pm 0.45}$	$71.76_{\pm 0.86}$	$51.21_{\pm 0.99}$	$67.26_{\pm 0.79}$
AD-GCL	$\mathbf{82.19}_{\pm 0.46}$	$\mathbf{69.18}_{\pm 0.95}$	$\mathbf{87.53}_{\pm 0.53}$	$\mathbf{82.11}_{\pm 0.20}$	$70.39_{\pm 0.48}$	$\mathbf{86.25}_{\pm 0.20}$	$\mathbf{79.62}_{\pm 0.66}$	$\mathbf{67.73}_{\pm 1.04}$	$\mathbf{82.77}_{\pm 1.29}$

Table 1: Performance comparison for AUC (%). TN and HN represent the AUC for tail nodes and head nodes, respectively. Bold indicates the optimal and underline indicates the suboptimal.

anomaly detection performance of the model from the structural imbalance perspective, we adhere to the convention of previous research (Liu et al. 2020; Liu, Nguyen, and Fang 2021) by dividing head and tail nodes based on a predefined degree threshold K and reporting the corresponding results. Please refer to the Appendix for more detailed parameter settings and experimental environment introduction.

Main Results and Analysis

We evaluate the anomaly detection performance of AD-GCL by conducting a comparison with 10 baseline methods. By calculating the area under the ROC curve, the AUC scores of the 6 benchmark datasets are compared as shown in Table 1. More experiments validating the effectiveness of AD-GCL are presented in Appendix, including report results for AUPRC and AP, comparisons of ROC curves, visualization of the distribution of abnormal scores, compared with naive data augmentation methods, the advantages of polynomial distribution sampling, and the sensitivity analysis of the readout function. Based on the results, we have the following observations:

We can intuitively observe that AD-GCL achieves the best anomaly detection performance on these six datasets. It outperforms the runner-up methods by 2.13%, 2.84%, 0.45%, 2.89%, 1.21%, and 2.71%, respectively. Of greater significance, AD-GCL achieves the best AUC scores for both tail nodes and head nodes on most of the datasets. Clustering method SCAN, matrix factorization method Radar and ONE show unsatisfactory performance compared to deep GAD methods. SCAN shows significant differences in the performance of head and tail nodes. Non-deep methods have inherent limitations in dealing with complex graph structures and high-dimensional and sparse features. Meanwhile, in datasets with sparse structures and features, such as Cora and Citeseer, the reconstruction quality of AdONE based on autoencoder methods tends to be poor, leading to completely collapsed solutions and consequently resulting in degraded performance.

Within the category of deep anomaly detection methods, we observe that GCL-based approaches CoLA, ANEMONE, and GRADATE exhibit superior performance. This suggests that the GCL-based paradigm effectively detects anomalies by leveraging the feature and structure information present in the graph. However, existing GCL-based methods focus on how to design contrast pairs of different granularities in graphs, such as node-subgraph or node-node, and they neglect to consider GAD from the perspective of structural imbalance. We first show the performance gap caused by structural imbalance through empirical research, and propose different strategies for imbalanced structures. Compared with existing GCL-based methods in Fig. 1, AD-GCL demonstrates significant performance improvements on nodes with varying degrees in Cora and Citeseer datasets. We further validate that AD-GCL assigns higher anomaly scores to anomaly nodes via violin plots, and the experiments shown in Appendix. Our method yields surprising results by not only enhancing the performance of tail nodes but also improving the performance of head nodes. This highlights the significant role played by the proposed scheme.

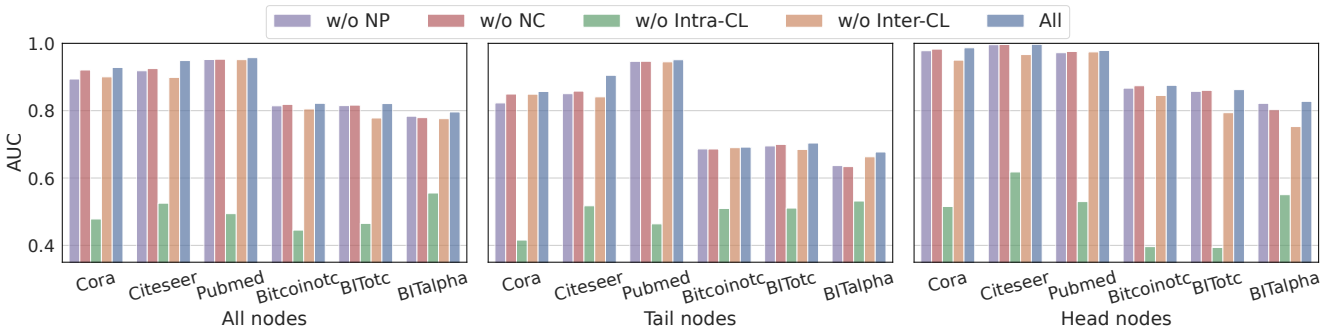
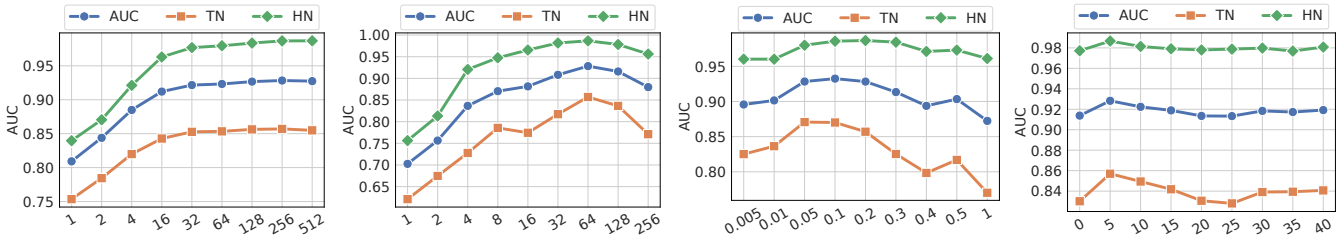


Figure 3: Ablation study on different variants.



(a) Sampling rounds R (b) Dimension d (c) Trade-off parameter α (d) Sliding window w

Figure 4: The parameter study of AD-GCL with varying (a) sampling rounds R, (b) dimension d, (c) trade-off parameter α and (d) sliding window w on Cora dataset w.r.t. AUC respectively.

Ablation Study

To validate the validity of each component in AD-GCL, we investigate the impact of neighbor pruning (NP), anomaly-guided neighbor completion (NC), intra-view contrast loss (Intra-CL), and inter-view contrast loss (Inter-CL) in our method. W/o means removing the component during the training stage. Figure 3 reports the results of our ablation study on six datasets. We can observe that AD-GCL consistently outperforms the other variants in terms of overall performance, as well as tail and head nodes. On the tail node, w/o NP, NC, Intra-CL and Inter-CL cause performance degradation in 2.46%, 1.87%, 30.62%, and 1.89% respectively. The performance is a drastic drop when w/o Intra-CL, which involves learning the latent relationships between nodes in graphs. This observation provides evidence that intra-view contrast plays a dominant role in GCL-based anomaly detection. Meanwhile, we find an improvement in the tail performance of the NP strategy acting on the head node, further proving that aligning head nodes with forged tail nodes enhances the model's ability to learn from genuine tail nodes. In summary, the absence of any component in AD-GCL leads to a degradation in performance, highlighting the importance of each component in the AD-GCL.

Parameter Study

Effect of sampling rounds R We investigate the effect of varying the value of R in Eq. (10), as shown in Figure 4a. The performance is subpar when only using lower sample rounds such as 1, 2, 4, etc. As the number of sampling rounds increases, the AUC score generally increases. However, when R reaches 256, further increases do not lead to significant performance changes.

Effect of hidden layer dimension d We visualize the effect of the hidden layer dimension d in Figure 4b. As the embedding dimension increases from 1 to 64, all three curves exhibit an overall upward trend. However, further increasing d, the performance begins to decline. This phenomenon can be attributed to the fact that the features in the Cora dataset are high-dimensional and sparse. When d becomes too large, it may retain noise information that is not beneficial for anomaly detection tasks.

Effect of trade-off parameter α In Figure 4c, we present the evaluation results for different values of α in Eq. (9). We observe similar conclusions to the ablation experiments. Firstly, increasing α leads to a significant performance improvement, demonstrating the effectiveness of $\mathcal{L}_{inter}$. Additionally, from Figure 3 we can clearly see the importance of $\mathcal{L}_{intra}$ for anomaly detection. However, when α becomes excessively large, $\mathcal{L}_{inter}$ dominates the overall loss, leading to a decline in performance.

Effect of sliding window w We study the sliding window w in the neighbor completion strategy, which incorporates anomaly scores from the previous w training rounds to compute the anomaly similarity distribution. Figure 4d shows the best performance is achieved when $w = 5$, and there is a significant performance drop when $w = 0$, demonstrating the importance of considering anomaly scores for anomaly detection. However, an excessively large w may introduce noise or smooth more accurate anomaly information from the current epoch, leading to performance degradation.

Conclusion

In this paper, we first investigate the limitations of anomaly detection methods based on GCL from the perspective of structural imbalance, revealing that prior works perform poor performance in detecting tail anomalies. To address the above challenges, we propose a new method AD-GCL. It filters the noisy edges of the head node through neighbor pruning and enlarges the neighborhood of the tail node through anomaly-guided neighbor completion. Finally, intra- and inter-view consistency losses of the original and enhanced graphs are introduced to enhance the representation. Extensive experiments on six benchmark datasets demonstrate that AD-GCL outperforms the competitors.

Acknowledgments

This research was partially supported by the National Key Research and Development Project of China No. 2021ZD0110700, the Key Research and Development Project in Shaanxi Province No. 2022GXLH01-03, the National Science Foundation of China No. (62250009, 62037001, 62302380, and 62476215), the China Postdoctoral Science Foundation No. 2023M742789, the Fundamental Scientific Research Funding No. (xzd012023061 and xpt012024003), and the Shaanxi Continuing Higher Education Teaching Reform Research Project No. 21XJZ014. Co-author Chen Chen consulted on this project on unpaid weekends for personal interests, and appreciated collaborators and family for their understanding.

References

Bandyopadhyay, S.; Lokesh, N.; and Murty, M. N. 2019. Outlier aware network embedding for attributed networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 12–19.

Bandyopadhyay, S.; N, L.; Vivek, S. V.; and Murty, M. N. 2020. Outlier resistant unsupervised deep architectures for attributed network embedding. In *Proceedings of the 13th international conference on web search and data mining*, 25–33.

Chen, J.; Zhu, G.; Yuan, C.; and Huang, Y. 2024. Boosting Graph Anomaly Detection with Adaptive Message Passing. In *The Twelfth International Conference on Learning Representations*.

Chen, Z.; Liu, B.; Wang, M.; Dai, P.; Lv, J.; and Bo, L. 2020. Generative adversarial attributed network anomaly detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1989–1992.

Ding, K.; Li, J.; Agarwal, N.; and Liu, H. 2021. Inductive anomaly detection on attributed networks. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, 1288–1294.

Ding, K.; Li, J.; Bhanushali, R.; and Liu, H. 2019. Deep anomaly detection on attributed networks. In *SDM*, 594–602. SIAM.

Dou, Y.; Liu, Z.; Sun, L.; Deng, Y.; Peng, H.; and Yu, P. S. 2020. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 315–324.

Duan, J.; Wang, S.; Zhang, P.; Zhu, E.; Hu, J.; Jin, H.; Liu, Y.; and Dong, Z. 2023. Graph anomaly detection via multi-scale contrastive learning networks with augmented view. In *AAAI*, volume 37, 7459–7467.

Goga, O.; Venkatadri, G.; and Gummadi, K. P. 2015. The doppelgänger bot attack: Exploring identity impersonation in online social networks. In *Proceedings of the 2015 internet measurement conference*, 141–153.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 9729–9738.

Hooi, B.; Shin, K.; Song, H. A.; Beutel, A.; Shah, N.; and Faloutsos, C. 2017. Graph-based fraud detection in the face of camouflage. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11(4): 1–26.

Hu, J.; Xiao, B.; Jin, H.; Duan, J.; Wang, S.; Lv, Z.; Wang, S.; Liu, X.; and Zhu, E. 2023. SAMCL: Subgraph-Aligned Multiview Contrastive Learning for Graph Anomaly Detection. *IEEE Transactions on Neural Networks and Learning Systems*.

Jin, M.; Liu, Y.; Zheng, Y.; Chi, L.; Li, Y.-F.; and Pan, S. 2021. Anemone: Graph anomaly detection with multi-scale contrastive learning. In *CIKM*, 3122–3126.

Kumar, S.; Hooi, B.; Makhija, D.; Kumar, M.; Faloutsos, C.; and Subrahmanian, V. 2018. Rev2: Fraudulent user prediction in rating platforms. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 333–341.

Kumar, S.; Spezzano, F.; Subrahmanian, V.; and Faloutsos, C. 2016. Edge weight prediction in weighted signed networks. In *ICDM*, 221–230. IEEE.

Li, J.; Dani, H.; Hu, X.; and Liu, H. 2017. Radar: Residual analysis for anomaly detection in attributed networks. In *IJCAI*, volume 17, 2152–2158.

Liao, J.; Li, J.; Chen, L.; Wu, B.; Bian, Y.; and Zheng, Z. 2023. SAILOR: Structural Augmentation Based Tail Node Representation Learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 1389–1399.

Liu, Y.; Ao, X.; Qin, Z.; Chi, J.; Feng, J.; Yang, H.; and He, Q. 2021a. Pick and choose: a GNN-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021*, 3168–3177.

Liu, Y.; Li, Z.; Pan, S.; Gong, C.; Zhou, C.; and Karypis, G. 2021b. Anomaly detection on attributed networks via contrastive self-supervised learning. *TNNLS*, 33(6): 2378–2392.

Liu, Z.; Li, Y.; Chen, N.; Wang, Q.; Hooi, B.; and He, B. 2023. A survey of imbalanced learning on graphs: Problems, techniques, and future directions. *arXiv preprint arXiv:2308.13821*.

Liu, Z.; Nguyen, T.-K.; and Fang, Y. 2021. Tail-gnn: Tail-node graph neural networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 1109–1119.

Liu, Z.; Zhang, W.; Fang, Y.; Zhang, X.; and Hoi, S. C. 2020. Towards locality-aware meta-learning of tail node embeddings on networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 975–984.

Ma, X.; Wu, J.; Xue, S.; Yang, J.; Zhou, C.; Sheng, Q. Z.; Xiong, H.; and Akoglu, L. 2021. A comprehensive survey on graph anomaly detection with deep learning. *TKDE*.

Pacini, C.; Hopwood, W.; Young, G.; and Crain, J. 2018. The role of shell entities in fraud and other financial crimes. *Managerial Auditing Journal*, 34(3): 247–267.

Park, J.; Song, J.; and Yang, E. 2022. Graphens: Neighbor-aware ego network synthesis for class-imbalanced node classification. In *The Tenth International Conference on Learning Representations, ICLR 2022*. International Conference on Learning Representations (ICLR).

Pazho, A. D.; Noghre, G. A.; Purkayastha, A. A.; Vempati, J.; Martin, O.; and Tabkhi, H. 2023. A Survey of Graph-based Deep Learning for Anomaly Detection in Distributed Systems. *IEEE Transactions on Knowledge and Data Engineering*.

Roy, A.; Shu, J.; Li, J.; Yang, C.; Elshocht, O.; Smeets, J.; and Li, P. 2023. GAD-NR: Graph Anomaly Detection via Neighborhood Reconstruction. *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM)*.

Tang, X.; Yao, H.; Sun, Y.; Wang, Y.; Tang, J.; Aggarwal, C.; Mitra, P.; and Wang, S. 2020. Investigating and mitigating degree-related biases in graph convoltuional networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1435–1444.

Tong, H.; Faloutsos, C.; and Pan, J.-Y. 2006. Fast random walk with restart and its applications. In *Sixth international conference on data mining (ICDM'06)*, 613–622. IEEE.

Van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv e-prints*, arXiv–1807.

Veličković, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep graph infomax. *ICLR*.

Wang, R.; Wang, X.; Shi, C.; and Song, L. 2022. Uncovering the Structural Fairness in Graph Contrastive Learning. *Advances in Neural Information Processing Systems*, 35: 32465–32473.

Wu, J.; He, J.; and Xu, J. 2019. Net: Degree-specific graph neural networks for node and graph classification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 406–415.

Wu, T.; Ren, H.; Li, P.; and Leskovec, J. 2020. Graph information bottleneck. *Advances in Neural Information Processing Systems*, 33: 20437–20448.

Xu, H.; Xiang, L.; Huang, F.; Weng, Y.; Xu, R.; Wang, X.; and Zhou, C. 2023a. Grace: Graph Self-Distillation and Completion to Mitigate Degree-Related Biases. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2813–2824.

Xu, X.; Yuruk, N.; Feng, Z.; and Schweiger, T. A. 2007. Scan: a structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 824–833.

Xu, Y.; Peng, Z.; Shi, B.; Hua, X.; and Dong, B. 2024. Learning dynamic graph representations through timespan view contrasts. *Neural Networks*, 176: 106384.

Xu, Y.; Shi, B.; Ma, T.; Dong, B.; Zhou, H.; and Zheng, Q. 2023b. CLDG: Contrastive Learning on Dynamic Graphs. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 696–707. IEEE.

Ye, G.; Yin, H.; Chen, T.; Chen, H.; Cui, L.; and Zhang, X. 2021. FENet: a frequency extraction network for obstructive sleep apnea detection. *IEEE Journal of Biomedical and Health Informatics*, 25(8): 2848–2856.

Zeng, L.; Li, L.; Gao, Z.; Zhao, P.; and Li, J. 2023. Imgcl: Revisiting graph contrastive learning on imbalanced node classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 11138–11146.

Zhang, J.; and et al. 2022. Reconstruction enhanced multi-view contrastive learning for anomaly detection on attributed networks. *IJCAI*.

Zheng, Q.; Xu, Y.; Liu, H.; Shi, B.; Wang, J.; and Dong, B. 2023. A Survey of Tax Risk Detection Using Data Mining Techniques. *Engineering*.