Graph Anomaly Detection via Feature Selection with Local Topological Residuals

Yazheng Zhao[1] , **Nannan Wu**[2*] , **Haoran Yin**[1] , **Yiming Zhao**[3]

[1]School of Computer Science and Technology, Tianjin University, Tianjin, China
[2]School of Cyber Science and Engineering, Tianjin University, Tianjin, China
[3]School of Artificial Intelligence, Tianjin University, Tianjin, China
{yazhengzhao, nannan.wu, yinhaoran, zhaoym}@tju.edu.cn

Abstract

Graph anomaly detection (GAD) aims to identify nodes that exhibit significant deviations from expected structural or attribute patterns, and has garnered increasing attention in recent years. Recent approaches for GAD have predominantly focused on local inconsistency mining, which refers to the difficulty of establishing high similarity relationships between anomalous nodes and their neighbors. While local inconsistency mining requires the incorporation of topological information, the use of Graph Neural Networks (GNNs) to introduce such information tends to homogenize connected nodes, thereby causing the loss of local anomalous signals. To address this challenge, we propose LTRGAD, a two-stage GAD framework that performs feature selection based on local feature-topological residuals (LTR). By processing features separately, LTRGAD effectively introduces topological information while preserving the original local anomalous patterns, enabling more accurate local anomaly detection. Subsequently, global anomaly detection is conducted on the entire graph by leveraging the results from the local detection phase. Extensive experiments on seven benchmark datasets demonstrate the effectiveness of the proposed LTRGAD framework.

1 Introduction

Due to its wide application in several domains, including financial fraud detection and intrusion detection in cybersecurity [Pourhabibi *et al.*, 2020; He *et al.*, 2022; Dong *et al.*, 2022; Huang *et al.*, 2022; Yang *et al.*, 2019], graph anomaly detection (GAD) has emerged as a highly prominent field within graph data mining. GAD aims to identify entities (e.g., nodes, edges or subgraphs) that deviate from normal patterns within a graph [Ma *et al.*, 2021; Pang *et al.*, 2021]. In this paper, our primary focus is on the detection of anomalous nodes. Recent works [Qiao and Pang, 2023; Chen *et al.*, 2024; Pan *et al.*, 2025] have been built on the concept of local inconsistency mining (LIM), which is grounded in the intuition that abnormal nodes tend to exhibit low similarity to their neighboring nodes. However, in the real-world anomaly dataset, normal nodes may also have a lower similarity to neighbor nodes [Qiao and Pang, 2023]. Under these circumstances, incorporating structural information to align the representations of benign nodes with those of their neighbors can effectively suppress their anomaly scores, thereby mitigating false positives. Consequently, it is imperative for LIM to leverage Graph Neural Networks (GNNs) to incorporate topological information to better distinguish these normal instances from true anomalies.

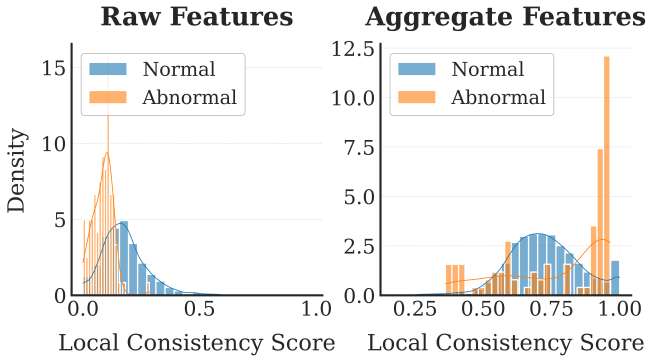

Figure 1: Local consistency is quantified as the mean similarity between a central node and its neighborhood. A comparative analysis is conducted on the Cora dataset, contrasting consistency scores calculated in the original feature space with those obtained after a single-layer neighborhood aggregation.

A key challenge lies the inherent conflict between the message passing of GNNs and the assumptions underlying LIM. As illustrated in Figure 1, the aggregation mechanism of GNNs tends to homogenize connected nodes [Chai *et al.*, 2022; Li *et al.*, 2019; Balcilar *et al.*, 2021], thereby increasing the consistency between a node and its neighborhood. This process effectively obfuscates local anomalous signals, making it significantly more challenging to distinguish local irregularities from their surrounding context.

While a multilayer perceptron (MLP) [Ramchoun *et al.*, 2016] could be employed to circumvent this conflict, such an approach disregards valuable anomalous signals embedded within the graph topology, ultimately resulting in limited applicability. Consequently, the wholesale abandon-

*Corresponding author.

ment of message passing is ill-advised, as it provides a vital mechanism for regularizing normal data. On the other hand, while attention mechanisms can be leveraged to mitigate message passing between anomalous nodes and their neighbors, such approaches are primarily effective only when anomalies exhibit pronounced dissimilarity to their local context. Since structural anomalies often manifest as cohesive clusters with similar attributes and synchronized behaviors, attention-based methods frequently fail to accurately detect these collective patterns.

Driven by these insights, we develop a new GAD framework aimed at harmonizing inherent conflict between neighborhood message passing and LIM mechanisms. In this paper, we propose local feature-topological residuals (LTR) to quantify the impact of aggregation operations on individual feature dimensions. Building upon this, we propose LTRGAD, a two-stage framework that performs local and global anomaly detection decoupled. In the local detection phase, LTRGAD leverages LTR to perform adaptive feature selection. Specifically, features minimally impacted by the aggregation operation are processed via a Graph Convolutional Network (GCN) [Kipf and Welling, 2016] to integrate structural context for structural anomaly detection, while the remaining features are fed into an MLP to detect primitive attribute anomaly. Through adaptive feature selection and specialized dual-path processing, LTRGAD effectively reconciles the inherent conflict between neighborhood message passing and LIM. Based on the resulting anomaly scores, we derive high-confidence normal and anomalous subsets. From the high-confidence normal set, we extract global normal contexts and normal prototype nodes to serve as references for the subsequent stage. During the global anomaly detection phase, LTRGAD employs a transformer-based architecture to facilitate long-range information propagation. LTRGAD implements a prototype-based querying mechanism, where each node retrieves information from its nearest prototypical node to construct a prototype query representation, which is subsequently integrated into a comprehensive global node representation. The representation space is optimized by maximizing the similarity between high-confidence normal nodes and the global normal context, while constraining their distance to prototypical nodes. Finally, anomaly scores are computed based on the nodes' deviation from the global context and their spatial distance to the prototypes. The final anomaly score is obtained by aggregating the scores from both stages.

The primary contributions of our paper are summarized as follows:

- We investigate the inherent challenges faced by current LIM-based GAD methods when handling GNN message passing. To address these, we introduce LTR as a metric to quantify the impact of neighborhood aggregation on individual feature dimensions. Subsequently, we design an adaptive feature selection mechanism guided by LTR to reconcile the conflict between the message passing of GNNs and the assumptions underlying LIM.
- From a global perspective, we introduce a novel prototypical querying mechanism integrated within a transformer-based framework. This mechanism retrieves representations by querying the relationship between a node and its nearest prototype node. The spatial distance between the node and these prototype nodes is subsequently utilized as a anomaly metric.
- Extensive experimental results on seven benchmark datasets, including four datasets injected with synthetic anomalies and three datasets with organic anomalies, demonstrate that our method achieves state-of-the-art performance compared to a series of baselines. Furthermore, ablation experiments confirm the effectiveness of feature selection based on LTR.

2 Related Work

Traditional GAD methods [Breunig *et al.*, 2000; Zhou *et al.*, 2021; Perozzi and Akoglu, 2016; Li *et al.*, 2017; Peng *et al.*, 2018] often underperform due to their limited capacity to integrate graph features and topology. Conversely, GNNs have revolutionized the field with superior representational power. While some supervised models exist [Dou *et al.*, 2020; Liu *et al.*, 2021a; Gao *et al.*, 2023; Zhang *et al.*, 2021], research primarily focuses on Unsupervised Graph Anomaly Detection (UGAD), which can be categorized into autoencoder-based methods, graph contrastive learning (GCL) methods, and LIM-based methods.

Autoencoder-based Methods. DOMINANT [Ding *et al.*, 2019a] established the graph autoencoders architecture, scoring anomalies via joint structural and attribute reconstruction error. Building on this, AnomalyDAE [Fan *et al.*, 2020] integrates an attention mechanism within the structural encoder to capture complex topological features, while ComGA [Luo *et al.*, 2022] integrates community aware mechanisms to mitigate intercommunity feature homogenization.

GCL-based Methods. CoLA [Liu *et al.*, 2021b] pioneered GCL for GAD by aligning node and local subgraph representations to derive similarity-based anomaly scores. Building on this, SL-GAD [Zheng *et al.*, 2021] integrates attribute regression with multi-view contrastive learning, while Sub-CR [Zhang *et al.*, 2022] combines multi-view contrastive learning and masked autoencoder reconstruction.

LIM-based Methods. While the motivations of some aforementioned methods are conceptually aligned with the LIM, more recent LIM-based GAD approaches specifically formulate anomaly scores based on the discrepancy between a node and its local neighborhood. TAM [Qiao and Pang, 2023] unveils the phenomenon of one-class homophily inherent in real-world anomaly datasets and formulates local node affinity as the primary metric for calculating anomaly scores. TAM addresses the conflict between message passing and LIM by pruning non-homophily edges. GADAM [Chen *et al.*, 2024] employs only an MLP for local anomaly detection to avoid message passing, and subsequently adjusts the message-passing framework based on the anomaly detection results. Both TAM and GADAM adjust their message-passing frameworks based on preliminary detection results. However, this dependency limits their performance across diverse datasets where initial predictions may be unreliable. In contrast, LTRGAD addresses the conflict between message

passing and LIM through adaptive feature selection, offering superior generalizability and effectiveness.

3 Preliminary

Attributed Graph. An attributed graph is formally denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ represents the set of n nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, and $\mathbf{X} \in \mathbb{R}^{n \times d}$ signifies the node attribute matrix. Specifically, the i-th row $\mathbf{X}_i \in \mathbb{R}^d$ of $\mathbf{X}$ characterizes the feature vector associated with node v_i. The topological structure of the graph is encoded by a binary adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$, where $\mathbf{A}_{i,j} = 1$ if $(v_i, v_j) \in \mathcal{E}$, and $\mathbf{A}_{i,j} = 0$ otherwise.

Problem Statement. Given an attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, the objective of GAD is to learn a scoring function $f : \mathcal{V} \to \mathbb{R}$ that assigns an anomaly score $\mathcal{S}_i = f(v_i)$ to each node $v_i \in \mathcal{V}$. This score quantifies the degree of deviation of node v_i from the normative patterns exhibited by the majority of nodes in the network. Nodes that significantly diverge from the consensus structural or attribute distributions are assigned higher scores, thereby facilitating their identification via a descending rank-order analysis.

4 Methodology

As shown in Figure 2, the overall workflow of LTRGAD consists of two stages. (1) Local Anomaly Detection Stage: This stage performs adaptive feature selection based on the LTR metric. Features are selectively channeled into either the MLP encoder or the GCN encoder to execute dual tasks for local anomaly detection. The output includes sets of high-confident normal and anomalous nodes, along with a global normal context and normal prototype nodes. (2) The Global Stage: This stage employs a transformer to facilitate global feature interaction. Building upon the transformer architecture, we compute the relationship between each node and its most relevant normal prototypes to generate prototype-query embeddings. These embeddings are then integrated to produce global node representations. Finally, the anomaly score is derived by considering both the node's similarity to the global normal context and its spatial distance from the prototype nodes. Next, we provide a detailed exposition of each stage within our proposed framework.

4.1 Local Feature Topological Residual

While the neighbor information brought by GNN aggregation is beneficial, it fundamentally conflicts with the mining of local inconsistency. To address this, we propose the local feature-topological residuals (LTR) to quantify the impact of GNN aggregation on various feature dimensions, formulated as follows:

$$\begin{aligned}\mathrm{LTR}^{(m)} = \frac{1}{|\mathcal{E}|} \sum_{(v_i,v_j)\in\mathcal{E}} \Big[&\mathrm{Sim}(\tilde{\mathbf{X}}_i^{(m)}, \tilde{\mathbf{X}}_j^{(m)}) \\ &- \mathrm{Sim}(\mathbf{X}_i^{(m)}, \mathbf{X}_j^{(m)}) \Big],\end{aligned} \tag{1}$$

where $\mathrm{Sim}(\cdot)$ denotes the cosine similarity function. $\tilde{\mathbf{X}} = (\hat{\mathbf{A}})^k \mathbf{X}$ represents the aggregated node feature matrix obtained by propagating information across k-hop neighborhoods, and $\tilde{\mathbf{X}}_i^{(m)}$ corresponds to the aggregated feature vector of the i-th node under the m-th attribute view. And $\hat{\mathbf{A}}$ is the symmetrically re-normalized adjacency matrix, defined as $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}$, where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and $\tilde{\mathbf{D}}$ denotes the corresponding degree matrix. In the subsequent experiments, the value of k is fixed at 2.

A higher value of $\mathrm{LTR}^{(m)}$ indicates that the m-th feature undergoes a more pronounced consistency transformation due to the aggregation operation, which proves detrimental to the efficacy of the anomaly detection task. Conversely, a lower $\mathrm{LTR}^{(m)}$ suggests that the m-th feature exhibits lower sensitivity to aggregation, thereby allowing for the beneficial integration of topological information through the aggregation process.

4.2 LTR-Based Local Anomaly Detection

For each feature $\mathbf{X}^{(m)}$, we calculate its LTR according to Eq. (1), followed by a normalization step to account for variations across diverse datasets. We then employ a threshold-based filtering mechanism to partition the features into two disjoint subsets, denoted as $\mathbf{X}_{low}$ and $\mathbf{X}_{high}$. Specifically, features satisfying $\mathrm{LTR} < \lambda$ are assigned to $\mathbf{X}_{low}$ and processed via a GCN to incorporate topological information. Conversely, the remaining features, denoted as $\mathbf{X}_{high}$, are handled by an MLP to identify original attribute anomalies.

Attribute Anomaly Detection via Contrastive Learning
Normal nodes typically exhibit high attribute similarity with their neighbors, whereas anomalous nodes manifest lower levels of such affinity. Leveraging this disparity, we employ contrastive learning to quantify the similarity between a node and its local neighborhood, thereby facilitating the identification of attribute anomalous.

MLP as encoder. We employ an L-layer MLP to obtain the node embedding matrix, as formulated in Eq. (2). Utilizing the MLP as an encoder preserves the independence of node attribute representations, thereby circumventing the over-smoothing issue typically induced by message-passing mechanisms.

$$\mathbf{H}_{MLP}^{(l)} = \sigma(\mathbf{H}_{MLP}^{(l-1)}\mathbf{W}^{(l-1)} + \mathbf{b}^{(l-1)}), \tag{2}$$

where $\mathbf{H}_{MLP}^{(0)} = \mathbf{X}_{high}$ and $\sigma(\cdot)$ is an activation function. The node representations yielded by the terminal MLP layer serve as the finalized attribute embeddings $\mathbf{Z}^{attr} = \mathbf{H}_{MLP}^{(L)}$.

Contrastive Learning. Motivated by the observation that normal nodes exhibit a high degree of similarity with their neighbors, we designate the first-order neighborhood $\mathcal{N}_i$ of node v_i as its positive sample set. Correspondingly, the negative sample set is constructed from the first-order neighbors of a randomly sampled non-neighboring node v_j (where v_j is not connected to v_i). Specifically, for a target node v_i, we define its positive representation $\mathbf{E}_i^P$ and negative representation $\mathbf{E}_i^N$ as follows:

$$\begin{aligned}\mathbf{E}_i^P &= \mathrm{Mean}\left(\{\mathbf{Z}_j^{attr} \mid v_j \in \mathcal{N}_i\}\right), \\ \mathbf{E}_i^N &= \mathrm{Mean}\left(\{\mathbf{Z}_p^{attr} \mid v_p \in \mathcal{N}_q, v_q \notin \mathcal{N}_i \cup \{v_i\}\}\right),\end{aligned} \tag{3}$$

where $\mathbf{Z}_j^{attr} = \mathbf{Z}^{attr}[j, :]$.

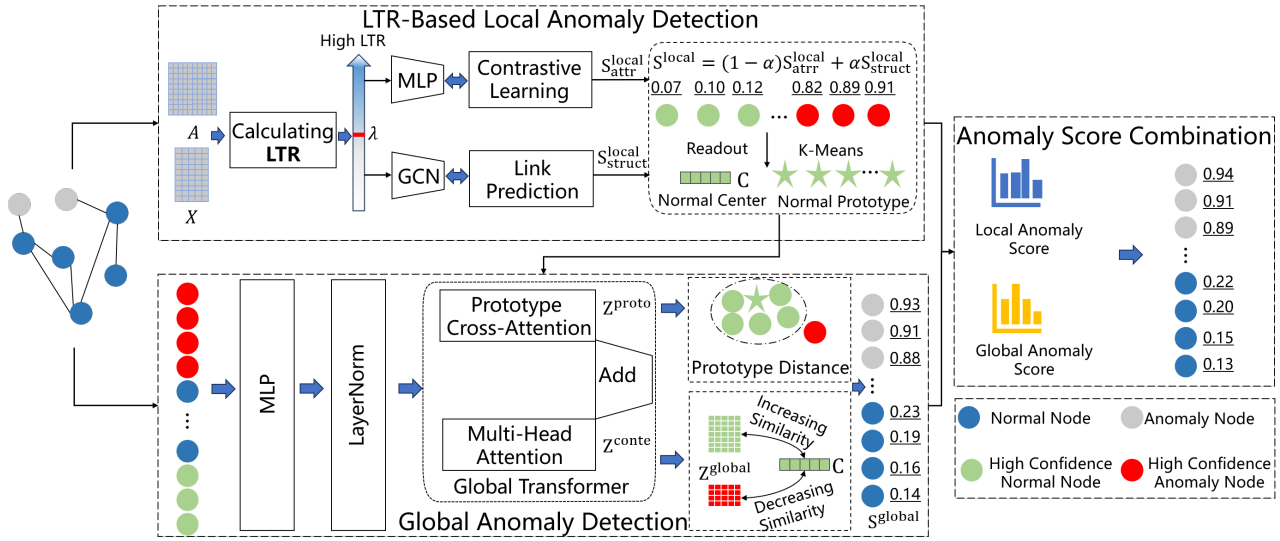

Figure 2: Architecture of the proposed LTRGAD. In the local stage, LTR is computed from the adjacency and feature matrices to guide feature partitioning. High-LTR features are sent to an MLP for attribute anomaly detection via contrastive learning, while low-LTR features are processed by a GCN for structural anomaly detection via link prediction. In the global stage, a transformer propagates information across the graph, and the model is optimized using global context alignment and prototype-based regularization to produce the final anomaly score.

To quantify the similarity between the target node and its corresponding positive and negative instances, we employ the inner product as the similarity discriminator, which is equivalent to cosine similarity after L2 normalization:

$$\begin{aligned} s_i^P &= Dis(\mathbf{Z}_i^{attr}, \mathbf{E}_i^P) = \mathbf{Z}_i^{attr}(\mathbf{E}_i^P)^T, \\ s_i^N &= Dis(\mathbf{Z}_i^{attr}, \mathbf{E}_i^N) = \mathbf{Z}_i^{attr}(\mathbf{E}_i^N)^T \end{aligned} \tag{4}$$

To distinguish anomalous nodes from normal ones, we employ the Binary Cross-Entropy (BCE) loss to optimize. Specifically, the objective is to maximize the similarity scores of positive instance pairs while minimizing those of negative instance pairs, thereby pulling normal nodes closer to their local context and pushing anomalous nodes away.

$$\mathcal{L}_{attr} = -\sum_{i=1}^{|\mathcal{V}|} \log(s_i^P) + \log(1 - s_i^N) \tag{5}$$

Then, the similarity score of the positive instance pair, denoted as s_i^P, is utilized as the raw attribute anomaly score:

$$\mathcal{S}_{attr}^{local}(v_i) = -Dis(\mathbf{Z}_i^{attr}, \mathbf{E}_i^P) \tag{6}$$

Structural Anomaly Detection via Link Prediction

Normal nodes tend to form robust and cohesive connections with their peers, reflecting a high degree of structural homophily. LTRGAD introduces the link reconstruction task as a self-supervised signal to compel the model to extract generalizable connectivity patterns directly from the graph topology. Intuitively, if the ground-truth edges of a node consistently receive low prediction confidence during the reconstruction process, it signifies that its structural relationships are poorly explained by the normal structural patterns captured by the encoder. Consequently, such a discrepancy indicates a high degree of structural anomalousness for the corresponding node.

GCN as encoder. LTRGAD integrates topological information via GCNs and employ a link reconstruction task to identify structural anomalies. Formally, the node representations at the l-th GCN layer can be generally expressed as:

$$\mathbf{H}_{GCN}^{(l)} = \phi\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}_{GCN}^{(l-1)}\mathbf{W}^{(l-1)}\right), \tag{7}$$

where $\phi(\cdot)$ denotes a non-linear activation function. $\mathbf{H}_{GCN}^{(l-1)}$ denotes the node embedding matrix from the preceding layer, with the initial input set as $\mathbf{H}_{GCN}^{(0)} = \mathbf{X}_{low}$.

Link Prediction. The node representations yielded by the terminal GCN layer serve as the finalized structural embeddings $\mathbf{Z}^{struct} = \mathbf{H}_{\text{GCN}}^{(L)}$. To model interactions between node pairs, we utilize the Hadamard product [Horn, 1990] as the interaction operator and employ a MLP as the decoder. For any node pair (v_i, v_j), the link prediction logit $\hat{y}_{ij}$ is computed as:

$$\hat{y}_{ij} = \text{MLP}(\mathbf{z}_i \odot \mathbf{z}_j), \tag{8}$$

where $\mathbf{z}_i, \mathbf{z}_j \in \mathbf{Z}^{struct}$ and $\odot$ denotes the element-wise product. $\hat{y}_{ij} \in \mathbb{R}$ serves as an unnormalized logit that quantifies the structural compatibility between the two nodes.

To provide supervision for the link reconstruction task, we construct the training sample set by designating all observed edges in the original graph as the positive set $\mathcal{E}_+$, while the negative set $\mathcal{E}_-$ is formed by uniformly sampling non-adjacent node pairs at random. We optimize the link reconstruction task using the BCE loss:

$$\begin{aligned} \mathcal{L}_{struct} = &- \mathbb{E}_{(v_i,v_j)\in\mathcal{E}_+} \left[\log \sigma(\hat{y}_{ij})\right] \\ &- \mathbb{E}_{(v_i,v_k)\in\mathcal{E}_-} \left[\log(1 - \sigma(\hat{y}_{ik}))\right], \end{aligned} \tag{9}$$

where $\sigma(\cdot)$ denotes the Sigmoid function.

We define the structural anomaly score of a node based on the link reconstruction consistency within its true neighborhood. Specifically, for a node v_i, the structural anomaly score is computed as:

$$\mathcal{S}_{struct}^{local}(v_i) = -\frac{1}{|\mathcal{N}(v_i)|} \sum_{v_j \in \mathcal{N}(v_i)} \hat{y}_{ij} \tag{10}$$

Local Anomaly Detection

The final local node representation is obtained by concatenating the embeddings derived from the two respective branches:

$$\mathbf{Z}^{local} = [\mathbf{Z}^{attr} \parallel \mathbf{Z}^{struct}], \tag{11}$$

where $\parallel$ denotes the concatenation operator.

To jointly optimize the local anomaly detection module, the total training objective is defined as the sum of the attribute loss and the structural loss:

$$\mathcal{L}_{local} = \mathcal{L}_{attr} + \mathcal{L}_{struct} \tag{12}$$

To aggregate the diagnostic information from dual perspectives, the final local anomaly score S^{local} is formulated as a weighted fusion of the attribute and structural components:

$$\mathcal{S}^{local} = (1-\alpha)\mathcal{S}_{attr}^{local} + \alpha \mathcal{S}_{struct}^{local}, \tag{13}$$

where $\alpha \in [0, 1]$ is a balancing coefficient. Based on the fused scores, we rank all nodes and identify the top $k_{nor}\%$ of nodes with the lowest scores to constitute a high-confidence normal set $\mathcal{V}_n$. To capture the representative characteristics of the normal class, a global normal context $\mathbf{C}$ is derived by averaging the local representations of nodes in $\mathcal{V}_n$:

$$\mathbf{C} = \frac{1}{|\mathcal{V}_n|} \sum_{v_i \in \mathcal{V}_n} \mathbf{Z}_i^{local} \tag{14}$$

Symmetrically, we select $k_{ano}\%$ of nodes with the highest anomaly scores to form a high-confidence abnormal set $\mathcal{V}_a$. To further account for the multi community distribution of normal patterns, we perform K-means clustering [MacQueen, 1967] on $\mathcal{V}_n$ to extract K distinctive normal prototype nodes $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_K\}$. These prototype nodes, together with the global context $\mathbf{C}$ and the two high-confidence sets, serve as refined supervised signals to guide the subsequent global anomaly detection stage.

4.3 Global Anomaly Detection

While effective in capturing neighborhood-level irregularities, the local detection module is inherently constrained by its finite receptive field, which may impede the identification of anomalies with global-scale deviations. To bridge this gap, we initiate a second stage of analysis that capitalizes on the high-confidence normal and abnormal priors distilled from the local phase. This stage enables the model to conduct a more comprehensive anomaly assessment from a holistic global perspective.

Global Contextual Feature Extraction. To capture long-range dependencies, we employ a transformer architecture that operates directly on the raw node features $\mathbf{X}$. We deliberately bypass the local embeddings generated in the first stage and utilize raw attributes as input to circumvent the over-smoothing issue. By re-engaging with the raw feature space, the global branch can extract distinct structural and semantic signals that are independent of the local topological bias.

We project $\mathbf{X}$ into Query $\mathbf{Q} = \mathbf{X}\mathbf{W}_Q$, Key $\mathbf{K} = \mathbf{X}\mathbf{W}_K$, and Value $\mathbf{V} = \mathbf{X}\mathbf{W}_V$ spaces via standard linear transformations. To ensure computational efficiency on large-scale graphs, we adopt a linear attention mechanism [Wu *et al.*, 2022]. The global contextual representation is formulated as:

$$\mathbf{Z}^{conte} = \frac{\phi(\mathbf{Q})(\phi(\mathbf{K})^T\mathbf{V})}{\phi(\mathbf{Q})\phi(\mathbf{K})^T\mathbf{1}}, \tag{15}$$

where $\phi(\cdot)$ is a non-negative feature mapping function.

Normality Alignment via Prototype Querying. We utilize the normal prototype nodes $\mathbf{P}$ obtained from the local anomaly detection stage to explicitly anchor the representation of normal data. Normal prototype query representation is formulated as:

$$\mathbf{A}^{proto} = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{P}^T}{\sqrt{d}}\right), \mathbf{Z}^{proto} = \mathbf{A}^{proto}\mathbf{P}, \tag{16}$$

where $\mathbf{A}^{proto}$ denotes the cross-attention weights between nodes and prototype nodes. The final global node embedding is formulated as the additive fusion of the contextual information and the normal prototype query:

$$\mathbf{Z}^{global} = \mathbf{Z}^{conte} + \mathbf{Z}^{proto} \tag{17}$$

Inspired by the observation that anomalies tend to exhibit a greater divergence from the global normal context $\mathbf{C}$ than normal nodes [Chen *et al.*, 2022], we quantify the degree of alignment by computing the similarity between node embeddings and the global center:

$$\mathcal{L}_{align} = -\frac{1}{|\mathcal{V}_n| + |\mathcal{V}_a|} \sum_{v_i \in \mathcal{V}_n} \log(g_i) + \sum_{v_j \in \mathcal{V}_a} \log(1 - g_j), \tag{18}$$

where $g_i = \text{Dis}(\mathbf{Z}_i^{global}, \mathbf{C}) = \mathbf{Z}_i^{global}\mathbf{C}^T$.

Furthermore, in accordance with the principle that normal nodes should align with their respective community-level normality [Francisquini *et al.*, 2022], we quantify the structural anomaly by measuring the Euclidean distance between the fused node embedding and its prototype-based query:

$$\mathcal{L}_{distance} = \sum_{u_i \in \mathcal{V}} \|\mathbf{Z}_i^{global} - \mathbf{Z}_i^{proto}\|^2 \tag{19}$$

The loss function of the global module is defined as:

$$\mathcal{L}_{global} = \mathcal{L}_{align} + \mathcal{L}_{distance} \tag{20}$$

The normality of a node is characterized by the extent to which its representation aligns with the global normal context $\mathbf{C}$. Similarly, a node is considered more normal if its representation lies closer to the learned normal prototype queries in the latent space. Based on these principles, we define the final global anomaly score as:

$$S^{global}(v_i) = -(\mathbf{Z}_i^{global}\mathbf{C}^T - |\mathbf{Z}_i^{global} - \mathbf{Z}_i^{proto}\|^2) \tag{21}$$

Ultimately, the final anomaly score is derived by fusing $\mathcal{S}^{local}$ and $\mathcal{S}^{global}$ using $\beta \in [0, 1]$ as a balancing factor to yield a comprehensive assessment:

$$\mathcal{S} = (1-\beta)\mathcal{S}^{local} + \beta\mathcal{S}^{global} \tag{22}$$

Method	Cora	Citeseer	Pubmed	ACM	Books	Disney	Enron
DOMINANT (2019)	0.8493	0.8391	0.8013	0.7452	0.5012	0.4873	0.6759
AnomalyDAE (2020)	0.8431	0.8264	0.8973	0.7516	0.5567	0.3234	0.7132
CoLA (2021)	0.8801	0.8891	0.9535	0.7783	0.3982	0.4237	<u>0.7245</u>
SL-GAD (2021)	0.8983	0.9106	0.9476	0.8538	0.5655	0.4703	0.5518
Sub-CR (2022)	0.9132	0.9310	<u>0.9629</u>	0.7245	0.5713	<u>0.5424</u>	0.6541
ComGA (2022)	0.8840	0.9167	0.9212	0.8496	0.5354	0.4831	0.7222
TAM (2023)	0.8907	0.7611	0.9321	0.4721	0.3774	0.4773	0.4531
GADAM (2024)	<u>0.9556</u>	<u>0.9415</u>	0.9337	<u>0.9567</u>	<u>0.5983</u>	0.4288	0.4483
HUGE (2025)	0.8778	0.8968	0.9456	0.9228	0.5465	0.3898	0.3207
LTRGAD	**0.9615**	**0.9675**	**0.9688**	**0.9696**	**0.7175**	**0.7020**	**0.8201**

Table 1: ROC-AUC comparison on seven datasets. The best and second-best results are in **bold** and <u>underlined</u>, respectively.

5 Experiments

5.1 Experimental Settings

Datasets. To comprehensively evaluate the proposed model, we conduct experiments on seven benchmark datasets. (1) Four widely used benchmark datasets: Cora, Citeseer, Pubmed [Sen *et al.*, 2008], and ACM [Tang *et al.*, 2008]. Given the absence of inherent anomalies in these graphs, we adopt the synthetic injection protocol established in prior work [Song *et al.*, 2007; Ding *et al.*, 2019b; Ding *et al.*, 2019a]. (2) Three real-world datasets that contain organic anomalies: Books [Sánchez *et al.*, 2013], Disneyc [Müller *et al.*, 2013], and Enron [Sánchez *et al.*, 2013]. Detailed descriptions of the datasets and anomaly injection approach are summarized in the supplementary material.

Baseline. To evaluate the performance, we benchmark our approach against nine representative methods. DOMINANT [Ding *et al.*, 2019a], AnomalyDAE [Fan *et al.*, 2020] and ComGA [Luo *et al.*, 2022] leverage autoencoder architectures to facilitate GAD. CoLA [Liu *et al.*, 2021b], SL-GAD [Zheng *et al.*, 2021], and Sub-CR [Zhang *et al.*, 2022] represent GAD approaches grounded in graph contrastive learning. TAM [Qiao and Pang, 2023], GADAM [Chen *et al.*, 2024] and HUGE [Pan *et al.*, 2025] are GAD methods that employ local inconsistency mining. We provide more details for these methods in the supplementary material.

Evaluation metrics. We evaluate the proposed framework and baseline methods using ROC-AUC, a widely adopted metric for anomaly detection [Ma *et al.*, 2021]. The ROC curve illustrates the trade-off between the true positive rate and false positive rate based on ground-truth labels. The AUC score represents the probability that an anomalous node is ranked higher than a normal one, with values closer to 1 indicating better performance.

Implementation details. We set the number of MLP layers in the local module to 1 and the GCN layer to 2. We set $k_{\text{nor}}\% = 50\%$ and $k_{\text{ano}}\% = 5\%$ for all datasets. It is noteworthy that LTRGAD exhibits strong robustness to the hyperparameters K and β. Consequently, we set $K = 10$ and $\beta = 0.5$ as the default values for our experiments. Both the local and global modules are trained for 100 epochs. Other implementation details are provided in the supplementary material.

5.2 Result and Analysis

As illustrated in Table 1, LTRGAD demonstrates a significant performance advantage over nine representative baselines in terms of ROC-AUC. LTRGAD achieves substantial improvements on challenging real-world anomaly datasets. Specifically, LTRGAD achieves absolute ROC-AUC improvements of 15.96, 11.92, and 9.56 percentage points on the Disney, Books, and Enron datasets, respectively. These results clearly demonstrate the superior capability of LTRGAD in effectively capturing and identifying real-world anomalies. LTRGAD demonstrates great compatibility across diverse graph domains, consistently achieving superior performance on datasets characterized by both injected and real-world anomaly patterns. This robustness distinguishes LTRGAD from existing paradigms that often struggle to maintain efficacy across varying anomaly distributions.

5.3 Ablation Study

To verify the contribution of each component within the LTRGAD framework, we conducted an ablation study by comparing the full model against four variants: (1) LTRGAD w/o Global, which removes the global detection stage; (2) LTRAGD w/o FS, which removes the feature selection module and feeds all features into the MLP encoder and GCN encoder simultaneously; (3) $\text{LTRGAD}_{\text{MLP}}$ w/o FS, which removes the feature selection module and directly feeds all features to an MLP encoder; and (4) $\text{LTRGAD}_{\text{GCN}}$ w/o FS, which similarly the removes feature selection module and uses a GCN encoder. The results in the Table 2 verify the importance of two key components of LTRGAD:

- **Global Module.** The complete LTRGAD consistently outperforms LTRGAD w/o Global across all datasets. This performance gap is particularly evident on the Books and Enron datasets, indicating that the long-distance dependencies and global context captured by the global module are beneficial for anomaly detection.
- **Feature Selection Module.** In the absence of feature selection, we observe a significant performance decline across all datasets, which validates the effectiveness of our feature selection module. Furthermore, we observed a difference in performance when using MLP encoders and GCN encoders independently within the local mod-

Method	Cora	Citeseer	ACM	Books	Disney	Enron
LTRGAD w/o Global	0.9531	<u>0.9540</u>	0.9468	<u>0.6522</u>	<u>0.6787</u>	<u>0.7655</u>
LTRGAD w/o FS	0.9551	0.9038	0.9612	0.5906	0.5056	0.5699
$\text{LTRGAD}_{\text{MLP}}$ w/o FS	<u>0.9556</u>	0.9025	<u>0.9657</u>	0.4399	0.5607	0.3887
$\text{LTRGAD}_{\text{GCN}}$ w/o FS	0.5042	0.6285	0.5194	0.6257	0.6511	0.6027
LTRGAD	**0.9615**	**0.9675**	**0.9696**	**0.7175**	**0.7020**	**0.8201**

Table 2: Ablation study results of LTRGAD on six datasets. The best and second-best results are in **bold** and <u>underlined</u>, respectively.

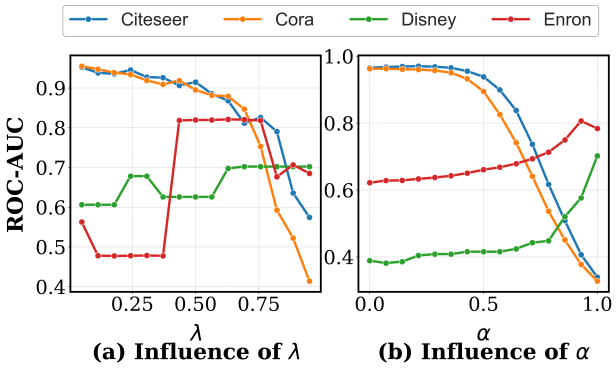

Figure 3: Influence of threshold λ and balancing factor α.

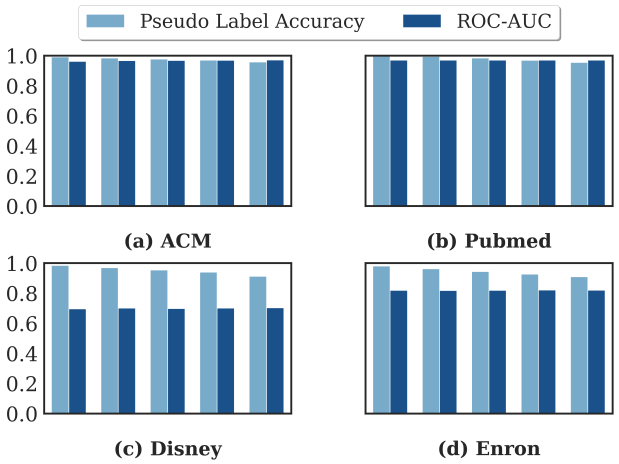

Figure 4: Pseudo label accuracy of high-confidence sets and the model performance of LTRGAD with different $k_{ano}\%$ on four benchmark datasets.

ule. Specifically, the $\text{LTRGAD}_{\text{MLP}}$ w/o FS exhibits superior performance on datasets with injected anomalies, whereas $\text{LTRGAD}_{\text{GCN}}$ w/o FS proves more effective on real-world anomaly datasets. This phenomenon indicates that injected anomalies are susceptible to the noise propagation inherent in GCNs, whereas real-world anomalies, characterized by strong attribute camouflage, necessitate the structural context provided by GCNs. Crucially, the full LTRGAD model consistently outperforms both variants across all benchmarks. This validates that our feature selection strategy provides LTRGAD with superior compatibility, making it equally effective for injected anomalies datasets and real world anomalies datasets.

5.4 Parameter Study

Thresholds λ and Balancing Factors α. As shown in figure 3, we investigate the impact of different thresholds λ and balancing factors α on the proposed LTRGAD framework across four datasets: Cora, Citeseer, Disney, and Enron. Specifically, for datasets featuring injected anomalies (Cora and Citeseer), the optimal ROC-AUC is typically achieved at lower values of λ and α. This suggests that the model benefits from a higher sensitivity to attribute-level deviations to capture artificial perturbations. Conversely, on real-world anomaly datasets (Disney and Enron) the model yields superior performance with larger λ and α. This phenomenon indicates that while naturally occurring anomalies may exhibit attribute distributions similar to normal instances, they often manifest more pronounced irregularities in their structural connectivity, necessitating a parameter configuration that prioritizes structural incongruity.

High-Confidence Anomaly Ratio $k_{ano}\%$. To evaluate the impact of the two high-confidence sets on LTRGAD, we conducted a sensitivity analysis of model performance and pseudo-label accuracy across various $k_{ano}\%$ values. Notably, as normal nodes constitute the vast majority of the dataset, we omitted the reporting of $k_{nor}\%$. As illustrated in Figure 4, the consistently high accuracy of the pseudo-labels validates the effectiveness of our local model. Furthermore, LTRGAD demonstrates strong robustness across different $k_{ano}\%$ settings. Consequently, in our practical implementation, we set $k_{\text{nor}}\% = 50\%$ and $k_{\text{ano}}\% = 5\%$ for all datasets.

6 Conclusion

In this paper, we investigate the inherent conflict between the message-passing mechanism of GNNs and the underlying assumptions of the LIM. To address this challenge, we propose a novel framework, LTRGAD, which employs adaptive feature selection guided by LTR to process features through dual paths, thereby effectively reconciling this conflict. Furthermore, LTRGAD incorporates a global module that leverages global normal context and normal prototypes to identify anomalies beyond the local perspective. Extensive experimental results demonstrate the superiority of LTRGAD, while ablation studies provide deeper insights into its inner workings. We hope our work provides insights for detecting anomalies in complex real-world scenarios.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Tianjin under Grant 25JCZDJC01300, and in part by the National Natural Science Foundation of China under Grant U2468204.

References

[Balcilar *et al.*, 2021] Muhammet Balcilar, Guillaume Renton, Pierre Héroux, Benoit Gaüzère, Sébastien Adam, and Paul Honeine. Analyzing the expressive power of graph neural networks in a spectral perspective. In *International conference on learning representations*, 2021.

[Breunig *et al.*, 2000] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM SIGMOD Conference*, 2000.

[Chai *et al.*, 2022] Ziwei Chai, Siqi You, Yang Yang, Shiliang Pu, Jiarong Xu, Haoyang Cai, and Weihao Jiang. Can abnormality be detected by graph neural networks? In *IJCAI*, pages 1945–1951, 2022.

[Chen *et al.*, 2022] Bo Chen, Jing Zhang, Xiaokang Zhang, Yuxiao Dong, Jian Song, Peng Zhang, Kaibo Xu, Evgeny Kharlamov, and Jie Tang. Gccad: Graph contrastive coding for anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8037–8051, 2022.

[Chen *et al.*, 2024] Jingyan Chen, Guanghui Zhu, Chunfeng Yuan, and Yihua Huang. Boosting graph anomaly detection with adaptive message passing. In *The Twelfth International Conference on Learning Representations*, 2024.

[Ding *et al.*, 2019a] Kaize Ding, Jundong Li, Rohit Bhanushali, and Huan Liu. Deep anomaly detection on attributed networks. In *Proceedings of the 2019 SIAM international conference on data mining*, pages 594–602. SIAM, 2019.

[Ding *et al.*, 2019b] Kaize Ding, Jundong Li, and Huan Liu. Interactive anomaly detection on attributed networks. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 357–365, 2019.

[Dong *et al.*, 2022] Linfeng Dong, Yang Liu, Xiang Ao, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Bilevel selection via meta gradient for graph-based fraud detection. In *International conference on database systems for advanced applications*, pages 387–394. Springer, 2022.

[Dou *et al.*, 2020] Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S Yu. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 315–324, 2020.

[Fan *et al.*, 2020] Haoyi Fan, Fengbin Zhang, and Zuoyong Li. Anomalydae: Dual autoencoder for anomaly detection on attributed networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5685–5689. IEEE, 2020.

[Francisquini *et al.*, 2022] Rodrigo Francisquini, Ana Carolina Lorena, and Mariá CV Nascimento. Community-based anomaly detection using spectral graph filtering. *Applied Soft Computing*, 118:108489, 2022.

[Gao *et al.*, 2023] Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM web conference 2023*, pages 1528–1538, 2023.

[He *et al.*, 2022] Li He, Guandong Xu, Shoaib Jameel, Xianzhi Wang, and Hongxu Chen. Graph-aware deep fusion networks for online spam review detection. *IEEE Transactions on Computational Social Systems*, 10(5):2557–2565, 2022.

[Horn, 1990] Roger A Horn. The hadamard product. In *Proc. symp. appl. math*, volume 40, pages 87–169, 1990.

[Huang *et al.*, 2022] Mengda Huang, Yang Liu, Xiang Ao, Kuan Li, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Auc-oriented graph neural network for fraud detection. In *Proceedings of the ACM web conference 2022*, pages 1311–1321, 2022.

[Kipf and Welling, 2016] Thomas Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *ArXiv*, abs/1609.02907, 2016.

[Li *et al.*, 2017] Jundong Li, Harsh Dani, Xia Hu, and Huan Liu. Radar: Residual analysis for anomaly detection in attributed networks. In *IJCAI*, volume 17, pages 2152–2158, 2017.

[Li *et al.*, 2019] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9267–9276, 2019.

[Liu *et al.*, 2021a] Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021*, pages 3168–3177, 2021.

[Liu *et al.*, 2021b] Yixin Liu, Zhao Li, Shirui Pan, Chen Gong, Chuan Zhou, and George Karypis. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE transactions on neural networks and learning systems*, 33(6):2378–2392, 2021.

[Luo *et al.*, 2022] Xuexiong Luo, Jia Wu, Amin Beheshti, Jian Yang, Xiankun Zhang, Yuan Wang, and Shan Xue. Comga: Community-aware attributed graph anomaly detection. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 657–665, 2022.

[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 transactions on knowledge and data engineering*, 35(12):12012–12038, 2021.

[MacQueen, 1967] J. MacQueen. Some methods for classification and analysis of multivariate observations. 1967.

[Müller *et al.*, 2013] Emmanuel Müller, Patricia Iglesias Sánchez, Yvonne Mülle, and Klemens Böhm. Ranking outlier nodes in subspaces of attributed graphs. *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 216–222, 2013.

[Pan *et al.*, 2025] Junjun Pan, Yixin Liu, Xin Zheng, Yizhen Zheng, Alan Wee-Chung Liew, Fuyi Li, and Shirui Pan. A label-free heterophily-guided approach for unsupervised graph fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12443–12451, 2025.

[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.

[Peng *et al.*, 2018] Zhen Peng, Minnan Luo, Jundong Li, Huan Liu, Qinghua Zheng, et al. Anomalous: A joint modeling approach for anomaly detection on attributed networks. In *IJCAI*, volume 18, pages 3513–3519, 2018.

[Perozzi and Akoglu, 2016] Bryan Perozzi and Leman Akoglu. Scalable anomaly ranking of attributed neighborhoods. In *Proceedings of the 2016 SIAM international conference on data mining*, pages 207–215. SIAM, 2016.

[Pourhabibi *et al.*, 2020] Tahereh Pourhabibi, Kok-Leong Ong, Booi H Kam, and Yee Ling Boo. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133:113303, 2020.

[Qiao and Pang, 2023] Hezhe Qiao and Guansong Pang. Truncated affinity maximization: One-class homophily modeling for graph anomaly detection. *Advances in Neural Information Processing Systems*, 36:49490–49512, 2023.

[Ramchoun *et al.*, 2016] Hassan Ramchoun, Youssef Ghanou, Mohamed Ettaouil, and Mohammed Amine Janati Idrissi. Multilayer perceptron: Architecture optimization and training. 2016.

[Sánchez *et al.*, 2013] Patricia Iglesias Sánchez, Emmanuel Müller, Fabian Laforet, Fabian Keller, and Klemens Böhm. Statistical selection of congruent subspaces for mining attributed graphs. In *2013 IEEE 13th international conference on data mining*, pages 647–656. IEEE, 2013.

[Sen *et al.*, 2008] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

[Song *et al.*, 2007] Xiuyao Song, Mingxi Wu, Christopher Jermaine, and Sanjay Ranka. Conditional anomaly detection. *IEEE Transactions on knowledge and Data Engineering*, 19(5):631–645, 2007.

[Tang *et al.*, 2008] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998, 2008.

[Wu *et al.*, 2022] Qitian Wu, Wentao Zhao, Zenan Li, David P Wipf, and Junchi Yan. Nodeformer: A scalable graph structure learning transformer for node classification. *Advances in Neural Information Processing Systems*, 35:27387–27401, 2022.

[Yang *et al.*, 2019] Yang Yang, Yuhong Xu, Yizhou Sun, Yuxiao Dong, Fei Wu, and Yueting Zhuang. Mining fraudsters and fraudulent strategies in large-scale mobile social networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(1):169–179, 2019.

[Zhang *et al.*, 2021] Ge Zhang, Jia Wu, Jian Yang, Amin Beheshti, Shan Xue, Chuan Zhou, and Quan Z Sheng. Fraudre: Fraud detection dual-resistant to graph inconsistency and imbalance. In *2021 IEEE international conference on data mining (ICDM)*, pages 867–876. IEEE, 2021.

[Zhang *et al.*, 2022] Jiaqiang Zhang, Senzhang Wang, and Songcan Chen. Reconstruction enhanced multi-view contrastive learning for anomaly detection on attributed networks. *arXiv preprint arXiv:2205.04816*, 2022.

[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 Transactions on Knowledge and Data Engineering*, 35(12):12220–12233, 2021.

[Zhou *et al.*, 2021] Shuang Zhou, Qiaoyu Tan, Zhiming Xu, Xiao Huang, and Fu-lai Chung. Subtractive aggregation for attributed network anomaly detection. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3672–3676, 2021.