LG-FGAD: An Effective Federated Graph Anomaly Detection Framework

Jinyu Cai[1] , **Yunhe Zhang**[2,3] , **Jicong Fan**[2,3*] and **See-Kiong Ng**[1]
[1]Institute of Data Science, National University of Singapore, Singapore
[2]Shenzhen Research Institute of Big Data, Shenzhen, China
[3]School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China
{jinyucai, seekiong}@nus.edu.sg, zhangyhannie@gmail.com, fanjicong@cuhk.edu.cn

Abstract

Graph anomaly detection (GAD), which aims to identify those graphs that are significantly different from other ones, has gained growing attention in many real-world scenarios. However, existing GAD methods are generally designed for centralized training, while in real-world collaboration, graph data is generally distributed across various clients and exhibits significant non-IID characteristics. To tackle this challenge, we propose a federated graph anomaly detection framework with local-global anomaly awareness (LG-FGAD). We first introduce a self-adversarial generation module and train a discriminator to identify the generated anomalous graphs from the normal graph. To enhance the anomaly awareness of the model, we propose to maximize/minimize the mutual information from local and global perspectives. Importantly, to alleviate the impact of non-IID problems in collaborative learning, we propose a dual knowledge distillation module. The knowledge distillation is conducted over both logits and embedding distributions, and only the student model engages in collaboration to preserve the personalization of each client. Empirical results on various types of real-world datasets prove the superiority of our method.

1 Introduction

Graph data has widely existed in many real-world scenarios, e.g., the medical, biological, and social network data [Aggarwal, 2011; Li *et al.*, 2021c; Li *et al.*, 2023a; Sun and Fan, 2024; Cai *et al.*, 2024a] is naturally graph-structured. Graph anomaly detection (GAD) [Pang *et al.*, 2021; Ma *et al.*, 2021; Duan *et al.*, 2023], which aims to identify abnormal patterns within graph data, is a fundamental problem in machine learning and has raised growing research interests. In this paper, we focus on the challenging anomaly detection problem of graph-level data. Recently, numerous GAD methods [Zhao and Akoglu, 2021; Qiu *et al.*, 2022; Ma *et al.*, 2022; Zhang *et al.*, 2024b] have been proposed. They typically utilize emerging graph neural networks (GNNs) [Kipf and Welling, 2017; Xu *et al.*, 2019; Chen *et al.*, 2023b; Chen *et al.*, 2023a; Wu *et al.*, 2024a] as the backbone, and combine with general anomaly detection methods like deep one-class classification (DeepSVDD) [Ruff *et al.*, 2018] to detect anomalous graphs. One may also adopt recent advances (e.g., [Cai and Fan, 2022; Zhao *et al.*, 2023; Liu *et al.*, 2023; Fu *et al.*, 2024]) in general anomaly detection to graph data.

However, existing GAD methods are generally designed for centralized training [Ma *et al.*, 2021; Ma *et al.*, 2023], which requires the collection of all training data. In real-world scenarios, data is often distributed among different clients of data owners. Unfortunately, data owners are often reluctant to share private data to collaboratively train a GAD model due to privacy leakage concerns, which hinder the application of those centralized GAD methods. Therefore, a promising research problem is *how to facilitate the collaboration with multiple data owners for training a GAD model while guaranteeing the data security and model efficiency*.

Federated learning (FL) [Li *et al.*, 2021b; Fu *et al.*, 2022; Qiao *et al.*, 2023; Zhang *et al.*, 2024a], an emerging technique that aims to facilitate secure collaborative learning, provides an intuitive solution to this issue. In classical FL methods, e.g., FedAvg [McMahan *et al.*, 2017] and FedProx [Li *et al.*, 2020], data privacy can be protected to some extent as it excludes the sharing of sensitive data information between clients. In addition, federated graph learning (FGL) [Xie *et al.*, 2021; Tan *et al.*, 2023; Huang *et al.*, 2023], a branch of FL that specializes in graph data, has also been investigated in recent years. FGL methods generally utilize GNNs [Kipf and Welling, 2017; Xu *et al.*, 2019; Sun *et al.*, 2024; Wu *et al.*, 2024b] and are promising for detecting anomalous graphs in FL scenarios. Therefore, combining the state-of-the-art FL or FGL methods with existing anomaly detection methods such as DeepSVDD is an intuitive solution to address the abovementioned problem.

Nevertheless, it may encounter the following challenges:

1. The above solutions identify anomalies by the learned representation on the graph level, which overlooks the factors within the node level that potentially cause anomalies.
2. The graph data distributed across different clients exhibits non-IID characteristics [Xie *et al.*, 2021] (see the specific setting in Section 4.1), which is a challenging issue.
3. The integration of DeepSVDD poses a great challenge

*Corresponding author.

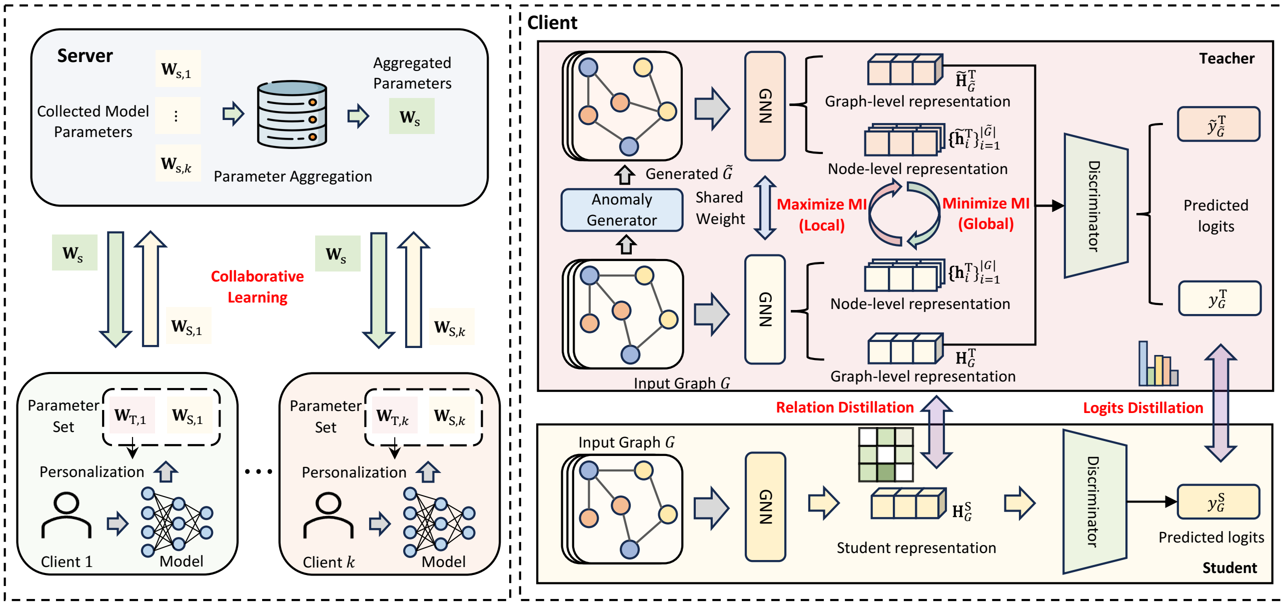


Figure 1: An illustration of the proposed LG-FGAD. In each client, a set of anomalous graphs is generated through the anomaly generator, and then a discriminator (anomaly detector) is trained in an adversarial manner. The local-global anomaly awareness module is introduced to enhance the discrimination ability of the anomaly detection model. Furthermore, a dual knowledge distillation (relation and logits) mechanism is proposed to preserve the personalization of local models (via teacher) and alleviate the non-IID problem (via student).

to learning a unified decision boundary on such non-IID graph data across different clients.

To address these challenges, we propose a local-global federated graph anomaly detection (LG-FGAD[1]) method in this paper. Figure 1 shows the network architecture. In each client, we introduce an adversarial graph anomaly detection framework based on generative adversarial network (GAN) [Goodfellow *et al.*, 2014] and graph isomorphism network (GIN) [Xu *et al.*, 2019]. The generator aims to generate a set of anomalous graphs while the discriminator is trained to serve as the anomaly detector for identifying those anomalies from normal graphs. To improve the discrimination ability for anomalous patterns within graph data, we propose a local-global anomaly awareness module that maximizes/minimizes the mutual information (MI) between normal and generated anomalous graphs from the local and global perspectives. Additionally, to alleviate the non-IID problems, we further propose a dual knowledge distillation mechanism, which introduces a student model to distill the knowledge of both embeddings and predicted logits from the teacher model. Notably, we let only the student models participate in collaborative learning between clients. Thus, the personalization of each client is preserved by the teacher model, and the collaboration efficiency is also improved since the student model simplifies the model complexity.

Our contributions are summarized as follows:

- We investigate the challenging graph anomaly detection task in graph data distributed across multiple clients (non-IID setting), and propose a novel LG-FGAD method to handle this issue.
- We propose a local-global anomaly awareness module, which enhances the discrimination ability via the mutual information maximization/minimization over the node/graph level between the normal and generated anomalous graphs.
- We propose a dual knowledge distillation mechanism to preserve the personalization of local clients, thereby alleviating the impact of non-IID problems.
- The proposed LG-FGAD framework significantly reduces the model parameters engaged in collaborative learning, thus improving communication efficiency.

We demonstrate the superiority of LG-FGAD through the comparison with state-of-the-art baselines on various data, e.g., medical, biological, and social network data.

2 Related Work

2.1 Graph Anomaly Detection

Graph anomaly detection (GAD) [Akoglu *et al.*, 2015; Ma *et al.*, 2021; Zheng *et al.*, 2021; Cai *et al.*, 2023] is a fundamental machine learning task on identifying abnormal patterns in graph-structured data, involving anomaly detection on node, edge, or graph. In this paper, we focus on anomaly detection on the entire graph. Recent advances lie in the integration of powerful GNN backbone [Kipf and Welling, 2017; Xu *et al.*, 2019] with general anomaly detection methods, e.g., DeepSVDD [Ruff *et al.*, 2018]. For instance, Zhao and Akoglu [2021] proposed one-class GIN (OCGIN), which integrates GIN and DeepSVDD to handle graph-level anomaly detection issues. Qiu *et al.* [2022] proposed one-class graph

[1]Code is available at https://github.com/wownice333/LG-FGAD

transformation learning (OCGTL), which overcomes the performance flip problem in OCGIN by neural transformation learning. Zhang *et al.* [2024b] improved DeepSVDD by standardizing the distribution of latent representation and further handled the soap bubble problem with a novel bi-hypersphere compression strategy. While these methods have shown efficacy in centralized settings, real-world graph data, e.g., social networks, is generally distributed in various clients, which poses significant challenges as the graph data across clients is often non-IID and heterogeneous. Particularly, adapting anomaly detection methods such as DeepSVDD in FL is challenging, as it is difficult to learn a uniform hypersphere that can accurately represent normal data patterns for non-IID data distributed in diverse clients. Therefore, it is of significant practical interest to study how to effectively detect anomalies on non-IID graphs.

2.2 Federated Learning

Federated learning (FL) [McMahan *et al.*, 2017] is an emerging paradigm in machine learning that involves training models across multiple decentralized devices or clients with local data, which has been proven effective in machine learning tasks such as image classification [Li *et al.*, 2021a; Sun *et al.*, 2023b], information retrieval [Pinelli *et al.*, 2023; Sun *et al.*, 2023a], and clustering [Qiao *et al.*, 2023; Zhang *et al.*, 2022; Chen *et al.*, 2022; Li *et al.*, 2023b; Cai *et al.*, 2024b], etc. FedAvg [McMahan *et al.*, 2017] is a classical FL method, where local models are trained on individual clients and then average their parameters to update a global model. FedProx [Li *et al.*, 2020] introduces a proximal term to the loss function in FedAvg, which improves the stability and performance in handling non-IID data. Federated graph learning (FGL) [Zhang *et al.*, 2021; Liu *et al.*, 2022; Wang *et al.*, 2022], which aims to collaboratively train robust GNN models and guarantee data privacy, has also been widely investigated recently. Xie *et al.* [2021] proposed clustered federated learning (GCFL), which dynamically identifies clusters of clients to reduce the heterogeneity and facilitate more effective federated learning. Tan *et al.* [2023] proposed FedStar, which addresses non-IID problems by separately encoding structural information and sharing it across clients, whereas the personalizations of the local models are preserved by graph feature knowledge. The potential of FL and FGL methods to handle anomaly detection on non-IID graphs is worth investigating. Consequently, in this paper, we study and evaluate their performance on various types of real-world graph data, e.g., medical, biological, and social network scenarios.

3 Methodology

3.1 Preliminary

Notation: A graph dataset can be represented as $D = \{G_1, \ldots, G_N\}$, which contains N graphs. Each graph G_i in the graph set is composed of a node set V_i and an edge set E_i, i.e., $G_i = \{V_i, E_i\}$. $\mathbf{A}_i \in \{0,1\}^{n_i \times n_i}$ is an adjacency matrix that is used to denote the topology of G_i, where n_i denotes the number in the node set, i.e., $n_i = |V_i|$. $\mathbf{X}_i \in \mathbb{R}^{n_i \times d}$ is the initial attributed feature of G_i. In federated learning setting, we use $D = \{D_1, \ldots, D_C\}$ to represent the dataset distributed in C clients.

Graph Isomorphism Network: In this study, we utilize Graph Isomorphism Network (GIN) [Xu *et al.*, 2019], a prevalent GNN backbone, to learn the graph-level representation for graph data. The representation learning in each layer of a GIN involves neighborhood aggregation and message propagation. Specifically, the aggregated neighborhood features in the k-th layer, expressed as $\mathbf{a}_v^{(k)}$, is derived by:

$$\mathbf{a}_v^{(k)} = \mathrm{AGGREGATE}(\{\mathbf{h}^{(k-1)}(u), u \in \tilde{\mathcal{N}}(v)\}), \tag{1}$$

where $\tilde{\mathcal{N}}(v)$ indicates the neighboring node set of node v, and $\mathrm{AGGREGATE}(\cdot)$ is the aggregation function. The node feature vector $\mathbf{h}_v^{(k)}$ in the k-th layer is derived by the combination of the aggregated feature and the feature learned in the $(k-1)$-th layer as follows:

$$\mathbf{h}_v^{(k)} = \sigma(\mathrm{COMBINE}(\mathbf{h}_v^{(k-1)}, \mathbf{a}_v^{(k)})), \tag{2}$$

where $\sigma(\cdot)$ is the activation function, such as ReLU. Notably, the feature $\mathbf{h}_v^{(0)}$ is initialized by the attributed feature $\mathbf{x}_v$ of node v. Then, the representation of an entire graph G can be obtained by combining the learned features of all its nodes:

$$\mathbf{h}_G = \mathcal{R}(\mathrm{CONCAT}(\mathbf{h}_v^{(k)}, k \in \{1, \ldots, K\}), v \in G), \tag{3}$$

where K denotes the number of latent layers in GIN. $\mathcal{R}(\cdot)$ is the readout function that aggregates node features to derive the graph-level representation. In this study, we opt for the sum readout strategy [Xu *et al.*, 2019]. $\mathrm{CONCAT}(\cdot)$ signifies the operation of feature concatenation across all layers. For simplicity, $\mathrm{GIN}(\cdot)$ is used to denote the GIN model comprising the above three key operations throughout this paper.

3.2 Self-Adversarial Generation for Detecting Anomalous Graphs

In this paper, we propose to learn an anomaly detection model via an adversarial framework between an anomalous graph generator and a discriminator (anomaly detector). To this end, we first propose to generate a set of anomalous graph $\tilde{D}_c$ via a self-adversarial generation strategy using the local dataset D_c in each client. To generate diverse anomalous graphs, we leverage the variational inference [Kipf and Welling, 2016] to guarantee the diversity in latent space. Specifically, we first map the graph data into a latent space following Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\sigma}_c^2)$ by:

$$\boldsymbol{\mu}_c = \mathrm{GIN}_{\boldsymbol{\mu}}(\mathbf{X}_c, \mathbf{A}_c), \quad \log \boldsymbol{\sigma}_c = \mathrm{GIN}_{\boldsymbol{\sigma}}(\mathbf{X}_c, \mathbf{A}_c), \tag{4}$$

where $\mathbf{X}_c$ and $\mathbf{A}_c$ denote the attributed features and adjacency matrix of dataset D_c. Then, $\boldsymbol{\mu}_c$ and $\boldsymbol{\sigma}_c$ can explicitly parameterize an inference model:

$$q(\tilde{\mathbf{Z}}_c|\mathbf{X}_c, \mathbf{A}_c) = \prod_{i=1}^{|D_c|} q(\mathbf{Z}_c^{(i)}|\mathbf{X}_c, \mathbf{A}_c), \tag{5}$$

where $q(\tilde{\mathbf{Z}}_c^{(i)}|\mathbf{X}_c, \mathbf{A}_c) = \mathcal{N}(\tilde{\mathbf{Z}}_c^{(i)}|\boldsymbol{\mu}_c^{(i)}, \mathrm{diag}({\boldsymbol{\sigma}_c^{(i)}}^2))$. To guarantee the propagation of the gradient information during the sampling process, we utilize the reparametrization trick [Kingma and Welling, 2013] to sample the latent representation $\tilde{\mathbf{z}}_c$:

$$\tilde{\mathbf{z}}_c = \boldsymbol{\mu}_c + \epsilon \exp(\boldsymbol{\sigma}_c),\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{1}), \tag{6}$$

where ϵ is a random noise factor follows the standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{1})$. Consequently, we can derive the generated adjacency matrix $\tilde{\mathbf{A}}_c$ and attributed features $\tilde{\mathbf{X}}_c$ by:

$$\tilde{\mathbf{A}}_c = \mathcal{T}(\tilde{\mathbf{Z}}_c \tilde{\mathbf{Z}}_c^\top),\ \ \tilde{\mathbf{X}}_c = \mathrm{MLP}(\tilde{\mathbf{Z}}_c), \tag{7}$$

where $\mathcal{T} : \mathbb{R} \to [0, 1]$ indicates a transformation function, e.g, Sigmoid$(\cdot)$, and $\mathrm{MLP}(\cdot)$ denotes a multi-layer perception based decoder. For simplicity, we can use a generator $\mathcal{G}(\cdot, \cdot)$ to represent the above generation process.

We then introduce a discriminator $\mathcal{D}(\cdot, \cdot)$ that aims to distinguish the normal graphs from the generated anomalous graphs. We project them into graph-level representations by:

$$\mathbf{H}_c = \mathcal{F}(\mathbf{X}_c, \mathbf{A}_c),\ \ \tilde{\mathbf{H}}_c = \mathcal{F}(\mathcal{G}(\mathbf{X}_c, \mathbf{A}_c)), \tag{8}$$

where $\mathcal{F}(\cdot, \cdot)$ denotes a weight-shared GNN model, e.g., GIN. The learned graph-level representation $\mathbf{H}_c$ and $\tilde{\mathbf{H}}_c$ serve as the input to the discriminator. Then, we can optimize the anomaly detection model via adversarial training of the following objectives, $\mathcal{L}^c_{\mathcal{G}}$ (generator) and $\mathcal{L}^c_{\mathcal{D}}$ (discriminator):

$$\begin{aligned}\mathcal{L}^c_{\mathcal{G}} = \underset{\mathbf{X}_c, \mathbf{A}_c \sim \mathbb{P}_{D_c}}{\mathbb{E}} &- [\mathcal{D}(\mathcal{F}(\mathcal{G}(\mathbf{X}_c, \mathbf{A}_c)))] \\ &+ \mathrm{KL}(q(\tilde{\mathbf{Z}}_c | \mathbf{X}_c, \mathbf{A}_c) || P(\mathbf{Z})),\end{aligned} \tag{9}$$

$$\mathcal{L}^c_{\mathcal{D}} = \underset{\mathbf{X}_c, \mathbf{A}_c \sim \mathbb{P}_{D_c}}{\mathbb{E}} [\mathcal{D}(\mathcal{F}(\mathcal{G}(\mathbf{X}_c, \mathbf{A}_c))) - \mathcal{D}(\mathcal{F}(\mathbf{X}_c, \mathbf{A}_c))]. \tag{10}$$

The model training is a classical min-max optimization, where the generator $\mathcal{G}$ takes the normal graphs as the input and aims to produce a set of high-quality anomalous graphs that resemble normal graphs to fool the discriminator. Conversely, the discriminator $\mathcal{D}$ aims to identify the anomalous graphs as much as possible. $P(\mathbf{Z}) = \prod_i p(\mathbf{Z}_i) = \prod_i \mathcal{N}(\mathbf{Z}_i | \mathbf{0}, \mathbf{I})$ is a prior latent distribution follows Gaussian distribution. The $\mathrm{KL}(\cdot || \cdot)$ term penalizes the KL-divergence between $q(\tilde{\mathbf{Z}}_c | \mathbf{X}_c, \mathbf{A}_c)$ and $P(\mathbf{Z})$, which encourages $\tilde{\mathbf{Z}}_c$ to be uniformly distributed in the latent space, thereby facilitating the diverse anomalous graphs generation. However, this solution has the following challenging problems:

- Considering only the graph level may not comprehensively reflect anomaly patterns, such as in the node level.
- The collaborative learning between different clients will pose severe non-IID problems as the data distribution across clients differs significantly.

3.3 Enhancing Anomaly Awareness with Local-Global Mutual Information

To address the first challenge, in this paper, we introduce a local-global anomaly awareness module based on mutual information (MI). For a normal graph $G \in D_c$, we can derive a set of node-level representations $\{\mathbf{h}_v\}_{v \in G}$ and graph-level representation $\mathbf{H}_G$ during the graph representation learning. Similarly, $\{\tilde{\mathbf{h}}_v\}_{v \in \tilde{G}}$ and $\tilde{\mathbf{H}}_{\tilde{G}}$ can be derived for a generated graph $\tilde{G} \in \tilde{D}_c$. From the local perspective, we propose to maximize the MI between the node-level and graph-level representations within a graph. Taking a normal graph as an instance, this can be quantified by defining a local MI estimator $\mathcal{I}_{\mathrm{local}}$ as follows:

$$\mathcal{I}_{\mathrm{local}}(\mathbf{h}_v; \mathbf{H}_G) := \mathbb{E}_{\mathbb{P}_G}[-\mathrm{sp}(-\mathcal{J}(\mathbf{h}_v, \mathbf{H}_G))], \tag{11}$$

where $\mathbb{P}_G$ is the normal graph distribution, $\mathrm{sp}(x) = \log(1 + e^x)$ denotes the softplus function, and $\mathcal{J}(\mathbf{x}, \mathbf{y}) = JS(\mathbf{x}\mathbf{y}^T)$ represents a JS-divergence based transformation function. Similarly, $\mathcal{I}_{\mathrm{local}}(\tilde{\mathbf{h}}_v, \tilde{\mathbf{H}}_{\tilde{G}})$ can be used to estimate the local MI within a generated anomalous graph. Then, from the global perspective, we aim to minimize the MI of the graph-level representations between normal graphs and generated anomalous graphs. To this end, we further introduce $\mathcal{I}_{\mathrm{global}}$, a global MI estimator:

$$\mathcal{I}_{\mathrm{global}}(\mathbf{H}_G; \tilde{\mathbf{H}}_{\tilde{G}}) := \mathbb{E}_{\mathbb{P}_{G \times \tilde{G}}}[-\mathrm{sp}(-\mathcal{J}(\mathbf{H}_G, \tilde{\mathbf{H}}_{\tilde{G}}))], \tag{12}$$

where $\mathbb{E}_{\mathbb{P}_{G \times \tilde{G}}}$ is the joint distribution of normal and generated graphs. Consequently, we propose the following objective function aims at enhancing the awareness of the model to the anomaly patterns across node level and graph level:

$$\begin{aligned}\mathcal{L}^c_{\mathrm{LGMI}} = &\sum_{G \in D_c} \frac{1}{|D_c|} \sum_{v \in G} -\mathcal{I}_{\mathrm{local}}(\mathbf{h}_v, \mathbf{H}_G) \\ &+ \sum_{\tilde{G} \in \tilde{D}_c} \frac{1}{|\tilde{D}_c|} \sum_{v \in \tilde{G}} -\mathcal{I}_{\mathrm{local}}(\tilde{\mathbf{h}}_v, \tilde{\mathbf{H}}_{\tilde{G}}) \\ &+ \sum_{G \in D_c} \frac{1}{|D_c|} \sum_{\tilde{G} \in \tilde{D}_c} \mathcal{I}_{\mathrm{global}}(\mathbf{H}_G, \tilde{\mathbf{H}}_{\tilde{G}})\end{aligned} \tag{13}$$

where $|D_c|$ denotes the number of graphs in the c-th client. This objective function is strategically designed: the two $\mathcal{I}_{\mathrm{local}}$ terms maximize the MI between node representation and graph representation within individual normal or anomalous graphs, while the $\mathcal{I}_{\mathrm{global}}$ term minimizes MI across normal and anomalous graphs, thus improving the discrimination ability to the normal and anomalous patterns.

3.4 Effective Collaborative Learning with Dual Knowledge Distillation

The second challenge points out the widely-existed heterogeneity and non-IID characteristics of graph data distributed across clients. This potentially adversely affects collaborative learning between clients, and results in unsatisfactory anomaly detection performance. To this end, we propose a dual knowledge distillation mechanism to preserve the personalization of local clients, thereby alleviating the influence of non-IID problems. Specifically, we regard the anomaly detection model defined in Section 3.2 and 3.3 as the teacher model, and introduce a student model that aims to distill knowledge from the teacher model and participate in collaborative learning. The network structure of the student model, as shown in Figure 1, is similar to the teacher model but with simplified hidden layers. Besides, the generator is not included in the student model as our goal is to distill knowledge about normal graphs.

The classical knowledge distillation [Hinton *et al.*, 2015; Zhang *et al.*, 2023] generally performs on the predicted logits. For a graph dataset D_c in a client, we can obtain the predicted logits $\mathbf{Y}_{c,\mathrm{t}}$ and $\mathbf{Y}_{c,\mathrm{s}}$ from the discriminator $\mathcal{D}_\mathrm{t}$ and $\mathcal{D}_\mathrm{s}$ of teacher and student models as follows:

$$\mathbf{Y}_{c,\mathrm{t}} = \mathcal{D}_\mathrm{t}(\mathbf{H}_c^\mathrm{t}), \ \ \mathbf{Y}_{c,\mathrm{s}} = \mathcal{D}_\mathrm{s}(\mathbf{H}_c^\mathrm{s}), \tag{14}$$

where $\mathbf{H}_c^\mathrm{t}$ and $\mathbf{H}_c^\mathrm{s}$ denote the graph representations learned in teacher and student models. Then, the logits distillation loss can be formalized as:

$$\ell_{\mathrm{logit}}^c = \frac{1}{|D_c|}\sum_{i=1}^{|D_c|} \mathrm{KL}\left(\mathrm{softmax}\left(\frac{\mathbf{Y}_{c,\mathrm{t}}^{(i)}}{\tau}\right), \mathrm{softmax}\left(\frac{\mathbf{Y}_{c,\mathrm{s}}^{(i)}}{\tau}\right)\right), \tag{15}$$

where $\mathrm{KL}(\cdot,\cdot)$ denotes the Kullback-Leibler divergence, τ is the temperature factor that determines the smoothness of knowledge distillation, and $\mathrm{softmax}(y_i/\tau) = \frac{\exp(y_i/\tau)}{\sum_j \exp(y_j/\tau)}$ is the softmax function.

However, the logits distillation is conducted in the output space and hardly distills the knowledge in the local-global anomaly awareness module in the latent space. Therefore, we further introduce an embedding distillation that focuses on distilling knowledge in graph representation learning. Due to the inconsistent architecture between the student and teacher model, we leverage a relation distillation strategy [Park *et al.*, 2019] instead of direct embedding distillation. Specifically, we define a distance function $\psi_{\mathrm{dis}}(\cdot,\cdot)$ to measure the Euclidian distance between two embeddings as follows:

$$\psi_{\mathrm{dis}}(\mathbf{H}_\mathrm{t}^{(i)}, \mathbf{H}_\mathrm{t}^{(j)}) = \frac{1}{\eta}\left\|\mathbf{H}_\mathrm{t}^{(i)} - \mathbf{H}_\mathrm{t}^{(j)}\right\|_2, \tag{16}$$

where we opt for the non-squared Euclidean norm to prevent the overemphasis on large distances. $\mathbf{H}_\mathrm{t}^{(i)} \in \mathbb{R}^{|D_c|\times K_\mathrm{t} d'}$ denotes the concatenated features across K_t latent layers with d' dimensions in the teacher model for simplicity, and the same definition for $\mathbf{H}_\mathrm{s}^{(i)} \in \mathbb{R}^{|D_c|\times K_\mathrm{s} d'}$. Besides, we let $\eta = \frac{1}{|D_c|}\sum_{i=1}^{|D_c|}\sum_{j=1}^{|D_c|}||\mathbf{H}_\mathrm{t}^{(i)} - \mathbf{H}_\mathrm{t}^{(j)}||_2$ to normalize the distance. Thus, for a graph dataset D_c in a client, the embedding relation distillation loss can be defined as:

$$\ell_{\mathrm{emb}}^c = \frac{1}{|D_c|}\sum_{i=1}^{|D_c|}\sum_{j=1}^{|D_c|} l_\mathrm{H}\left(\psi_{\mathrm{dis}}(\mathbf{H}_\mathrm{t}^{(i)}, \mathbf{H}_\mathrm{t}^{(j)}), \psi_{\mathrm{dis}}(\mathbf{H}_\mathrm{s}^{(i)}, \mathbf{H}_\mathrm{s}^{(j)})\right). \tag{17}$$

l_H denotes the Huber loss that is robustness to outliers. It is formalized by:

$$l_\mathrm{H}(x,y) = \begin{cases} \frac{1}{2}(x-y)^2 & \text{for } |x-y| \le 1, \\ |x-y| - \frac{1}{2} & \text{otherwise.} \end{cases} \tag{18}$$

Combining the two terms, we can define the loss function of the dual knowledge distillation module as follows:

$$\mathcal{L}_{\mathrm{KD}}^c = \ell_{\mathrm{logit}}^c + \ell_{\mathrm{emb}}^c. \tag{19}$$

Subsequently, the overall objective function of the proposed LG-FGAD is formalized by:

$$\mathcal{L}_{\mathrm{total}} = \frac{1}{C}\sum_{c=1}^{C}\frac{|D_c|}{|D|}\left(\mathcal{L}_\mathcal{G}^c + \mathcal{L}_\mathcal{D}^c + \beta\mathcal{L}_{\mathrm{LGMI}}^c + \gamma\mathcal{L}_{\mathrm{KD}}^c\right) \tag{20}$$

where C denotes the number of clients, $|D|$ denotes the total number of graph across all clients. β and γ are two trade-off parameters. We summarize the effect of each loss as follows:

1. $\mathcal{L}_\mathcal{G}^c$ and $\mathcal{L}_\mathcal{D}^c$ are loss functions for optimizing the generator and discriminator in the adversarial model.
2. $\mathcal{L}_{\mathrm{LGMI}}^c$ aims to maximize/minimize the MI between normal and generated anomalous graphs from the local and global perspectives.
3. $\mathcal{L}_{\mathrm{KD}}^c$ distills knowledge from the teacher model through both relation and logits distillation, which aims to handle the non-IID problem.

Let $\{\mathbf{W}_\mathrm{t}^{(c)}\}_{c=1}^C$ and $\{\mathbf{W}_\mathrm{s}^{(c)}\}_{c=1}^C$ be the parameter sets of teacher model and student model in C clients. During the collaborative learning in each communication round, we let only the student parameter sets $\{\mathbf{W}_\mathrm{s}^{(c)}\}_{c=1}^C$ uploaded to the server, while keeping the teacher parameter sets $\{\mathbf{W}_\mathrm{t}^{(c)}\}_{c=1}^C$ in the local clients for preserving the client personalization. Therefore, the collaborative learning between clients on the server can be formalized as:

$$\bar{\mathbf{W}}_\mathrm{s} = \sum_{c=1}^{C}\frac{|D_c|}{|D|}\mathbf{W}_\mathrm{s}^{(c)}, \tag{21}$$

where $\bar{\mathbf{W}}_s$ represents the aggregated parameter set in the server, and it will be distributed to each client to update the local student model, which is eventually used to detect anomalies. The total training process of LG-FGAD is included in the **Appendix A**.

3.5 Theoretical Complexity Analysis

Assume we have N graphs with maximal m nodes and $|E|$ edges across all graphs, the number of GIN layers is L, and the maximal dimensions between the original attribute and latent representation is represented by $\bar{d}$. Besides, let the maximal latent dimension of the MLP-based discriminator denoted by d_h, and K_t and K_s be the number of latent layers in the teacher and student models. We then analyze the time and space complexity of LG-FGAD in a single client as follows:

- **Time complexity**: The time complexity of a GIN backbone is $\mathcal{O}(NL(m\bar{d}^2 + |E|\bar{d}))$, and for the MLP-based discriminator is $\mathcal{O}(\bar{d}d_h)$. Therefore, the time complexity of the overall framework is $\mathcal{O}(NL(m\bar{d}^2 + |E|\bar{d}) + (K_\mathrm{t} + K_\mathrm{s})\bar{d}d_h)$, where includes the teacher model (anomaly generator, weight-shared GIN backbone, and discriminator) and the student model (GIN backbone and discriminator).
- **Space complexity**: For the GIN backbone, the space complexity is $\mathcal{O}(L\bar{d}(1+\bar{d}))$, which includes the weight and bias matrices in each GIN layer. Similarly, $\mathcal{O}((K_\mathrm{t} + K_\mathrm{s})\bar{d}(1+d_h))$ is the space complexity of teacher and student models. Therefore, the overall space complexity is $\mathcal{O}(L\bar{d}(1+\bar{d}) + (K_\mathrm{t} + K_\mathrm{s})\bar{d}(1+d_h))$.

Moreover, we also analyze the complexity of communication in collaborative learning. According to the definition in our model, only the student model is involved in collaborative learning. Therefore, the communication complexity in terms of time and space within a single communication round are $\mathcal{O}(NL(m\bar{d}^2+|E|\bar{d})+\bar{d}d_h)$ and $\mathcal{O}(L\bar{d}(1+\bar{d})+K_\mathrm{s}\bar{d}(1+d_h))$ respectively.

Methods	IMDB-BINARY		MUTAG		DD		AIDS	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
Self-train	41.58±1.34	42.45±1.29	70.00±3.54	70.00±0.00	52.39±1.31	62.88±1.76	79.48±0.32	72.64±0.58
FedAvg [McMahan *et al.*, 2017]	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 [Li *et al.*, 2020]	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 [Xie *et al.*, 2021]	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 [Tan *et al.*, 2023]	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
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**	**99.57±0.06**	**97.55±1.34**

Table 1: AUCs and F1-Scores (mean (%) ± std (%)) under the single-dataset setting. Note that the best performance is marked in **Bold**.

Methods	MOLECULES		BIOCHEM		SMALL		MIX	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
Self-train	61.26±2.91	57.66±0.00	50.81±0.22	53.68±0.57	59.64±0.11	61.30±0.77	51.94±0.42	54.71±0.00
FedAvg [McMahan *et al.*, 2017]	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 [Li *et al.*, 2020]	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 [Xie *et al.*, 2021]	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 [Tan *et al.*, 2023]	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
LG-FGAD	**70.84±0.47**	**65.72±0.59**	**67.98±0.16**	**61.52±0.81**	**66.39±0.91**	**62.22±1.30**	**62.26±0.97**	**56.52±0.39**

Table 2: AUCs and F1-Scores (mean (%) ± std (%)) under the multi-dataset setting. Note that the best performance is marked in **Bold**.

4 Experiment

4.1 Experimental Setup

Datasets. We conduct experiments on two types of datasets to evaluate the anomaly detection performance:

- **Single-dataset (IID):** We experiment on four single-datasets: IMDB-BINARY, MUTAG, DD, and AIDS. In this setting, the dataset is distributed across multiple clients, and each client possesses a unique subset of that dataset.
- **Multi-dataset (Non-IID):** We experiment on four multi-datasets, including MOLECULES (molecule), BIOCHEM (biology), SMALL (small molecule and protein) and MIX (mixed data types), each of the multi-datasets contains several sub-datasets with different data distributions. In this setting, multiple datasets are distributed to different clients, and each client holds a specific sub-dataset.

The datasets used in the experiment are from publicly available real-world graph benchmarks TUDataset[2], and their construction details are illustrated in **Appendix B**.

Baseline Methods. As the anomaly detection on non-IID graphs is still under-explored, we construct baselines via the combination of DeepSVDD [Ruff *et al.*, 2018] with state-of-the-art FL methods (FedAvg [McMahan *et al.*, 2017], FedProx [Li *et al.*, 2020]) and FGL methods (GCFL [Xie *et al.*, 2021], FedStar [Tan *et al.*, 2023]). Note that the network structure of the GIN backbone used in each baseline is the same as our method to ensure a fair comparison.

Network Structure. We leverage a 4-layer GIN network as the backbone for each baseline, and the latent dimension is fixed to 64. The teacher GIN backbone of LG-FGAD is the same for fairness and followed by a 4-layer MLP discriminator, while the student network employs the 2-layer GIN and 3-layer MLP discriminator to simplify the mode complexity.

[2]https://chrsmrrs.github.io/datasets/

Parameter Settings. We describe the parameter settings in the experiment here.

- For the proposed LG-FGAD, we use Rmsprop as the optimizer and fixed learning rate $\alpha = 0.001$. Besides, we adopt the optimal search grid strategy with all parameters varying in the range $[1e^{-3}, 1e^{2}]$. Regarding some fixed hyperparameters embedded in the architecture, the coefficient of the KL-Divergence regularization term is fixed as $1e^{-4}$, and the temperature of knowledge distillation is set to 1.0. The range of gradient clipping is limited to $[-0.01, 0.01]$.
- For baseline methods, we use Adam as the optimizer, and the learning rate is set to $\alpha = 0.001$. The percentile to draw the final decision boundary in DeepSVDD for AIDS is fixed at 0.3, and others are fixed at 0.001. Other parameter settings follow the default setting in related papers. Note that the batch size and the training epochs are set to 128 and 200 for our method and other baselines, respectively.

Evaluation Metrics. We utilize two popular metrics, i.e., Area Under the Curve (AUC) and F1-Score, for performance evaluation. Besides, to ensure the persuasiveness of the results, we report the performance with means and standard deviations by executing each algorithm 10 times.

Implementation. We implement LG-FGAD based on PyTorch Geometric [Fey and Lenssen, 2019] library in practice. All experiments in this paper are run on the platform with NVIDIA Tesla A100 GPU and AMD EPYC 7532 CPU.

4.2 Comparison with State-of-the-art Methods

To provide comprehensive evaluation, we conduct experiments on various graph data types in real-world scenarios (e.g., molecule, biology, and social network). Tables 1 and 2 present the anomaly detection performance under two different settings, where we compare the proposed LG-FGAD with several state-of-the-art FL and GFL based graph anomaly detection methods. We can observe that LG-FGAD obtains

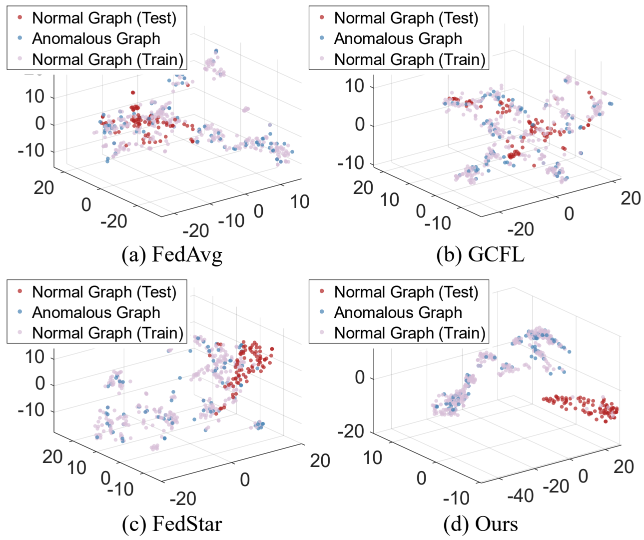


Figure 2: The t-SNE visualization of our method and three baselines on MOLECULES.

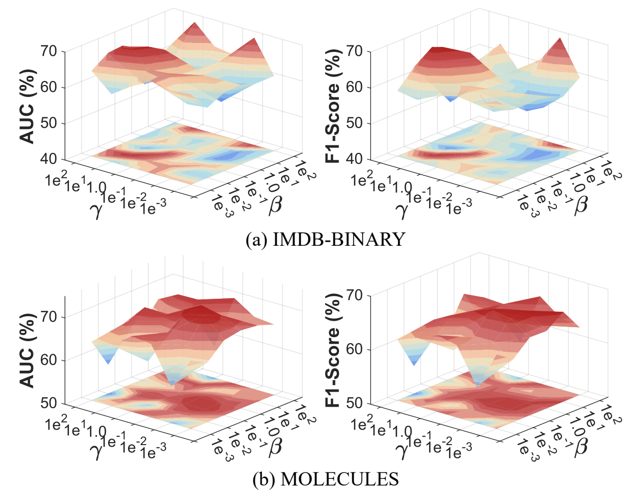


Figure 3: Impact of parameters β and γ on IMDB-BINARY and MOLECULES.

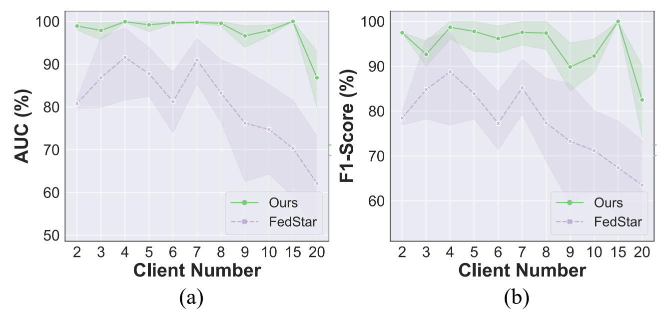


Figure 4: Parameter sensitivity of different numbers of clients on AIDS dataset between FedStar and Ours.

remarkable anomaly detection performance in the single-dataset setting and the more challenging multi-dataset setting. For instance, LG-FGAD outperforms the runner-up FedStar on IMDB-BINARY by more than 10% in AUC and F1-Score, and also 10% higher in both metrics on MOLECULES. Besides, the baseline methods show different levels of effectiveness for different data types. For instance, GCFL outperforms FedStar on IMDB-BINARY in AUC, and the opposite observation is found on DD. While LG-FGAD exhibits great adaptivity to different data types, consistently outperforms other baselines on both metrics. To provide a more intuitive comparison, we leverage t-SNE [Van der Maaten and Hinton, 2008] to visualize the graph representations learned by LG-FGAD and other baselines, which is presented in Figure 2. Compared with other baselines, it is evident that the latent embeddings of anomalous and normal graphs in LG-FGAD are more discriminative and clearly separated in the latent space.

4.3 Parameter Sensitivity Analysis

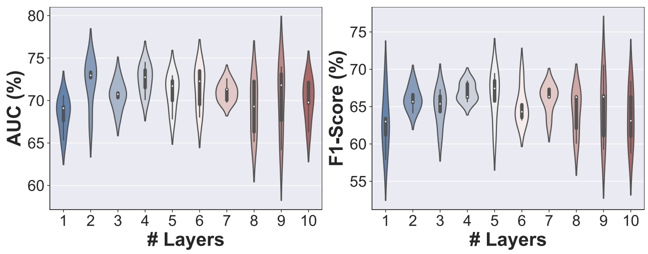
Impact of Hyper-Parameters β and γ. Figure 3 illustrates the impact of varying β and γ on the anomaly detection performance across IMDB-BINARY and MOLECULES datasets. Note that we vary the values of β and γ in a wide ranges, i.e., $[1e^{-3}, 1e^{2}]$. Regions shaded closer to red indicate higher performance, whereas those closer to blue signify lower ones. It can be observed that optimal performance generally obtains within a moderate range of these hyper-parameters, excessively large or small values tend to degrade the effectiveness of the model. Furthermore, LG-FGAD maintains relatively stable performance across a broad spectrum of β and γ values. These observations fully demonstrate the stability of LG-FGAD.

Impact of Client Number C. Figure 4 shows the impact of varying client number C on the anomaly detection performance of LG-FGAD and FedStar. Note that the value of C ranges from 2 to 20. It is evident that both methods suffer from performance decline in an excessive C value, e.g., 20. More importantly, LG-FGAD consistently outperforms state-of-the-art FedStar, and has significantly less performance fluctuation with different C, which further demonstrates the robustness of the proposed LG-FGAD.

Impact of the Student GNN Layers. Figure 5 presents the influence of different numbers of the student GNN layers. The experimental results reveal that an overly shallow network may lead to limited information mining capability. Conversely, excessively deep networks require additional training epochs to fully optimize the objective function. Optimal results are generally achieved with a moderate number of layers, which tends to offer stability. Furthermore, improved performance is observed when the student network employs more layers than the teacher network, as it may potentially generate additional information. However, this benefit is offset by significantly increased communication costs.

4.4 Ablation Study

To verify the individual contributions of each component in LG-FGAD, we conduct a comprehensive ablation study. This study involves the construction of three degradation models



(a) (b)

Figure 5: The impact of different numbers of student GNN layers on the IMDB-BINARY dataset.

Methods	MUTAG		MOLECULES	
	AUC	F1-Score	AUC	F1-Score
Base Model	82.50±5.00	63.81±6.67	56.26±0.00	56.34±0.07
w/o $\mathcal{L}_{\mathrm{LGMI}}$	86.25±1.25	83.50±2.50	64.68±5.87	62.26±4.16
w/o $\mathcal{L}_{\mathrm{KD}}$	85.75±1.25	85.00±5.00	68.22±0.54	64.08±0.52
LG-FGAD	**87.50±4.37**	**85.00±0.00**	**70.84±0.47**	**65.72±0.59**

Table 3: Ablation study results (mean(%) ± std(%)).

of LG-FGAD, each removing specific components:

1. **Base Model**: This variant excludes both the local-global anomaly awareness and dual knowledge distillation modules, remaining solely the adversarial model.
2. **w/o** $\mathcal{L}_{\mathrm{LGMI}}$: This variant removes the local-global anomaly awareness module.
3. **w/o** $\mathcal{L}_{\mathrm{KD}}$: This variant removes the dual knowledge distillation module.

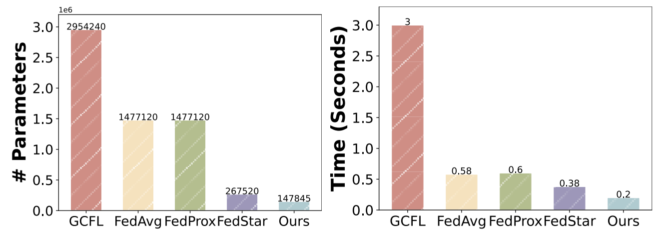
Table 3 shows the performance of LG-FGAD and its variants on MUTAG and MOLECULES, where the experimental setting is the same as described in Section 4.1. The base model exhibits sub-optimal performance on both datasets, suggesting that the mere presence of the GAN-structured model is insufficient for effective anomaly detection on non-IID graphs distributed in multiple clients. Furthermore, a certain performance decrease is observed when removing the corresponding module. i.e., w/o $\mathcal{L}_{\mathrm{LGMI}}$ or w/o $\mathcal{L}_{\mathrm{KD}}$, which fully demonstrates the effectiveness of each component.

4.5 Experimental Complexity Analysis

Here, we analyze the space and time complexity of LG-FGAD, reflected by **1)** the number of parameters involved in the collaborative learning and **2)** Time spent in each communication round. Figure 6 shows the experimental results, where other baselines are included for comparison. It can be observed that the proposed LG-FGAD significantly reduces the parameter size and communication time compared with other baselines, e.g., GCFL, FedAvg, and FedProx.

5 Conclusion

We presented LG-FGAD, a novel method for federated graph anomaly detection. By introducing a local-global anomaly awareness module and a dual knowledge distillation mechanism, LG-FGAD effectively enhances the discriminator ability of the model from both local and global perspectives. Importantly, LG-FGAD not only preserves client personalization but also improves communication efficiency in collaborative learning. Empirical results across diverse real-world datasets, including medical, biological, and social networks, demonstrated the superiority of the proposed LG-FGAD. Nevertheless, one limitation of this work is that it does not currently incorporate differential privacy, which, however, is beyond the scope of this work and could be a future work.



(a) (b)

Figure 6: Space and time complexity (200 communication rounds).

Ethical Statement

There are no ethical issues.

Acknowledgments

This research is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-018). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of National Research Foundation, Singapore.

This research is also supported by the National Natural Science Foundation of China (Grant No.62376236), the Guangdong Provincial Key Lab of Mathematical Foundations for Artificial Intelligence (2023B1212010001), Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone Project (No.HZQSWS-KCCYB2024016), and the funding UDF01001770 of The Chinese University of Hong Kong, Shenzhen.

Contribution Statement

Jinyu Cai and Yunhe Zhang contributed equally to this work as co-first authors.

References

[Aggarwal, 2011] Charu C Aggarwal. *An introduction to social network data analytics*. Springer, 2011.

[Akoglu *et al.*, 2015] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29:626–688, 2015.

[Cai and Fan, 2022] Jinyu Cai and Jicong Fan. Perturbation learning based anomaly detection. *NeurIPS*, 35, 2022.

[Cai *et al.*, 2023] Jinyu Cai, Yunhe Zhang, and Jicong Fan. Self-discriminative modeling for anomalous graph detection. *arXiv preprint arXiv:2310.06261*, 2023.

[Cai *et al.*, 2024a] Jinyu Cai, Yi Han, Wenzhong Guo, and Jicong Fan. Deep graph-level clustering using pseudo-label-guided mutual information maximization network. *NCAA*, pages 1–16, 2024.

[Cai *et al.*, 2024b] Jinyu Cai, Yunhe Zhang, Shiping Wang, Jicong Fan, and Wenzhong Guo. Wasserstein embedding learning for deep clustering: A generative approach. *IEEE T-MM*, 2024.

[Chen *et al.*, 2022] Man-Sheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S. Yu. Efficient orthogonal multi-view subspace clustering. In *SIGKDD*, pages 127–135. ACM, 2022.

[Chen *et al.*, 2023a] Zhaoliang Chen, Lele Fu, Shunxin Xiao, Shiping Wang, Claudia Plant, and Wenzhong Guo. Multi-view graph convolutional networks with differentiable node selection. *ACM T-KDD*, 18(1):1–21, 2023.

[Chen *et al.*, 2023b] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95:109–119, 2023.

[Duan *et al.*, 2023] Jingcan Duan, Siwei Wang, Pei Zhang, En Zhu, Jingtao Hu, Hu Jin, Yue Liu, and Zhibin Dong. Graph anomaly detection via multi-scale contrastive learning networks with augmented view. In *AAAI*, volume 37, pages 7459–7467, 2023.

[Fey and Lenssen, 2019] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

[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.

[Fu *et al.*, 2024] Dazhi Fu, Zhao Zhang, and Jicong Fan. Dense projection for anomaly detection. In *AAAI*, volume 38, pages 8398–8408, 2024.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 27, 2014.

[Hinton *et al.*, 2015] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[Huang *et al.*, 2023] Wenke Huang, Guancheng Wan, Mang Ye, and Bo Du. Federated graph semantic and structural learning. In *IJCAI*, pages 139–143, 2023.

[Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[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.*, 2021a] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *CVPR*, pages 10713–10722, 2021.

[Li *et al.*, 2021b] 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 T-KDE*, 2021.

[Li *et al.*, 2021c] Rui Li, Xin Yuan, Mohsen Radfar, Peter Marendy, Wei Ni, Terrence J O'Brien, and Pablo M Casillas-Espinosa. Graph signal processing, graph neural network and graph learning on biological data: a systematic review. *IEEE Reviews in Biomedical Engineering*, 16:109–135, 2021.

[Li *et al.*, 2023a] Xingfeng Li, Zhenwen Ren, Quansen Sun, and Zhi Xu. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognition*, 134:109083, 2023.

[Li *et al.*, 2023b] Xingfeng Li, Yinghui Sun, Quansen Sun, Zhenwen Ren, and Yuan Sun. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100:101941, 2023.

[Liu *et al.*, 2022] Rui Liu, Pengwei Xing, Zichao Deng, Anran Li, Cuntai Guan, and Han Yu. Federated graph neural networks: Overview, techniques and challenges. *arXiv preprint arXiv:2202.07256*, 2022.

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

[Ma *et al.*, 2021] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE T-KDE*, 2021.

[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 *WSDM*, 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 *SIGKDD*, 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 *AISTATS*, pages 1273–1282. PMLR, 2017.

[Pang *et al.*, 2021] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

[Park *et al.*, 2019] Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. Relational knowledge distillation. In *CVPR*, pages 3967–3976, 2019.

[Pinelli *et al.*, 2023] Fabio Pinelli, Gabriele Tolomei, and Giovanni Trappolini. Flirt: Federated learning for information retrieval. In *SIGIR*, pages 3472–3475, 2023.

[Qiao *et al.*, 2023] Dong Qiao, Chris Ding, and Jicong Fan. Federated spectral clustering via secure similarity reconstruction. In *NeurIPS*, 2023.

[Qiu *et al.*, 2022] Chen Qiu, Marius Kloft, Stephan Mandt, and Maja Rudolph. Raising the bar in graph-level anomaly detection. In *IJCAI*, pages 2196–2203, 2022.

[Ruff *et al.*, 2018] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *ICML*, pages 4393–4402. PMLR, 2018.

[Sun and Fan, 2024] Yan Sun and Jicong Fan. Mmd graph kernel: Effective metric learning for graphs via maximum mean discrepancy. In *ICLR*, 2024.

[Sun *et al.*, 2023a] Yuan Sun, Zhenwen Ren, Peng Hu, Dezhong Peng, and Xu Wang. Hierarchical consensus hashing for cross-modal retrieval. *IEEE T-MM*, 2023.

[Sun *et al.*, 2023b] Yuan Sun, Xu Wang, Dezhong Peng, Zhenwen Ren, and Xiaobo Shen. Hierarchical hashing learning for image set classification. *IEEE T-IP*, 32:1732–1744, 2023.

[Sun *et al.*, 2024] Ziheng Sun, Chris Ding, and Jicong Fan. Lovász principle for unsupervised graph representation learning. *NeurIPS*, 36, 2024.

[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 *AAAI*, volume 37, pages 9953–9961, 2023.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(11):2579–2605, 2008.

[Wang *et al.*, 2022] Zhen Wang, Weirui Kuang, Yuexiang Xie, Liuyi Yao, Yaliang Li, Bolin Ding, and Jingren Zhou. Federatedscope-gnn: Towards a unified, comprehensive and efficient package for federated graph learning. In *SIGKDD*, pages 4110–4120, 2022.

[Wu *et al.*, 2024a] Zhihao Wu, Zhaoliang Chen, Shide Du, Sujia Huang, and Shiping Wang. Graph convolutional network with elastic topology. *Pattern Recognition*, page 110364, 2024.

[Wu *et al.*, 2024b] Zhihao Wu, Zhao Zhang, and Jicong Fan. Graph convolutional kernel machine versus graph convolutional networks. *NeurIPS*, 36, 2024.

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

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

[Zhang *et al.*, 2021] Huanding Zhang, Tao Shen, Fei Wu, Mingyang Yin, Hongxia Yang, and Chao Wu. Federated graph learning–a position paper. *arXiv preprint arXiv:2105.11099*, 2021.

[Zhang *et al.*, 2022] Chao Zhang, Huaxiong Li, Caihua Chen, Xiuyi Jia, and Chunlin Chen. Low-rank tensor regularized views recovery for incomplete multiview clustering. *IEEE T-NNLS*, 2022.

[Zhang *et al.*, 2023] Lei Zhang, Lele Fu, Tong Wang, Chuan Chen, and Chuanfu Zhang. Mutual information-driven multi-view clustering. In *CIKM*, pages 3268–3277, 2023.

[Zhang *et al.*, 2024a] Lei Zhang, Lele Fu, Chen Liu, Zhao Yang, Jinghua Yang, Zibin Zheng, and Chuan Chen. Towards few-label vertical federated learning. *ACM T-KDD*, 2024.

[Zhang *et al.*, 2024b] Yunhe Zhang, Yan Sun, Jinyu Cai, and Jicong Fan. Deep orthogonal hypersphere compression for anomaly detection. In *ICLR*, 2024.

[Zhao and Akoglu, 2021] Lingxiao Zhao and Leman Akoglu. On using classification datasets to evaluate graph outlier detection: Peculiar observations and new insights. *Big Data*, 2021.

[Zhao *et al.*, 2023] Yue Zhao, Guoqing Zheng, Subhabrata Mukherjee, Robert McCann, and Ahmed Awadallah. Admoe: Anomaly detection with mixture-of-experts from noisy labels. In *AAAI*, volume 37, pages 4937–4945, 2023.

[Zheng *et al.*, 2021] Yu Zheng, Ming Jin, Yixin Liu, Lianhua Chi, Khoa T Phan, and Yi-Ping Phoebe Chen. Generative and contrastive self-supervised learning for graph anomaly detection. *IEEE T-KDE*, 2021.