FedCIGAR: A Personalized Reconstruction Approach for Federated Graph-Level Anomaly Detection

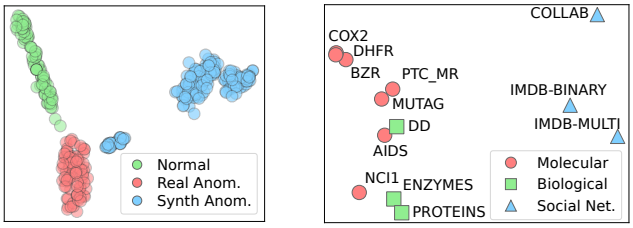
Yunfeng Zhao[1], **Yixin Liu**[2], **Qingfeng Chen**[1*], **Shiyuan Li**[2], **Yue Tan**[2] and **Shirui Pan**[2]
[1]Guangxi University
[2]Griffith University
{yunf.zhao, qingfeng}@gxu.edu.cn,
{yixin.liu, shiyuan.li, yue.tan, s.pan}@griffith.edu.au

Abstract

Graph-level anomaly detection (GLAD) is crucial for ensuring the reliability of graph-driven applications by identifying abnormal graphs that deviate from the majority. Considering the privacy concerns in distributed scenarios, federated graph-level anomaly detection (FedGLAD) has emerged as a promising solution to enable collaborative detection without sharing raw data. However, existing methods suffer from poor generalization due to the reliance on unrealistic synthetic anomalies and insufficient personalization capabilities under data heterogeneity. To address these challenges, we propose a novel **Fed**erated graph-level anomaly detection approach with **C**luster-adapt**I**ve **GA**ted **R**econstruction (FedCIGAR). Specifically, we design a reconstruction-based paradigm trained on normal graphs to avoid synthetic data. Furthermore, we introduce a client-side node contribution gating mechanism and a server-side sliding window-based clustering strategy to tackle data heterogeneity. Extensive experiments demonstrate that FedCIGAR achieves superior performance and robustness in contrast to state-of-the-art methods.

1 Introduction

Graph-structured data have been widely used in a variety of real-world scenarios, such as bioinformatics, small molecules, and social networks [Stokes *et al.*, 2020; Barabasi and Oltvai, 2004; Easley *et al.*, 2010]. To ensure the reliability and robustness of graph-driven applications, graph-level anomaly detection (GLAD) is a crucial task on graph data [Zhao and Akoglu, 2023; Luo *et al.*, 2022; Qiu *et al.*, 2022], which aims to identify the abnormal graphs that are different from the majority. Recently, most GLAD methods [Kim *et al.*, 2024; Zhang *et al.*, 2022; Ma *et al.*, 2023; Liu *et al.*, 2023] follow a centralized training principle, which requires collecting all graph data for model training. However, in real-world scenarios, graph data are usually distributed among different users. Owing to privacy protection concerns [Sarasamma *et al.*, 2005], users are often unwilling to share their data, which greatly hinders the application of centralized GLAD methods in practical scenarios.



(a) Synthetic abnormal (b) Client data distributions

Figure 1: (a): T-SNE visualization of normal graphs, synthetic anomalies, and real anomalies learned by LG-FGAD on AIDS dataset. (b): Heterogeneous client data distributions across different datasets.

As an effective distributed learning paradigm, federated learning (FL) [Huang *et al.*, 2023; Fu *et al.*, 2022; Li *et al.*, 2021] provides a promising pathway to address the above limitation. Recently, an emerging branch of studies has attempted to incorporate FL and GLAD to mitigate the impact of centralized training and facilitate anomaly detection in distributed environments. These federated graph-level anomaly detection (FedGLAD) methods perform federated optimization on specially designed anomaly detection models to enable effective GLAD under distributed data settings. For instance, FGAD [Cai *et al.*, 2024b] employs a teacher-student architecture with knowledge distillation to maintain the personalization of local models and enable collaborative learning among clients. To improve anomaly detection in the absence of annotated anomalous graphs, it uses self-generated anomalies to simulate realistic anomalous cases for model training. LG-FGAD [Cai *et al.*, 2024a] indicates that node-level representations may induce potential anomalies and proposes a local-global anomaly awareness module, which captures anomalous patterns at both node and graph levels to enhance the discrimination ability.

Despite their effectiveness, the existing FedGLAD methods treat anomaly detection as a binary classification task and employ artificially generated anomalies to mimic abnormal samples, which leads to their ***Limitation 1*** *- poor generalization capacity to realistic and heterogeneous anomaly cases*. Considering the sparsity of anomalous graph data, current

*Corresponding author.

GLAD methods [Ma *et al.*, 2022; Liu *et al.*, 2023] usually adopt a one-class training setting [Yin *et al.*, 2024], where abnormal samples are unavailable for model training. Under this constraint, to train the binary classification-based detection models without annotated samples, the existing FedGLAD methods use an additional generative model to synthesize pseudo-anomalous samples. Nevertheless, these synthetic anomalies may fail to reflect the true distribution of real-world abnormal graphs, which hinders their ability to capture diverse and realistic anomaly characteristics. As visualized in Fig. 1(a), learned by LG-FGAD [Cai *et al.*, 2024a], the representations of synthetic anomalies significantly deviate from those of real anomalies, indicating the inherent limitation of such a data synthesis paradigm.

In addition to the limited distribution modeling capability of local GLAD models, the FedAvg-style global aggregation paradigm in existing FedGLAD methods further intensifies ***Limitation 2*** *– insufficient personalization capability under heterogeneous client data*. As demonstrated in Fig. 1(b), client data in real-world scenarios usually exhibits heterogeneous distributions (see Appendix A for details), which makes it challenging for one global model to capture all diverse anomaly patterns [Xie *et al.*, 2021]. However, current FedGLAD methods adopt a FedAvg-like paradigm [McMahan *et al.*, 2017] that globally updates a single server model for all clients, which results in limited adaptability to heterogeneous client data. Even if the knowledge distillation-based collaborative learning mechanism in FGAD and LG-FGAD can preserve the personalization to some extent, the globally shared student model may still fail to adapt well to the diverse anomaly distributions across clients.

To tackle the aforementioned limitations, in this paper, we propose **FedCIGAR**, a **Fed**erated graph-level anomaly detection approach with **C**luster-adapt**I**ve **GA**ted **R**econstruction. To address ***Limitation 1***, we design a reconstruction-based GLAD model that can be directly trained on normal graph data without relying on synthetic anomalies. This formulation eliminates the need for costly data generation and avoids the adverse impact of low-quality pseudo-anomalous samples. To better capture structural and attributive anomaly patterns, we propose a *feature–structure joint reconstruction model* for GLAD that independently encodes structural and feature information into a unified latent space and then reconstructs the original features and graph structures accordingly. To handle ***Limitation 2***, we enhance the personalization capability of both the client-side detection model and the server-side aggregation mechanism. At the client side, we introduce a *node contribution gating mechanism* that equips each client with an independent local model to assess the contribution of different nodes to anomaly scores. This mechanism alleviates the limitation of global reconstruction models under heterogeneous client data and preserves client-specific anomaly characteristics. At the server side, we introduce a *sliding window–based clustering aggregation strategy* that dynamically groups clients with similar data distributions and aggregates their updates in a cluster-adaptive manner. In this way, FedCIGAR effectively balances global collaboration and local personalization under heterogeneous client data. In summary, this paper makes the following contributions:

- **New FedGLAD Paradigm.** We, for the first time, introduce a reconstruction-based paradigm for FedGLAD. Compared to classification-based methods, our approach avoids synthetic anomaly generation and achieves better generalization to different anomaly patterns.
- **Data Heterogeneity Handling.** We tackle the data heterogeneity across clients from both server and client perspectives by crafting a node contribution-based model and a cluster-adaptive server aggregation mechanism.
- **Superior Detection Performance.** We conduct extensive experiments on multiple datasets and multi-dataset settings, where FedCIGAR achieves excellent detection performance and robustness.

2 Preliminaries

Notations. Let $\mathcal{D} = \{D_1, \ldots, D_C\}$ represent the federated datasets distributed in C clients, where $\mathcal{D}_c = \{\mathcal{G}_1, \ldots, \mathcal{G}_{N_c}\}$ represents the local dataset held by client c and contains N_c graphs. Each graph $\mathcal{G}_i \in \mathcal{D}$ has a node set $\mathcal{V}_i$ and an edge set $\mathcal{E}_i$. The connection between nodes of graph $\mathcal{G}_i$ can be represented by an adjacency matrix $\mathbf{A}_i \in \{0,1\}^{n_i \times n_i}$, where $n_i = |\mathcal{V}_i|$ is the nodes number in graph $\mathcal{G}_i$. The features matrix of graph $\mathcal{G}_i$ can be described by $\mathbf{X}_i \in \mathbb{R}^{n_i \times m}$ where the j-th row $\mathbf{x}_{i,j} \in \mathbb{R}^d$ represents the feature vector for node $v_j \in \mathcal{V}_i$.
Problem Formulation. This paper concentrates on the one-class (a.k.a. unsupervised) federated graph-level anomaly detection (FedGLAD) problem. Given a federated dataset $\mathcal{D}$, where each client possesses its own graph set $\mathcal{D}_c$, and it can be divided into anomalous graph set $\mathcal{D}_a$ and normal graph set $\mathcal{D}_n$, with $\mathcal{D}_c = \mathcal{D}_a \cup \mathcal{D}_n$, $\mathcal{D}_a \cap \mathcal{D}_n = \emptyset$, and $|\mathcal{D}_a| = N_a \ll |\mathcal{D}_n| = N_n$. Formally, the objective of FedGLAD is to learn a set of scoring functions $f_{\mathcal{W}} = \{f_{\mathcal{W}^{(1)}}, \ldots, f_{\mathcal{W}^{(C)}}\}$ (where $\mathcal{W}^{(c)}$ is the set of learnable parameters of client c), which assign anomaly scores to indicate the abnormality of graphs from each client. The overall training objective can be regarded as minimizing the following loss function:

$$\min_{\mathcal{W}^{(1)},\ldots,\mathcal{W}^{(C)}} \frac{1}{C}\sum_{c=1}^{C}\frac{|D_c|}{|D|}\big(\ell_c(y_n, f_{\mathcal{W}^{(c)}}(\mathcal{G}_n)) + \ell_c(y_a, f_{\mathcal{W}^{(c)}}(\mathcal{G}_a))\big), \tag{1}$$

where $\{\mathcal{G}_n; y_n\}$ represents the normal graph labeled with y=1, and $\{\mathcal{G}_a; y_a\}$ represents the anomalous graph labeled with y=0. Here, $f_{\mathcal{W}^{(c)}}(\cdot)$ and $\ell_c(\cdot)$ separately denote the c-th client scoring function and loss function.

3 Methodology

In this section, we introduce FedCIGAR, a novel reconstruction-based method for federated graph-level anomaly detection (FedGLAD). The overall pipeline of FedCIGAR is illustrated in Fig. 2. To capture diverse anomaly patterns without relying on real anomalous samples, we introduce a *feature-structure joint reconstruction model* (Sec. 3.1), which explicitly captures the attributive and structural information of the graph and reconstructs them from a fused latent

[1]Code is available at https://github.com/yunf-zhao/FedCIGAR

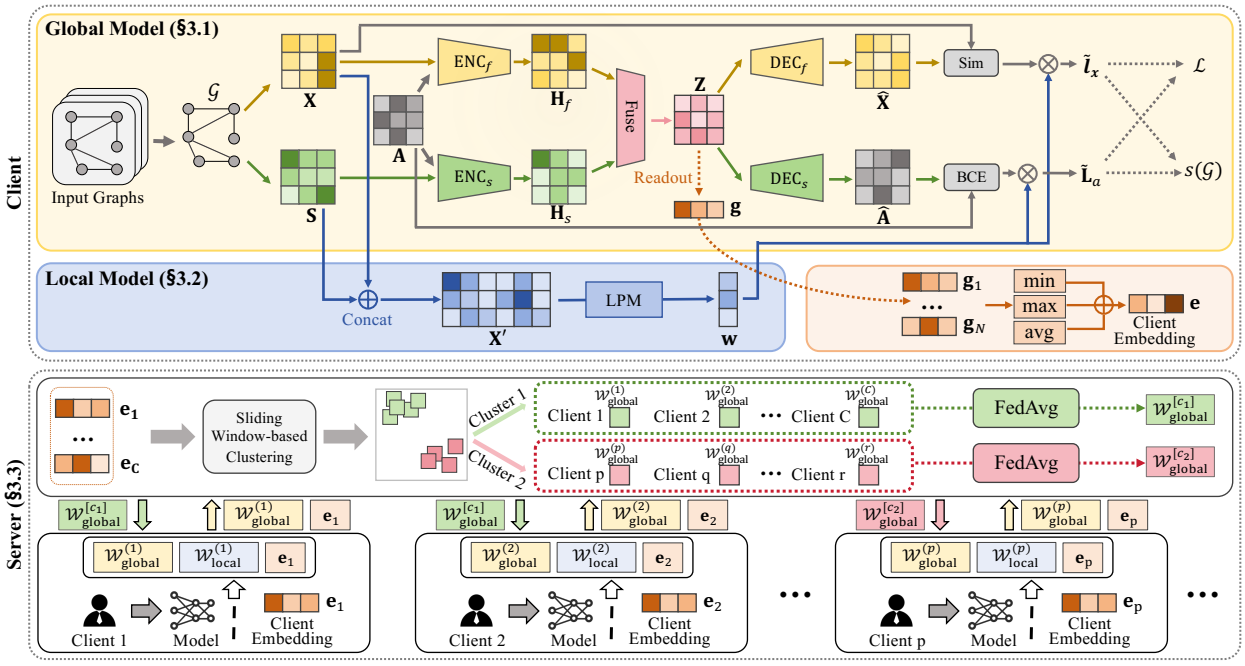


Figure 2: The overall pipeline of FedCIGAR.

space. Once well trained, the reconstruction error can serve as an effective indicator of graph abnormality. To alleviate the impact of data heterogeneity across clients, we design a *node contribution gating module* (Sec. 3.2) to gauge each node's contribution to graph anomaly under personalization, which allows each client to maintain adaptive anomaly scoring tailored to its local data distribution. Furthermore, we introduce a *sliding window-based clustering aggregation strategy* (Sec. 3.3) at the server side, which adaptively groups clients with similar behaviors and performs cluster-wise aggregation to better accommodate heterogeneous client data.

3.1 Feature-Structure Joint Reconstruction

In order to train the binary classification-based federated anomaly detection models using only normal data, the existing FedGLAD methods (e.g., FGAD [Cai *et al.*, 2024b] and LG-FGAD [Cai *et al.*, 2024a]) adopt data generation techniques to synthesize pseudo-anomalous samples for supervision. However, such generated anomalies usually fail to reflect real-world anomaly patterns, leading to suboptimal detection performance. To address the limitations caused by relying on synthetic anomalies, FedCIGAR abandons the binary classification paradigm and instead employs a reconstruction-based FedGLAD model. The core idea of our detection model is to separate normal and anomalous graph patterns based on the reconstruction assumption, i.e., the reconstruction model trained on purely normal data can accurately rebuild normal graphs while producing large reconstruction errors for anomalous graphs [Kim *et al.*, 2024]. In order to effectively capture both structural and attributive patterns of graphs, we carefully design a feature–structure joint reconstruction model for GLAD, which is detailed in the following paragraphs. For simplicity, we only describe the reconstruction process of an individual graph $\mathcal{G}$ and omit its index (e.g., i) in the notation.

Dual-Channel Encoder. The existing reconstruction models for GLAD [Kim *et al.*, 2024; Liu *et al.*, 2023] typically use GNN-based autoencoder to aggregate *raw node features* for graph reconstruction. However, the low-dimensional features of the real-world graph datasets (e.g., atom types) may provide limited semantic information. In some specific domains (e.g., social networks), node features are even absent, which provides limited informative resources for modeling data patterns. To mitigate the impact of poor-quality features and explicitly capture structural patterns, we introduce structure encoding as an additional input of our reconstruction model. Specifically, for a graph $\mathcal{G}$, its structure encoding matrix is denoted as $\mathbf{S} \in \mathbb{R}^{n \times d_s} = \text{concat}[\mathbf{S}^{\text{DSE}}, \mathbf{S}^{\text{RWSE}}]$, where each row $\mathbf{s}_j$ is the concatenation of $\mathbf{s}_j^{\text{DSE}}$, the one-hot encoding of node v_j's degree, and $\mathbf{s}_j^{\text{RWSE}}$, the random walk diffusion-based structure encoding [Dwivedi *et al.*, 2022] of node v_j. Taking the structural encoding as an auxiliary input not only enhances the informativeness of the original features but also injects structural knowledge from both local (i.e., node degree) and global (i.e., random-walk characteristics) perspectives.

After obtaining the structural encoding, we employ two parallel channels to learn feature and structure representations:

$$\mathbf{H}_f = \text{ENC}_f(\mathbf{X}, \mathbf{A}), \quad \mathbf{H}_s = \text{ENC}_s(\mathbf{S}, \mathbf{A}), \tag{2}$$

where $\text{ENC}_f(\cdot)$ and $\text{ENC}_s(\cdot)$ denote the GNN-based feature encoder and structure encoder, respectively, while $\mathbf{H}_f$ and $\mathbf{H}_s$ respectively denote the feature and structure representations of graph $\mathcal{G}$. Such a decoupled design allows the model to learn complementary representations, capturing attribute semantics and structural patterns independently.

Fused Decoder. After obtaining the feature and structure rep-

resentations, we fuse them into a shared latent space, where the complementary information from both domains is integrated to form a unified representation for downstream reconstruction, i.e., $\mathbf{Z} = \mathrm{MLP}\Big(\mathrm{concat}\left[\mathbf{H}_f, \mathbf{H}_s\right]\Big)$. Taking $\mathbf{Z}$ as input, we use a GNN-based feature decoder $\mathrm{DEC}_f(\cdot)$ and a dot-product-based structure decoder $\mathrm{DEC}_s(\cdot)$ to reconstruct the raw feature matrix and adjacency matrix, respectively:

$$\hat{\mathbf{X}} = \mathrm{DEC}_f(\mathbf{Z}), \quad \hat{\mathbf{A}} = \mathrm{DEC}_s(\mathbf{Z}), \tag{3}$$

where $\hat{\mathbf{X}}$ denotes the reconstructed feature matrix and $\hat{\mathbf{A}}$ denotes the reconstructed adjacency matrix.

Statistics-Aware Reconstruction. Although minimizing the average reconstruction loss over all nodes in a graph can facilitate model optimization, recent studies [Kim *et al.*, 2024] pinpoint that such a strategy may suffer from the "reconstruction flip" phenomenon, i.e., the model yields even lower reconstruction errors in particular nodes on unseen test graphs than on training graphs. As a result, the reconstruction error may fail to faithfully indicate the abnormality of test graphs under federated settings. To bridge the gap, we introduce a statistics-aware reconstruction strategy for our GLAD model, which not only incorporates the mean reconstruction loss but also considers the standard deviation of reconstruction errors. For model training, the loss function can be written by:

$$\mathbf{l}_x = \left[1 - \frac{\mathbf{X}_{i,:}^{\top}\hat{\mathbf{X}}_{i,:}}{\|\mathbf{X}_{i,:}\|_2 \cdot \|\hat{\mathbf{X}}_{i,:}\|_2}\right]_{i\in[|\mathcal{V}|]} \in \mathbb{R}^{|\mathcal{V}|}, \tag{4}$$

$$\mathbf{L}_a = \Big[\mathbf{A}_{i,j}\log\hat{\mathbf{A}}_{i,j} + (1-\mathbf{A}_{i,j})\log(1-\hat{\mathbf{A}}_{i,j})\Big]_{i,j\in[|\mathcal{V}|]} \in \mathbb{R}^{|\mathcal{V}|^2}, \tag{5}$$

$$\mathcal{L} = \alpha\big(\mathrm{avg}(\mathbf{l}_x) + \beta\ \mathrm{std}(\mathbf{l}_x)\big) + (1-\alpha)\big(\mathrm{avg}(\mathbf{L}_a) + \beta\ \mathrm{std}(\mathbf{L}_a)\big), \tag{6}$$

where $||\cdot||$ represents the $l2$ norm, $\hat{\mathbf{X}}_{i,:}$ stands for i-th node's reconstruction feature and $\hat{\mathbf{A}}_{i,j}$ represents the reconstructed adjacency entry between node i and j, while α and β are the balance hyperparameters. When testing, the anomaly score s of graph $\mathcal{G}$ can be directly calculated by the mean and standard deviation of the reconstruction error, i.e., $s(\mathcal{G}) = \mathcal{L}$.

Through additionally modeling the standard deviation of reconstruction errors, the model can encourage stable and uniform reconstruction errors on all nodes in a graph, thereby alleviating the reconstruction flip issue by suppressing excessively high or low reconstruction errors. Meanwhile, variance modeling enables FedCIGAR to better capture the distributional characteristics of reconstruction errors, which leads to more reliable anomaly characterization.

3.2 Node Contribution Gating

By collaboratively optimizing a global anomaly detection model in a federated manner without sharing raw data, clients can benefit from collective knowledge while preserving data privacy. Although such collaborative learning can train a global GLAD model to capture broader data patterns, it may overlook client-specific anomaly characteristics under heterogeneous data distributions. On one client, anomalies may predominantly stem from a small subset of structurally influential nodes (e.g., popular users in social networks), whereas on another client, anomalies may arise from nodes with specific features (e.g., certain types of atoms in molecular graphs). In this case, a global reconstruction model may be insufficient to capture the heterogeneous anomaly behaviors exhibited by different clients, leading to sub-optimal performance.

To address this issue, we introduce a node contribution gating module as the personalized model for each client. Specifically, a local personalized model is employed to estimate node-wise contribution weights conditioned on the local data distribution. Unlike the reconstruction model that is globally shared, the parameters of the local personalized model are maintained independently on each client, enabling client-specific adaptation to heterogeneous anomaly patterns. For each client, the input of the local personalized model is constructed by concatenating the feature and structure encoding matrices, i.e., $\mathbf{X}' = \mathrm{concat}(\mathbf{X}, \mathbf{S})$. Afterwards, the GNN-based local personalized model $\mathrm{LPM}(\cdot)$ is built to estimate the contribution of each node within its local context:

$$\mathbf{w} = \mathrm{sigmoid}\big(\mathrm{LPM}\big(\mathbf{X}', \mathbf{A}\big)/\tau\big) \in \mathbb{R}^n, \tag{7}$$

where $\mathbf{w}$ denotes the contribution weight vector (in which each element indicates the contribution of a corresponding node) and τ is the temperature hyperparameter to adjust the smoothness of the node-wise weights.

Given the column weight vector $\mathbf{w}$, the reconstruction errors in Eqs. (4) and (5) can be further reweighted as:

$$\tilde{\mathbf{l}}_x = \mathbf{l}_x \odot \mathbf{w}, \quad \tilde{\mathbf{L}}_a = \mathbf{L}_a \odot (\mathbf{w}\mathbf{w}^T), \tag{8}$$

where $\odot$ indicates element-wise product. The reweighted errors, i.e., $\tilde{\mathbf{l}}_x$ and $\tilde{\mathbf{L}}_a$, can be further used to calculate the loss function and anomaly score, replacing the original reconstruction errors $\mathbf{l}_x$ and $\mathbf{L}_a$.

According to such a gating mechanism, the local model can learn to adaptively weight node contribution according to the local data patterns. For instance, the atoms that are closely associated with abnormal behaviors in a client with molecular data can be assigned higher weights. In this manner, the model realizes personalized anomaly detection on each client while continuing to benefit from the powerful global knowledge learned by the GLAD model at server.

3.3 Sliding Window-based Clustering Aggregation

While the above gating mechanism can alleviate data heterogeneity via introducing personalized local models, the client-level heterogeneity across different data distributions can still persist if we rely on a single global reconstruction model. For instance, the reconstruction process for molecular graphs can be significantly different from that for social network graphs, and using one unified global model may fail to adapt to such fundamentally diverse data characteristics. A promising solution to this issue is clustered FL, which groups clients with similar data distributions and performs cluster-wise model aggregation. However, existing clustered FL algorithms typically rely on high-dimensional variables for communication, such as model parameters [Wang *et al.*, 2024] or gradients [Xie *et al.*, 2021], and thus often converge slowly due to their top-down bi-partitioning mechanisms, which inevitably incur high communication costs. Although FLT [Jamali-Rad *et al.*, 2022]

acknowledged the above issue, its static clustering strategy, which performs clustering only once at model initialization, may fail to capture changes in dynamic cluster relations during multi-round communication.

To address the above issues, we propose a sliding window-based clustering aggregation strategy for FedGLAD. Our strategy consists of two steps: client representation construction and dynamic clustering. In the first step, we introduce a low-dimensional and high-density client representation to compactly characterize clients' behavioral patterns and reduce communication overhead. Then, according to these representations, we dynamically cluster clients with similar characteristics for aggregation.

Client Representation Construction. To describe the characteristics of each client, we compute statistics over its graph representations to construct a client representation. Based on the node-level representations $\mathbf{Z}$ learned from the reconstruction model, which capture both feature and structure information, we apply a readout function to obtain the graph-level representation $\mathbf{g}$ for graph $\mathcal{G}$. Further, for each client c, we collect the graph representations of all training samples, denoted as: $\mathbb{G}_c = \{\mathbf{g}_i \mid i = 1, \ldots, N_c\}$. Then, the embedding of client c is calculated by statistically aggregating its graph representations:

$$\mathbf{e}_c = [\text{mean}(\mathbb{G}_c) \| \max(\mathbb{G}_c) \| \min(\mathbb{G}_c)], \tag{9}$$

where $\mathbf{e}_c$ is the embedding of client c.

Dynamic Clustering-based Aggregation. After obtaining the low-dimensional client representations, we design the server aggregation strategy based on them. Compared to binary clustering [Xie *et al.*, 2021], k-means optimizes a global partition across all clients, with greater robustness and stability [Ghosh *et al.*, 2020]. At communication round t, the server performs k-means on the stacked client representations $\mathbf{E} = [\mathbf{e}_1; \ldots; \mathbf{e}_C]$ to obtain the clustering assignment $\mathcal{A}_t = \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ where K denotes the number of clusters. Unfortunately, k-means still requires periodic updates to accommodate evolving client representations, incurring considerable computational costs on the server side.

To mitigate re-clustering issue, we design a two-phase clustering module with update and silent phases. The server maintains a sliding window $\mathcal{Q} = \mathcal{A}_1, \ldots, \mathcal{A}_L$ to track historical clusterings, where L is the maximum length. At communication round t, the server first establishes whether $\mathcal{Q}$ is full. If not, the module enters the update phase: the server computes the current clustering $\mathcal{A}_t$ via k-means and evaluates clustering stability by calculating the minimum Adjusted Rand Index (ARI) with historical clusterings:

$$R_t = \min_{\mathcal{A}_i \in \mathcal{Q}} ARI(\mathcal{A}_t, \mathcal{A}_i). \tag{10}$$

If $R_t < \theta$, where θ is a predefined stability threshold, the current clustering deviates substantially from historical patterns, and the server clears $\mathcal{Q}$ and re-initializes it with $\mathcal{A}_t$. Otherwise, the server appends $\mathcal{A}_t$ to $\mathcal{Q}$. Once $\mathcal{Q}$ is full, the module enters the silent phase: the server clears $\mathcal{Q}$ and directly reuses the final stable clustering $\mathcal{A}_L$ for the subsequent P rounds without executing k-means, thereby substantially reducing server-side computational overhead while preserving stable client groupings. The overall training and testing algorithms and complexity analysis are provided in Appendices B.1 and B.2, respectively.

4 Experiments

4.1 Experimental Setup

Datasets. Following the setting in Cai et al. [2024a], we validate the performance of our method on two types of datasets. (1) The single-dataset setting indicates IID scenarios, where a single dataset is randomly distributed across multiple clients, each of which holds a subset of the dataset. We employ four graph classification datasets, including AIDS, DD, MUTAG, and IMDB-BINARY. (2) The multi-dataset setting indicates non-IID scenarios, where multiple datasets consist of several single datasets. Each client holds a specific graph dataset. We employ four multi-dataset collections, including MOLECULES, BIOCHEM, SMALL, and Mix. The first class in each dataset is treated as anomalous, and more dataset details are illustrated in Appendix C.1.

Baselines. We compare FedCIGAR with seven baselines including (1) Self-train, where each client trains locally without any communications; (2) FedAvg [McMahan *et al.*, 2017] and (3) FedProx [Li *et al.*, 2020], the widely used federated learning methods; (4) GCFL [Xie *et al.*, 2021] and (5) FedStar [Tan *et al.*, 2023], two federated graph learning methods; (6) FGAD [Cai *et al.*, 2024b] and (7) LG-FGAD [Cai *et al.*, 2024a], two state-of-the-art FedGLAD methods. Part of the results are borrowed from [Cai *et al.*, 2024b; Cai *et al.*, 2024a]. For datasets not included in these papers, we use random search to select the hyperparameters.

Evaluation Metrics and Implementation. Following the existing FedGLAD method [Cai *et al.*, 2024a], we consider two evaluation metrics: Area Under the Curve (AUC) and F1-Score. We report the average score across 10 trials. More implementation details are given in Appendix C.2.

4.2 Performance Comparison

The comparison results on four single-dataset and four multi-dataset are reported in Table 1 and Table 2. From these results, we have the following observations. ❶ FedCIGAR demonstrates superior performance compared to state-of-the-art approaches across four single-dataset, with the sole exception of a slight reduction in AUC on the AIDS dataset(0.24%). In most cases, it achieves better results for both AUC and F1-Score compared to the best baseline (e.g., an increase of ↑10.02% in AUC and ↑6.67% in F1-Score on the MUTAG dataset). These results demonstrate the effectiveness of our framework in the single-dataset setting. ❷ In the multi-dataset scenarios, client data from various domains, where non-IID problem is further aggravating, FedCIGAR remains significantly superior to other methods. For instance, on BIOCHEM, FedCIGAR outperforms the strongest baseline, LG-FGAD, by 4.22% in AUC and 3.51% in F1-Score. These results suggest that FedCIGAR effectively captures shared information across cross-domain clients, enhancing detection performance.

4.3 Ablation Study

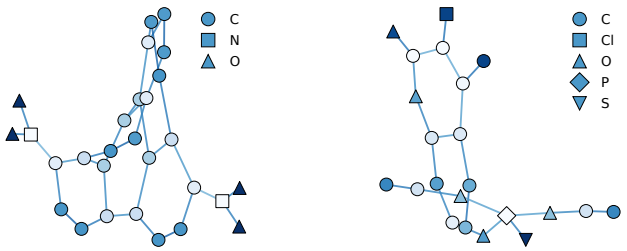
To verify the effectiveness of each component in FedCIGAR, we conduct an ablation study to compare FedCIGAR with

Methods	IMDB-BINARY		MUTAG		DD		AIDS	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
Self-train	65.01 ± 1.53	60.83 ± 1.55	94.44 ± 2.08	89.58 ± 3.61	80.11 ± 0.12	74.52 ± 0.31	98.35 ± 0.16	97.89 ± 0.0
FedAvg	40.96 ± 3.44	45.95 ± 1.66	85.83 ± 2.36	85.00 ± 0.00	30.83 ± 0.44	34.16 ± 1.52	61.26 ± 0.54	58.80 ± 0.60
FedProx	41.86 ± 0.32	45.70 ± 1.96	84.17 ± 2.36	85.00 ± 0.00	30.81 ± 0.43	33.91 ± 1.45	61.18 ± 0.37	58.80 ± 0.60
GCFL	56.98 ± 5.56	45.44 ± 2.25	83.33 ± 3.11	78.33 ± 4.71	38.56 ± 0.13	30.06 ± 1.10	74.31 ± 0.91	70.86 ± 1.76
FedStar	54.76 ± 1.28	52.40 ± 1.04	85.42 ± 10.82	83.33 ± 10.27	55.88 ± 0.92	55.95 ± 1.98	86.75 ± 2.11	82.77 ± 1.23
FGAD	67.90 ± 7.74	65.08 ± 5.27	83.75 ± 3.64	80.24 ± 4.68	81.03 ± 0.89	75.33 ± 0.98	99.25 ± 0.11	98.14 ± 0.62
LG-FGAD	66.40 ± 3.55	63.10 ± 3.15	87.50 ± 4.37	85.00 ± 0.00	81.66 ± 0.16	75.72 ± 0.27	$\mathbf{99.57 \pm 0.06}$	97.55 ± 1.34
FedCIGAR	$\mathbf{70.25 \pm 1.01}$	$\mathbf{65.16 \pm 1.07}$	$\mathbf{97.52 \pm 1.71}$	$\mathbf{91.67 \pm 0.00}$	$\mathbf{83.13 \pm 0.79}$	$\mathbf{76.79 \pm 0.51}$	99.33 ± 0.09	$\mathbf{98.84 \pm 0.33}$

Table 1: Anomaly detection performance in terms of AUC and F1-Score (in percent, mean±std) under the single-dataset setting.

Methods	MOLECULES		BIOCHEM		SMALL		MIX	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
Self-train	70.35 ± 0.56	66.38 ± 0.68	68.3 ± 0.74	61.67 ± 1.31	57.92 ± 4.56	51.12 ± 4.11	59.8 ± 0.33	52.14 ± 0.32
FedAvg	54.41 ± 3.21	55.57 ± 1.46	47.49 ± 1.04	51.24 ± 1.42	48.90 ± 0.60	52.77 ± 0.64	47.96 ± 0.61	53.89 ± 0.83
FedProx	57.93 ± 2.14	55.91 ± 0.44	46.04 ± 0.49	51.71 ± 1.59	48.89 ± 0.50	52.42 ± 0.27	46.79 ± 0.63	53.50 ± 0.88
GCFL	58.86 ± 1.09	57.80 ± 1.20	51.44 ± 1.18	54.88 ± 1.67	53.93 ± 0.51	57.68 ± 0.02	51.46 ± 0.96	55.35 ± 1.14
FedStar	57.03 ± 2.02	55.54 ± 0.86	47.80 ± 0.48	53.21 ± 0.84	51.09 ± 2.00	55.28 ± 1.92	51.68 ± 1.61	54.89 ± 1.11
FGAD	62.00 ± 5.13	60.00 ± 4.78	61.41 ± 1.78	55.20 ± 1.77	62.38 ± 1.98	56.27 ± 2.23	59.58 ± 0.31	52.46 ± 0.59
LG-FGAD	70.84 ± 0.47	65.72 ± 0.59	67.98 ± 0.16	61.52 ± 0.81	66.39 ± 0.91	$\mathbf{62.22 \pm 1.30}$	62.26 ± 0.97	$\mathbf{56.52 \pm 0.39}$
FedCIGAR	$\mathbf{74.22 \pm 1.37}$	$\mathbf{69.52 \pm 1.46}$	$\mathbf{72.20 \pm 1.95}$	$\mathbf{65.03 \pm 1.40}$	$\mathbf{70.03 \pm 1.21}$	61.11 ± 1.12	$\mathbf{63.77 \pm 2.04}$	54.96 ± 1.68

Table 2: Anomaly detection performance in terms of AUC and F1-Score (in percent, mean±std) under the multi-dataset setting.



(a) MUTAG (b) PTC_MR

Figure 3: Visualization of node weight.

Method	MOLECULES	BIOCHEM	MIX	MUTAG
FedCIGAR	$\mathbf{74.22 \pm 1.37}$	$\mathbf{72.20 \pm 1.95}$	$\mathbf{63.77 \pm 2.04}$	$\mathbf{97.52 \pm 1.71}$
w/o Struct	70.70 ± 1.06	68.16 ± 3.24	60.45 ± 1.94	93.98 ± 0.65
w/o Gate	63.32 ± 1.36	62.86 ± 3.52	58.87 ± 0.56	96.76 ± 0.66
w/o Cluster	73.83 ± 1.28	71.13 ± 1.29	61.85 ± 1.92	97.38 ± 0.66

Table 3: Performance of FedCIGAR and variants in terms of AUC.

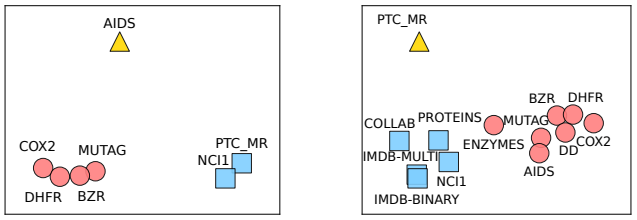
three variants: ❶ **w/o SE:** removing the structure encoding branch; ❷ **w/o Gating:** removing the anomaly gate module; and ❸ **w/o Cluster:** replacing the sliding clustering module with the FedAvg parameter aggregation method.

As shown in Table 3, all components contribute to the overall performance. ❶ Removing the structure encoding (**w/o SE**) leads to significant performance degradation, demonstrating that high-quality node representations are crucial for effectively distinguishing anomalies. ❷ Removing the gating module (**w/o Gate**) results in a substantial performance drop across multi-datasets, which confirms the importance of capturing the client's specific patterns. ❸ The single-dataset shows a light decrease compared to the multi-dataset after removing the sliding clustering mechanism (**w/o Cluster**), which confirms that the naive FedAvg algorithm fails in non-IID scenarios.

4.4 Visualization Analysis

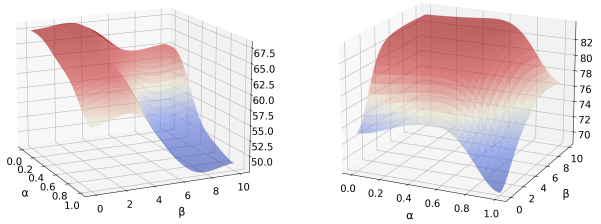
Node Contribution. To investigate the weight allocation mechanism of the gating module in FedCIGAR, we visualize the node contribution weights on the MUTAG and PTC_MR datasets from SMALL. As shown in Fig. 3(a), FedCIGAR tends to assign extreme weights to N and O atoms, and assign uniform weights for C atoms. This is consistent with the patterns of MUTAG dataset, in which the "NO_2" functional group is a key structural cue for distinguishing mutagenic molecules. Fig. 3(b) shows the weight allocation for PTC_MR dataset, where FedCIGAR primarily detects anomalies through special atoms (e.g., Cl, O, and S) rather than the carbon chain. In summary, the gating module enables FedCIGAR to adaptively capture node pattern most relevant to anomalies.

Cluster. To better understand the clustering mechanism of FedCIGAR, we visualize the client embedding distributions on both MOLECULES and MIX datasets in Fig. 4. We can witness that FedCIGAR adaptively partitions clients into multiple homogeneous clusters, where compact intra-cluster embeddings and clear inter-cluster separation. These results indicates that the clustering mechanism can explicitly distinguish homogeneous and heterogeneous clients based on their representations, enabling effective aggregation of shared anomaly patterns across datasets and even domains, thereby improving



(a) MOLECULES (b) MIX

Figure 4: Visualization of Cluster.



(a) SMALL (b) DD

Figure 5: The sensitivity of FedCIGAR in terms of α and β.

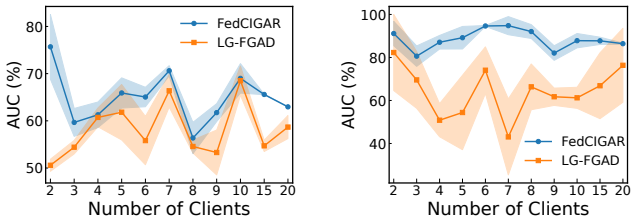
local anomaly detection performance.

4.5 Parameter Analysis

Balance Hyperparameters α and β. Fig. 5 presents the sensitivity of α and β on SMALL and DD datasets, showing clear dataset-dependent trends. SMALL favors smaller α and β, indicating the preference for structural information and reconstruction mean in anomaly detection, whereas DD prefers higher α and β values, relying more on feature information and reconstruction variance. These divergences reveal pronounced heterogeneity in anomaly patterns across graph datasets, and FedCIGAR's multi-aspect modeling improves performance while reducing reliance on any single metric. **Client Number C.** To evaluate the performance of FedCIGAR across different FedGLAD scenarios, we split IMDB-BINARY and MUTAG datasets into 2-20 shards, with each shard assigned to a client. As shown in Fig. 6, increasing the number of clients causes fluctuations in the performance of both algorithms, while FedCIGAR consistently outperforms the state-of-the-art baseline LG-FGAD, demonstrating its robustness. More results can be found in Appendix D.1-D.2.

5 Related Work

Graph-level anomaly detection (GLAD) aims to detect the abnormal graphs that are different from the majority [Zhang *et al.*, 2022; Niu *et al.*, 2023; Sutherland *et al.*, 2003]. Since anomaly labels are difficult to obtain in many practical scenarios [Li *et al.*, 2026a; Pan *et al.*, 2026a], mainstream GLAD studies primarily focus on a one-class learning setting, i.e., models are trained using only normal instances while anomalies are detected as deviations at test time. In recent years, graph neural networks (GNNs) have been widely applied across various tasks[Li *et al.*, 2026b; Miao *et al.*, 2025; Zheng *et al.*, 2022; Qian *et al.*, 2026; Tan *et al.*, 2025] and have become the dominant paradigm for GLAD, with representative work exploring reconstruction-based approaches [Kim *et al.*, 2024], inter-graph relational mining [Ma *et al.*, 2023] and interpretable frameworks [Liu *et al.*, 2023] (see Appendix E.1 for detailed review). Despite their strong performance, these GNN-based methods [Liu *et al.*, 2026; Pan *et al.*, 2026b; Pan *et al.*, 2025; Zhao *et al.*, 2025; Chen *et al.*, 2025] rely on centralized training, where all graph data are collected at a central server. However, in real-world scenarios, graph data are often distributed across multiple clients with non-IID characteristics, highlighting the significance of developing collaborative and privacy-preserving GLAD approaches in a federated setting.



(a) IMDB-BINARY (b) MUTAG

Figure 6: The sensitivity of FedCIGAR in terms of client number C.

Federated graph learning (FGL) extends federated learning (FL) to graph-structured data with distributed and privacy-preserving training [Guo *et al.*, 2023; Wang *et al.*, 2022]. Owing to the diversity of graph data, the data heterogeneity problem (i.e., non-IID data distributions across different clients) is an important challenge in FGL. Prior works have approached this problem through different aspects, such as client clustering [Xie *et al.*, 2021] and structure-aware knowledge sharing [Tan *et al.*, 2023] (see Appendix E.2 for detailed review). Despite their remarkable success, most existing studies focus on node or graph classification, while few studies explore FGL for GLAD. Among these, FGAD [Cai *et al.*, 2024b] and LG-FGAD [Cai *et al.*, 2024a] employ knowledge distillation but rely on the FedAVG-style collaborative learning strategy, which may fail on heterogeneous datasets. To address these issues, we introduce a sliding clustering and anomaly gate module to design an FGL algorithm for GLAD.

6 Conclusion

In this paper, we propose FedCIGAR, a novel reconstruction-based framework for federated graph-level anomaly detection (FedGLAD), which addresses the detection challenges under heterogeneous graph data. FedCIGAR uses a feature-structure joint reconstruction mechanism to capture anomaly patterns without annotated labels. It also integrates a contribution gating module with a sliding window-based clustering aggregation strategy to effectively alleviate data heterogeneity. Extensive experiments on real-world datasets from various domains demonstrate that FedCIGAR consistently outperforms state-of-the-art methods in terms of detection performance.

Acknowledgments

This work was partially supported by the Specific Research Project of Guangxi for Research Bases and Talents (GuiKe AD24010011), the Key Research Development Program Project of Guangxi (GuiKe AB25069095).

Contribution Statement

Yunfeng Zhao and Yixin Liu contributed equally to this work as co-first authors.

References

[Barabasi and Oltvai, 2004] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.

[Cai *et al.*, 2024a] Jinyu Cai, Yunhe Zhang, Jicong Fan, and See-Kiong Ng. Lg-fgad: An effective federated graph anomaly detection framework. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3760–3769, 2024.

[Cai *et al.*, 2024b] Jinyu Cai, Yunhe Zhang, Zhoumin Lu, Wenzhong Guo, and See-Kiong Ng. Towards effective federated graph anomaly detection via self-boosted knowledge distillation. In *ACM Multimedia 2024*, 2024.

[Chen *et al.*, 2025] Qingfeng Chen, Shiyuan Li, Yixin Liu, Shirui Pan, Geoffrey I Webb, and Shichao Zhang. Uncertainty-aware graph neural networks: A multihop evidence fusion approach. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[Dwivedi *et al.*, 2022] Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. In *International Conference on Learning Representations*, 2022.

[Easley *et al.*, 2010] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets: Reasoning about a highly connected world*, volume 1. Cambridge university press Cambridge, 2010.

[Fu *et al.*, 2022] Xingbo Fu, Binchi Zhang, Yushun Dong, Chen Chen, and Jundong Li. Federated graph machine learning: A survey of concepts, techniques, and applications. *ACM SIGKDD Explorations Newsletter*, 24(2):32–47, 2022.

[Ghosh *et al.*, 2020] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. *Advances in neural information processing systems*, 33:19586–19597, 2020.

[Guo *et al.*, 2023] Kun Guo, Yutong Fang, Qingqing Huang, Yuting Liang, Ziyao Zhang, Wenyu He, Liu Yang, Kai Chen, Ximeng Liu, and Wenzhong Guo. Globally consistent federated graph autoencoder for non-iid graphs. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3768–3776, 2023.

[Huang *et al.*, 2023] Wenke Huang, Guancheng Wan, Mang Ye, and Bo Du. Federated graph semantic and structural learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3830–3838, 2023.

[Jamali-Rad *et al.*, 2022] Hadi Jamali-Rad, Mohammad Abdizadeh, and Anuj Singh. Federated learning with taskonomy for non-iid data. *IEEE transactions on neural networks and learning systems*, 34(11):8719–8730, 2022.

[Kim *et al.*, 2024] Sunwoo Kim, Soo Yong Lee, Fanchen Bu, Shinhwan Kang, Kyungho Kim, Jaemin Yoo, and Kijung Shin. Rethinking reconstruction-based graph-level anomaly detection: limitations and a simple remedy. *Advances in Neural Information Processing Systems*, 37:95931–95962, 2024.

[Li *et al.*, 2020] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[Li *et al.*, 2021] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, and Bingsheng He. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3347–3366, 2021.

[Li *et al.*, 2026a] Shiyuan Li, Yixin Liu, Yu Zheng, Xiaofeng Cao, Shirui Pan, and Heng Tao Shen. Towards one-for-all anomaly detection for tabular data. In *International Conference on Machine Learning (ICML)*, 2026.

[Li *et al.*, 2026b] Shiyuan Li, Yixin Liu, Yu Zheng, Mei Li, Quoc Viet Hung Nguyen, and Shirui Pan. OFA-MAS: One-for-all multi-agent system topology design based on mixture-of-experts graph generative models. In *Proceedings of the ACM Web Conference*, 2026.

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

[Liu *et al.*, 2026] 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 Transactions on Knowledge and Data Engineering*, 2026.

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

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

[Ma *et al.*, 2023] Xiaoxiao Ma, Jia Wu, Jian Yang, and Quan Z Sheng. Towards graph-level anomaly detection

via deep evolutionary mapping. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pages 1631–1642, 2023.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[Miao *et al.*, 2025] Rui Miao, Yixin Liu, Yili Wang, Xu Shen, Yue Tan, Yiwei Dai, Shirui Pan, and Xin Wang. Blindguard: Safeguarding llm-based multi-agent systems under unknown attacks. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, 2025.

[Niu *et al.*, 2023] Chaoxi Niu, Guansong Pang, and Ling Chen. Graph-level anomaly detection via hierarchical memory networks. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 201–218. Springer, 2023.

[Pan *et al.*, 2025] Junjun Pan, Yu Zheng, Yue Tan, and Yixin Liu. A survey of generalization of graph anomaly detection: From transfer learning to foundation models. In *The 16th IEEE International Conference on Knowledge Graphs*, 2025.

[Pan *et al.*, 2026a] Junjun Pan, Yixin Liu, Rui Miao, Kaize Ding, Yu Zheng, Quoc Viet Hung Nguyen, Alan Wee-Chung Liew, and Shirui Pan. Explainable and fine-grained safeguarding of llm multi-agent systems via bi-level graph anomaly detection. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, 2026.

[Pan *et al.*, 2026b] Junjun Pan, Yixin Liu, Chuan Zhou, Fei Xiong, Alan Wee-Chung Liew, and Shirui Pan. Correcting false alarms from unseen: Adapting graph anomaly detectors at test time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026.

[Qian *et al.*, 2026] Yanyu Qian, Yue Tan, Yixin Liu, Wang Yu, and Shirui Pan. Dynhd: Hallucination detection for diffusion large language models via denoising dynamics deviation learning. *arXiv preprint arXiv:2603.16459*, 2026.

[Qiu *et al.*, 2022] Chen Qiu, Marius Kloft, Stephan Mandt, and Maja Rudolph. Raising the bar in graph-level anomaly detection. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2196–2203, 2022.

[Sarasamma *et al.*, 2005] Suseela T Sarasamma, Qiuming A Zhu, and Julie Huff. Hierarchical kohonenen net for anomaly detection in network security. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(2):302–312, 2005.

[Stokes *et al.*, 2020] Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

[Sutherland *et al.*, 2003] Jeffrey J Sutherland, Lee A O'brien, and Donald F Weaver. Spline-fitting with a genetic algorithm: A method for developing classification structure-activity relationships. *Journal of chemical information and computer sciences*, 43(6):1906–1915, 2003.

[Tan *et al.*, 2023] Yue Tan, Yixin Liu, Guodong Long, Jing Jiang, Qinghua Lu, and Chengqi Zhang. Federated learning on non-iid graphs via structural knowledge sharing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 9953–9961, 2023.

[Tan *et al.*, 2025] Yue Tan, Xiaoqian Hu, Hao Xue, Celso De Melo, and Flora Salim. Bisecle: Binding and separation in continual learning for video language understanding. *Advances in Neural Information Processing Systems*, 38:33752–33782, 2025.

[Wang *et al.*, 2022] Binghui Wang, Ang Li, Meng Pang, Hai Li, and Yiran Chen. Graphfl: A federated learning framework for semi-supervised node classification on graphs. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 498–507. IEEE, 2022.

[Wang *et al.*, 2024] Yingcheng Wang, Songtao Guo, Dewen Qiao, Guiyan Liu, and Mingyan Li. Fedsg: A personalized subgraph federated learning framework on multiple non-iid graphs. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(5):3678–3690, 2024.

[Xie *et al.*, 2021] Han Xie, Jing Ma, Li Xiong, and Carl Yang. Federated graph classification over non-iid graphs. *Advances in neural information processing systems*, 34:18839–18852, 2021.

[Yin *et al.*, 2024] Jiaxin Yin, Yuanyuan Qiao, Zitang Zhou, Xiangchao Wang, and Jie Yang. Mcm: Masked cell modeling for anomaly detection in tabular data. In *The Twelfth International Conference on Learning Representations*, 2024.

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

[Zhao and Akoglu, 2023] Lingxiao Zhao and Leman Akoglu. On using classification datasets to evaluate graph outlier detection: Peculiar observations and new insights. *Big Data*, 11(3):151–180, 2023.

[Zhao *et al.*, 2025] Yunfeng Zhao, Yixin Liu, Shiyuan Li, Qingfeng Chen, Yu Zheng, and Shirui Pan. Freegad: A training-free yet effective approach for graph anomaly detection. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pages 4379–4389, 2025.

[Zheng *et al.*, 2022] Yu Zheng, Ming Jin, Yixin Liu, Lianhua Chi, Khoa T Phan, and Yi-Ping Phoebe Chen. From unsupervised to few-shot graph anomaly detection: A multi-scale contrastive learning approach. *arXiv preprint arXiv:2202.05525*, 2022.