Addressing the Distortion of Community Representations in Anomaly Detection on Attributed Networks

Enbo He*
Harbin Engineering University
Harbin, Heilongjiang, China
enochhe@hrbeu.edu.cn

Yitong Hao*
Harbin Engineering University
Harbin, Heilongjiang, China
haoyitong@hrbeu.edu.cn

Yue Zhang
Harbin Engineering University
Harbin, Heilongjiang, China
zycg87@sina.com

Guisheng Yin
Harbin Engineering University
Harbin, Heilongjiang, China
yinguishengabc@163.com

Lina Yao
CSIRO's Data61
Sydney, NSW, Australia
The University of New South Wales
Sydney, NSW, Australia
lina.yao@unsw.edu.au

ABSTRACT

Anomaly detection on attributed networks, especially in the unsupervised scenario, has garnered significant attention. And the Contrastive Learning (CL)-based methods have emerged as one of the state-of-the-art approaches for this task. However, existing CL-based methods face a critical challenge: anomalous nodes infiltrate the sampled local communities, leading to the distortion of community representation which fundamentally limits the discriminative ability. Our theoretical analysis reveals that this distortion is caused by two main mechanisms: the cross contamination and the aggregation bias. And the key oversight is to treat all community members equally and ignore the relative reliability of nodes. To address these issue, we propose a CL-based ANomaly detectIon Method on Attributed networks targeted at mitigating community distortions to enhance anomaly discrimination (ANIMA for short), which incorporates a Truncation-Restriction community encoder (TRC-Encoder) with an elaborate heuristic prior instruction to detect and suppress anomalous contributions during community representation learning. Comprehensive experiments on 7 datasets demonstrate that ANIMA outperforms 10 SOTA methods by 2.25-8.8% AUC, validating the effectiveness of our approach in mitigating community distortions and enhancing anomaly discrimination.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**.

KEYWORDS

Anomaly Detection, Attributed Networks, Graph Neural Network

ACM Reference Format:
Enbo He, Yitong Hao, Yue Zhang, Guisheng Yin, and Lina Yao. 2025. Addressing the Distortion of Community Representations in Anomaly Detection on Attributed Networks. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746252.3761418

1 INTRODUCTION

Tasks on networks have drawn much attention in recent years [20], particularly for anomaly detection on attributed network because there are many complex interaction patterns between entities hard to process [9]. The difficulty lies in the existence of multiple anomaly types, including attribute anomalies, characterized by nodes with atypical attribute conditions, and structural anomalies, referring to nodes with aberrant positions or connections within the network topology [3]. Given the scarcity of real labels in anomaly detection scenarios, many works focus on anomaly detection under the unsupervised condition.

There have been many studies on the unsupervised anomaly detection on attributed networks. In the early research, many non-deep methods were proposed. Metrics that measure abnormality are their major focus. For instance, [1] proposes a node isolation metric. [24] considers node clustering difficulties. Both [10] and [14] use residual analysis to calculate node anomaly scores. Due to their reliance on human anomaly assessment, these approaches have limited anomaly detection capacity. Thanks to the increasing development of GNN allows for simultaneous topological and attribute knowledge extraction [23]. And numerous researchers have been prompted to integrate GNNs into the anomaly detection on attributed networks [9]. For example, the Auto-Encoder based methods [3, 25], the One Class-SVM based methods [22, 28], and the Contrastive Learning based (CL-based) methods [4, 8, 11, 26, 27].

The Existing Challenge of the Current CL-based Methods: Although many studies have demonstrated that the CL-based methods have emerged as the most advanced unsupervised methodology [4, 8, 11, 26, 27], fundamental limitations in existing methodologies warrant urgent attention. These methods typically estimate the probability of a node being homologous to a community by

*Both authors contributed equally to this research.

CIKM '25, November 10–14, 2025, Seoul, Republic of Korea
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2040-6/2025/11.
https://doi.org/10.1145/3746252.3761418

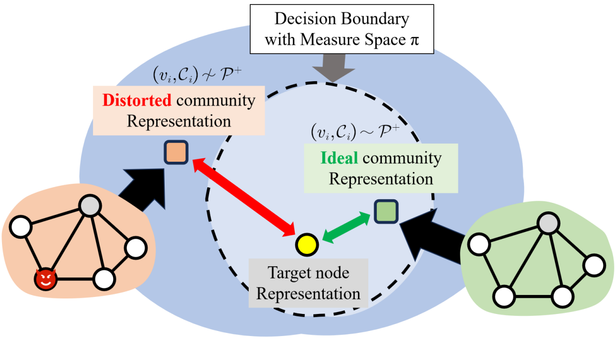

Figure 1: The negative effect of anomalous nodes infiltrating the community. Both the left (orange background) and right (green background) communities are believed to be natural communities sampled for a normal node (yellow circle). Their main difference is that the orange community has aberrant nodes, distorting the community representation. The anomalous node caused the community to be misestimated outside the decision border, making it indistinguishable from negative communities.

sampling local communities via RWR[18] and using GNNs for representation learning, followed by a bilinear discriminator. However, this pipeline suffers from a critical flaw: anomalous nodes inevitably infiltrate sampled communities, causing the **distortion of community representation**, which is the major focus of this work. As shown in Figure 1, when anomalous nodes are involved in the community, they propagate information that misrepresents the relationship between the positive community and the decision boundary, leading to misestimate the probability of homology between the target node and the community it is truly located in.

Our theoretical analysis (see §3) shows that the distortion arises through two mechanisms: (1) **Cross Contamination**: Anomalous nodes corrupt neighboring node representations via GNN propagation and conceal their own characteristics using normal node information, both due to the message passing mechanism. (2) **Aggregation Bias**: Simple averaging during as the readout method bias the community representation from the ideal representation.

Mathematically, in a community with P nodes and m anomalies, the normalized distortion degree $\tau(\mathbf{e}_i) = \frac{\|\mathbf{e}_i - \mathbf{e}_i^{ideal}\|_F}{\|\mathbf{e}_i^{worst} - \mathbf{e}_i^{ideal}\|_F}$ increases linearly with m under mean aggregation which limits the ability to distinguish the node-community pairs. The key issue is that all community members are treated equally, ignoring their relative reliability. This motivates designing a specialized community encoder to detect and suppress anomalous contributions during community representation learning.

In this work, we not only descriptively explain the sources of distortions, but also theoretically reveal the effectiveness of operation like truncation and restriction for mitigating distortions. Combining these motivations, we propose a CL-based **AN**omaly detect**I**on **M**ethod on **A**ttributed networks targeted at mitigating community distortions to enhance anomaly discrimination (**ANIMA** for short). Due to the intrinsic limitations of unsupervision, we design a heuristic affinity matrix as the prior to instruct the community encoder of ANIMA to mitigate the distortion. In addition, to prevent valid information from being overly culled, we adopt an auxiliary task imposed on the discriminator to enhance the expressiveness of the community representations.

The main contributions of this paper are:

- Theoretical Foundation: We theoretically analyze the community representation distortion under anomaly infiltration in CL-based methods, with quantitative bounds on expected distortion.
- Methodological Innovation: We proposed ANIMA, a novel framework that incorporates a community encoder with functionality like truncation and restriction to effectively detect and suppress anomalous contributions during community representation learning.
- Empirical Validation: Comprehensive experiments on 7 datasets show ANIMA outperforms 12 SOTA methods up to 11.26% AUC, with complementary experiments demonstrating the practical contribution of the various components of ANIMA and their consistency with theory.

2 PROBLEM FORMULATION

In this paper, the bold lowercase letter (e.g. $\mathbf{x}$) and uppercase letter(e.g. $\mathbf{X}$) are adopted to indicate vectors and matrices, respectively. The calligraphic fonts (e.g. $\mathcal{V}$) are used to denote sets and distributions, which is very easy to distinguish in the body of the text. The i−th row of a matrix $\mathbf{X}$ is denoted by $\mathbf{x}_i$ and the (i, j)−th element of $\mathbf{X}$ is denoted by $\mathbf{X}_{i,j}$.

Definition 1. Attributed Networks. *Given an attributed network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V} = \{v_i, \cdots, v_N\}$ is the set of nodes, the number of nodes $|\mathcal{V}|$ is N, where the set of normal nodes is $\mathcal{V}_+$, the set of abnormal nodes is $\mathcal{V}_-$, and $\mathcal{V} = \mathcal{V}_+ \cup \mathcal{V}_-$. $\mathcal{N}(v_i)$ is the node set of the neighbors of v_i. The $\mathcal{E} = \{e_1, \cdots, e_M\}$ is the set of edges and the number of edges $|\mathcal{E}|$ is M. The topology information is presented by the adjacent matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. If there is an edge connecting between the v_i and v_j, $\mathbf{A}_{i,j} = 1$. Otherwise, $\mathbf{A}_{i,j} = 0$. The attribute matrix is denoted by $\mathbf{X} \in \mathbb{R}^{N \times f}$, and the attributed of v_i is $\mathbf{x}_i \in \mathbb{R}^f$ where f is the dimension of the attribute vector.*

Problem 1. Anomaly Detection on Attributed Networks. *For an attributed network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, the aim is to learn a scoring function $score(\cdot)$ for qualifying the degree of each nodes' abnormality. To be specific, the larger the anomaly score $s_i = score(v_i)$ indicates the node v_i is more likely to be an anomaly.*

3 BACKGROUND AND MOTIVATION ANALYSIS

In this section, we start with delving into the workflow of traditional CL-based anomaly detection methods. Subsequently, by formalizing the concept of community distortion, we reveal the intrinsic limitations of existing methods and provide theoretical underpinnings for the design of ANIMA.

3.1 CL-Based Method for Anomaly Detection

In essence, CL-based methods aim to estimate the probability that a node-community pair $(v_i, \mathcal{C}_i)$ belongs to the positive distribution $\mathcal{P}^+$ [4, 8, 11, 26, 27]. Positive samples are drawn from $\mathcal{P}^+$ as

$(v_i, C_i^+) \sim \mathcal{P}^+$, implying homology between the node and community. To distinguish positive and negative pairs, negative samples are drawn from the negative distribution $\mathcal{P}^-$, formed by selecting another node as the target for positive sampling, resulting in $(v_i, C_i^-) \sim \mathcal{P}^-$. Here, C_i^+ and C_i^- denote positive and negative communities, respectively. Formally, the positive and negative distributions are defined as follows:

$$\mathcal{P}^+ = \mathbb{P}(C_i|v_i) \cdot \mathbb{P}(v_i) \tag{1}$$

$$\mathcal{P}^- = \mathbb{P}(C_j|v_j) \cdot \mathbb{P}(v_j|v_i)|_{v_j \in \mathcal{V} \setminus v_i} \cdot \mathbb{P}(v_i) \tag{2}$$

where $\mathbb{P}(v_i)$ denotes the probability density of randomly selecting node v_i from the node distribution, while C_i and C_j represent the communities obtained through positive sampling centered at nodes v_i and v_j, respectively. In practice, existing CL-based methods use Random Walk with Restart (RWR [18]) for P-step sampling to obtain fixed-size communities. To estimate the probability of a pair (v_i, C_i) belonging to the positive distribution $\mathcal{P}^+$, they use a node encoder and a community encoder to learn representations, and a discriminator for estimation. This estimation process can be expressed as:

$$\mathbb{P}((v_i, C_i) \sim \mathcal{P}^+) = \Pi_\pi(\Phi_\phi(v_i), \Psi_{[\phi,\psi]}(C_i)) \tag{3}$$

where, $\Phi(\cdot)$ denotes the node encoder, $\Psi(\cdot)$ represents the community encoder, and $\Pi(\cdot, \cdot)$ is the discriminator for estimation. The subscripts indicate the respective parameters.

It should be noted that the node encoder and community encoder share parameters ϕ to map nodes and communities into the same semantic space. Community encoder may also involve private parameters ψ due to the aggregation operation. The community representation is obtained by first encoding community nodes with a GNN, followed by a readout operation:

$$\Psi_{[\phi,\psi]}(C_i) = Readout_\psi(GNN_\phi(C_i)) \in \mathbb{R}^d \tag{4}$$

where d is the dimension of the representation. The most common readout method is average pooling, which implies $\psi = \emptyset$. Subsequently, the node encoding is performed via a direct mapping by a MLP:

$$\Phi_\phi(v_i) = MLP_\phi(v_i) \in \mathbb{R}^d \tag{5}$$

At last, it is common for the estimation to use a discriminator consisting of a bilinear transformation:

$$\Pi_\pi(\Phi_\phi(v_i), \Psi_{[\phi,\psi]}(C_i)) = Bilinear_\pi\left(\Phi_\phi(v_i), \Psi_{[\phi,\psi]}(C_i)\right) \in (0,1) \tag{6}$$

The model aims to separate the two distributions by minimizing the negative log-likelihood. The training objective can be equivalently formulated as:

$$\min_{\pi,\phi,\psi} -\{\mathbb{E}_{(v_i,C_i^+)\sim\mathcal{P}^+}[log(s_i^+)] + \mathbb{E}_{(v_i,C_i^-)\sim\mathcal{P}^-}[log(1-s_i^-)]\} \tag{7}$$

Here, $s_i^+ = \mathbb{P}((v_i, C_i^+) \sim \mathcal{P}^+)$ and $s_i^- = \mathbb{P}((v_i, C_i^-) \sim \mathcal{P}^+)$. This objective effectively trains the model to assign higher estimated probability (closer to 1) to positive community-node pairs and lower estimated probability (closer to 0) to negative pairs. Since anomalous nodes violate the inherent homology between nodes and positive communities, the model experiences increased ambiguity in distinguishing between positive and negative communities for these nodes. Therefore, in the anomaly inference procedure, the following method is adopted to compute the anomaly score:

$$score(v_i) = \mathbb{E}_{(v_i,C_i^+)\sim\mathcal{P}^+,(v_i,C_i^-)\sim\mathcal{P}^-}[s_i^- - s_i^+]|_{\mathbb{P}(v_i)=1} \tag{8}$$

The score is normalized by setting $\mathbb{P}(v_i) = 1$, indicating that the score is computed for a specific node v_i without considering the overall node distribution.

3.2 The Challenge of Community Representation Distortion

Despite the demonstrated effectiveness of CL-based methods, intrinsic limitations remain to be addressed. As discussed in the previous section, traditional CL-based methods employ RWR for sampling. However, due to the nature of unsupervised anomaly detection, it is impossible to ensure the consistency of labels within the sampled positive communities. This issue consequently leads to distortion of community representations from their ideal counterparts by the cross contamination and the aggregation bias.

To delve into the impact of anomalous node infiltration, we begin by assuming that the representations of normal and anomalous nodes are independently and identically sampled from two multivariate gaussian distributions, namely $\mathbf{h}_i^+ \sim \mathcal{N}(\mu^+, \Sigma)$ and $\mathbf{h}_j^- \sim \mathcal{N}(\mu^-, \Sigma)$. Here, $\mu^+, \mu^- \in \mathbb{R}^d$ are the mean vectors, Σ is the covariance matrix, and $\mu^+ \neq \mu^-$ implicitly signifies the distinct characteristics between the two distributions. For a given community C_i with $|\mathcal{V}_{C_i}| = P$ nodes, where m nodes are anomalous ($|\mathcal{V}_{C_i} \cap \mathcal{V}_-| = m$), the representation $\mathbf{e}_i$ of C_i under the mean aggregation of traditional CL-based methods is computed as:

$$\mathbf{e}_i = \frac{1}{P}\left(\sum_{v_i \in \mathcal{V}_{C_i} \cap \mathcal{V}_+}^{P-m} \mathbf{h}_i^+ + \sum_{v_j \in \mathcal{V}_{C_i} \cap \mathcal{V}_-}^{m} \mathbf{h}_j^-\right) \tag{9}$$

Under ideal conditions, where no anomalous nodes are involved in the aggregation, the community representation is:

$$\mathbf{e}_i^{ideal} = \frac{1}{P-m} \sum_{v_i \in \mathcal{V}_{C_i} \cap \mathcal{V}_+}^{P-m} \mathbf{h}_i^+ \tag{10}$$

To quantify the distortion, a natural approach is to calculate the euclidean distance between $\mathbf{e}_i$ and $\mathbf{e}_i^{\text{ideal}}$. We refer to this as the Absolute Distortion Degree.

Definition 2. Absolute Distortion Degree. Given a community C_i with representation $\mathbf{e}_i$, the absolute distortion degree is defined as the euclidean distance between $\mathbf{e}_i$ and $\mathbf{e}_i^{\text{ideal}}$, formalized as:

$$ADD(\mathbf{e}_i) = \left\|\mathbf{e}_i - \mathbf{e}_i^{ideal}\right\|_F \tag{11}$$

Correspondingly, to provides a standardized metric to quantify the impact of anomalous node infiltration, we introduce the concept of Normalized Distortion Degree as follows:

Definition 3. Normalized Distortion Degree. Given a community C_i with representation $\mathbf{e}_i$, the normalized distortion degree is defined as the ratio of the absolute distortion to the maximum possible distortion, formalized as:

$$\tau(\mathbf{e}_i) = \frac{ADD(\mathbf{e}_i)}{\left\|\mathbf{e}_i^{worst} - \mathbf{e}_i^{ideal}\right\|_F} = \frac{\left\|\mathbf{e}_i - \mathbf{e}_i^{ideal}\right\|_F}{\left\|\mathbf{e}_i^{worst} - \mathbf{e}_i^{ideal}\right\|_F} \tag{12}$$

Here, the $\mathbf{e}_i^{worst}$ is the representation when all nodes in the community are anomalous, defined as:$\mathbf{e}_i^{worst} = \frac{1}{m}\sum_{v_i \in \mathcal{V}_{C_i} \cap \mathcal{V}_-}^{m} \mathbf{h}_i^-$.

The normalized distortion degree measures the relative deviation normalized by the maximum possible distortion. This normalization ensures that $\tau(\mathbf{e}_i)$ is invariant to the scale of the underlying distributions and provides a standardized metric to quantify the impact of anomalous node infiltration on community representation.

Based on the definitions and assumptions presented above, we formally state the following Theorem 1.

Theorem 1: Upper Bound on Expected Normalized Distortion Degree. *For a given community C_i with P nodes, where m nodes are anomalous ($m < P$), the expected value of the normalized distortion degree $\mathbb{E}[\tau(\mathbf{e}_i)]$ is bounded above by:*

$$\mathbb{E}\left[\tau(\mathbf{e}_i)\right] \le \frac{1}{P}\sqrt{\left[\left(\frac{m}{P-m}\right)^2 + 1\right]\frac{Tr\left[\Sigma\right]}{\|\mu^- - \mu^+\|_F^2} + m^2} \tag{13}$$

The proof is displayed in §6.1. By examining the upper bound in Theorem 1, it is evident that the distortion of $\mathbf{e}_i$ is contingent upon the number of anomalous nodes m, which is referred as the "aggregation bias" in this work. This observation underscores that the infiltration of anomalous nodes into communities is the primary cause of the distortion.

To further illustrate the phenomenon, we designed an experiment by manually selecting the proportion of node neighbor labels to elucidate the origin and impact of community distortion. Specifically, we take the experimental setting of $P = 10$ on the dataset Amazon, and for pairs taken from the positive distribution, we manually control the number of anomalous nodes in the community by replacing either the true neighbor or the fake neighbor attributes, and compute the degree of distortion in the community representation. The results are shown in the Figure 2. It is evident that the proportion of anomalous nodes within a community is positively correlated with the degree of distortion, which in turn is negatively correlated with the model's training performance. This visualization not only validates the phenomenon described in Theorem 1 but also highlights the significant impact of community distortion on CL-based methods.

To prevent anomalous nodes from infiltrating the community representation under the bias aggregation, an intuitive solution is to suppress the proportion of anomalous nodes in the aggregation process by introducing a restriction coefficient $r \ll 1$. The community representation after restriction becomes:

$$\mathbf{e}_i^r = \frac{(1-r)}{P-m}\sum_{v_i \in \mathcal{V}_{C_i} \cap \mathcal{V}_+}^{P-m} \mathbf{h}_i^+ + \frac{r}{m}\sum_{v_j \in \mathcal{V}_{C_i} \cap \mathcal{V}_-}^{m} \mathbf{h}_j^- \tag{14}$$

To demonstrate the effectiveness of restriction in alleviating distortion, we present Theorem 2.

Theorem 2: The effectiveness of restriction. *For a given community C_i with P nodes, where m nodes are anomalous ($m < P$), the expected value of the normalized distortion degree $\mathbb{E}[\tau(\mathbf{e}_i^r)]$ when using the restriction coefficient r is bounded above by:*

$$\mathbb{E}\left[\tau\left(\mathbf{e}_i^r\right)\right] \le r\sqrt{\frac{m^2 + (P-m)^2}{(P-m)^2\, m^2}\frac{Tr\left(\Sigma\right)}{\|\mu^- - \mu^+\|_F^2} + 1} \tag{15}$$

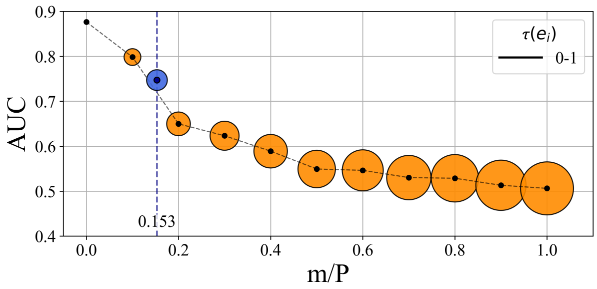

Figure 2: Empirical relationship between the number of anomalous nodes m in the community and distortion $\tau(e_i)$ and AUC. The size of the bubbles indicates the degree of distortion. The blue bubble represents the actual distortion on the Amazon dataset.

The proof is displayed in §6.2. Theorem 2 demonstrates that by introducing a restriction coefficient r, the expected normalized distortion degree can be significantly reduced with a speed of $O(r)$ and provides a theoretical foundation for the effectiveness of the restriction mechanism in mitigating community distortion caused by anomalous nodes.

Theorems 1 and 2 assume the term $\|\mu^- - \mu^+\|_F^2$ is constant, but in practice, CL-based methods use GNNs with message-passing, leading to cross-contamination where normal and anomalous node representations mix, which reduces distribution separability. To analyze this phenomenon, we assume the following mixed connectivity pattern within the community: on average, each normal node has n_- anomalous neighbors, and each anomalous node has n_+ normal neighbors. Considering only the message-passing mechanism of GNNs, the update rule for each node's representation is given by:$h_i' = \frac{1}{d_i+1}\left(h_i + \sum_{v_j \in \mathcal{N}(v_i)} h_j\right)$, we observe that: the expected distance between updated normal and anomalous representations contracts as:

$$\begin{aligned}\left\|\mu^{-\prime} - \mu^{+\prime}\right\|_F &= \left\|\mathbb{E}\left[\mathbf{h}^{+\prime}\right] - \mathbb{E}\left[\mathbf{h}^{-\prime}\right]\right\|_F \\ &= \left|1 - \frac{n_-}{d^+ + 1} - \frac{n_+}{d^- + 1}\right| \cdot \left\|\mu^- - \mu^+\right\|_F\end{aligned} \tag{16}$$

where d^+ and d^- denote the average degrees of normal and anomalous nodes respectively. And $\mu^{-\prime}$ and $\mu^{+\prime}$ represent the mean vector of the two distributions. updated by GNN. This contraction effect demonstrates how cross-contamination reduces the discriminative power of node representations, creating a compounding distortion effect in community aggregation. Combining the insights from Theorem 1, we further elaborate on the impact of Cross-Contamination on community representations: Cross-Contamination leads to an increase in the upper bound of the expected normalized distortion of community representations. Our analysis suggests that strategic truncation on edge (or edge pruning) can mitigate these effects. Consider truncating k most suspicious connections in each neighborhood. The updated separability becomes:

$$\left\|\mu_{tr}^{-\prime} - \mu_{tr}^{+\prime}\right\|_F = \left|1 - \frac{n_- - k}{d^+ - k + 1} - \frac{n_+ - k}{d^- - k + 1}\right| \cdot \left\|\mu^- - \mu^+\right\|_F \tag{17}$$

where the subscript tr denotes the truncation. We can see that the term $\frac{n_- - k}{d^+ - k + 1}$(normal nodes receiving anomalous messages) and

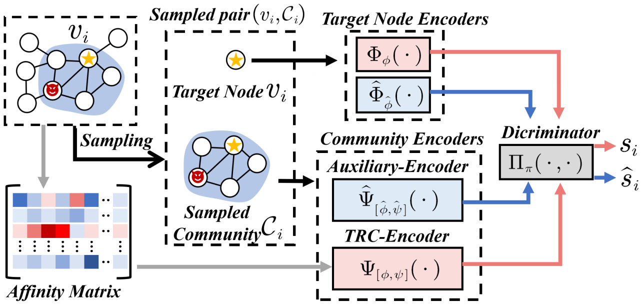

Figure 3: The Proposed ANIMA.

$\frac{n_+ - k}{d^- - k + 1}$ (anomalous nodes receiving normal messages) decay with k, effectively suppressing cross-contamination. And the coefficient $\left|1 - \frac{n_- - k}{d^+ - k + 1} - \frac{n_+ - k}{d^- - k + 1}\right|$ increases toward 1 as k grows, restoring the original separability $\left\|\mu^- - \mu^+\right\|_F^2$ compromised by mixed connectivity. This result demonstrates that the truncation is able to alleviate the distortion of the community distortion.

These theoretical insights form the foundation of ANIMA's architecture. The proportional relationship between anomalous infiltration and representation distortion explains why existing CL methods struggle with community quality.

4 THE PROPOSED ANIMA

In this section, we will follow the workflow of CL-based anomaly detection methods to introduce the individual components of ANIMA with training and score inference. The main workflow diagram of ANIMA is displayed in Figure 3.

4.1 Community Representation Learning

Based on the preceding discussion, it has been established that reducing the involvement of anomalous information in community representations can effectively enhance the performance of CL-based method. However, in the context of unsupervised learning, label information is unavailable, therefore, there is no prior information for the model to determine how to prevent distortion. To address this challenge, we propose an affinity matrix to serve as a prior reference for the model to perform the truncation and the restriction.

4.1.1 Node-Centric Instructions as the Prior Information. Graph-structured data relies on the homophily assumption, where connected nodes are typically similar. However, in anomaly detection, anomalous nodes disrupt this assumption. Thus, measuring affinity to estimate the reliability is a plausible design approach. To the point of this, ANIMA constructs an affinity matrix $\mathbf{M}$ that captures both direct and hypothetical relationships between nodes within communities for the fact that node v_i may not have direct edges with all nodes in community C_i. This matrix helps to correct distortion in community representations by considering node-node relationships for truncation and node-community relationships for restriction. Specifically, $\mathbf{M}_{i,j}$ indicates the reliability of message passing from v_j to v_i and the likelihood of v_j belonging to the community centered at v_i. Therefore, the design of $\mathbf{M}$ is node-centric, focusing on the perspective of each node.

Previous methods generally adopt cosine similarity for normalization [7, 17]. Such approach first maps all node attributes onto a hypersphere and then computes the normalized dot product similarity between node attributes, i.e., $\mathbf{M}^{cos}_{i,j} = \mathbf{x}_i / |\mathbf{x}_i| \cdot \mathbf{x}_j^T / |\mathbf{x}_j| \in (-1, 1)$. This design measures reliability from a global perspective and fails to provide effective information on the individual relationships between nodes and communities. To overcome this, ANIMA uses dot-product similarity to compute the similarity matrix $\mathbf{S} = \mathbf{X} \cdot \mathbf{X}^T$, followed by a row-wise min-max normalization to obtain $\mathbf{M}$:

$$\mathbf{M}_{i,j} = \frac{\mathbf{S}_{i,j} - \mathbf{S}_{i,\underset{j}{argmin}\{\mathbf{S}_{i,j}\}}}{\mathbf{S}_{i,\underset{j}{argmax}\{\mathbf{S}_{i,j}\}} - \mathbf{S}_{i,\underset{j}{argmin}\{\mathbf{S}_{i,j}\}}} \tag{18}$$

Here, $\mathbf{S}_{i,\underset{j}{argmin}\{\mathbf{S}_{i,j}\}}$ and $\mathbf{S}_{i,\underset{j}{argmax}\{\mathbf{S}_{i,j}\}}$ represent the minimum and maximum affinities of node v_i with all other nodes in the network, respectively. This approach ensures that the affinities are specific to each node and maintains the relative importance of nodes within communities.

4.1.2 Truncation - Restriction Community Encoder (TRC-Encoder). As previously analyzed, correcting community representations involves truncation and restriction. We design a truncation-restriction community encoder $\Psi(\cdot)$ to exclude anomalies from C_i. Both operations affect C_i^+ and C_i^-, but the impact on negative communities is minimal due to the low affinity between abnormal nodes and their communities, such a perspective will be confirmed in the experiment §5.3.

The truncation in TRC-Encoder is implemented via a gating mechanism attached to the community node encoding process. In principle, any type of GNN can be chosen; to maintain consistency with prior work, we continue to select GCN as the representation learner of the in-community node:

$$\mathbf{H}_{C_i} = \sigma\left(\tilde{\mathbf{D}}_{C_i}^{-\frac{1}{2}} \tilde{\mathbf{A}}_{C_i} \odot GATE\left(\mathbf{M}_{C_i}\right) \tilde{\mathbf{D}}_{C_i}^{-\frac{1}{2}} \mathbf{X}_{C_i} \mathbf{W}_{\phi}\right) \in \mathbb{R}^{P \times d} \tag{19}$$

where $\tilde{\mathbf{A}}_{C_i} = \mathbf{A}_{C_i} + \mathbf{I}$ represents the topology of community C_i with self-loops and $\mathbf{M}_{C_i}$ is a submatrix extracted from $\mathbf{M}$ based on the nodes within the community. The matrix $\tilde{\mathbf{D}}_{C_i}$ denotes the adjusted degree matrix after truncation, while $\mathbf{X}_{C_i}$ is the attribute matrix of nodes within the community. For multi-layer designs, $\mathbf{X}_{C_i}$ can be replaced by the embeddings from the previous encoder layer. The activation function $\sigma(\cdot)$ is PReLU, and $\odot$ denotes the Hadamard product, used to apply the gating mechanism to the adjacency matrix. The gating function $GATE(\cdot) \in \mathbb{R}^{P \times P}$ transforms the affinity matrix $\mathbf{M}_{C_i}$ into a gating matrix that selectively blocks or reduces the influence of certain edges.

The gating function can adopt either soft or hard truncation. Soft truncation learns edge truncation probabilities for probabilistic filtering, but ANIMA uses hard truncation since the affinity matrix already has enough discriminative power.Specifically, the gating function is defined as:

$$GATE\left(\mathbf{M}_{C_i}\right) = \left[1 - \mathbb{1}\left[\mathbf{M}_{C_i} < \varepsilon\right]\right] \wedge \left[1 - \mathbb{1}\left[\mathbf{M}_{C_i} < \varepsilon\right]\right]^T \tag{20}$$

In this context, $\wedge$ denotes the element-wise logical OR operation between matrices and such operation maintain symmetry of the gating matrix. ε represents the truncation threshold as a hyper-parameter, and $\mathbb{1}[\cdot]$ here is the indicating function. These operations aim to maintain the symmetry of the $GATE(\cdot)$ function while achieving "Unilateral Trust Truncation." This means that if one node at both ends of an edge is suspicious, bidirectional message passing is stopped, effectively reducing cross-contamination.

After obtaining the node representations within a community, the next objective is to derive the community representation. To achieve this, TRC-Encoder introduces a restriction parameter vector $\mathbf{r}_{C_i} = [r_v | v \in \mathcal{V}_{C_i}] \in \mathbb{R}^P$ that satisfies $\mathbf{r}_{C_i} \cdot \mathbb{1}^T = 1$, where each element represents the restriction degree of a node within community C_i. The community representation is computed as:

$$\Psi_{[\phi,\psi]}(C_i) = \sum_{v \in \mathcal{V}_{C_i}} r_v \cdot \mathbf{h}_v \tag{21}$$

Under the unsupervised learning paradigm, it is infeasible to determine whether each node should be excluded from the aggregation process. Therefore, leveraging the affinity information introduced earlier as a prior, we adaptively compute $\mathbf{r}_{C_i}$ via a linear transformation:

$$\mathbf{r}_{C_i} = Softmax\left(\mathbf{m}_{C_i}\mathbf{W}_{\psi} + \mathbf{b}\right) \tag{22}$$

where the $\mathbf{m}_{C_i} \in \mathbb{R}^P$ is the affinity vector for community C_i, derived from the node-centric affinity matrix $\mathbf{M}$. Each element m_v of $\mathbf{m}_{C_i}$ represents the reliability score of node v within the community. $\mathbf{W}_{\psi} \in \mathbb{R}^{P \times P}$ is a learnable weight matrix that transforms the affinity vector into the restriction parameter vector. $\mathbf{b} \in \mathbb{R}^P$ is a learnable bias vector. The Softmax function ensures that the elements of $\mathbf{r}_{C_i}$ are normalized to sum to 1.

4.2 Node Representation Learning and Discriminator Training

After obtaining the community representations, it is conventional to train a discriminator to estimate $\mathbb{P}((v_i, C_i) \sim \mathcal{P}^+)$. ANIMA employs a discriminator training scheme with an auxiliary task. Since the previous distortion mitigation methods were based on heuristic prior rather than labels, the design of $\Psi_{[\phi,\psi]}(\cdot)$ might inadvertently discard useful information. To address this issue, an additional community encoder $\hat{\Psi}_{[\hat{\phi},\hat{\psi}]}(\cdot)$ without truncation and restriction is used. And both community encoders are correspondingly equipped with node encoders, $\Phi_{\phi}(\cdot)$ and $\hat{\Phi}_{\hat{\phi}}(\cdot)$, as in Eq(5). The estimation based on $\hat{\Psi}_{[\hat{\phi},\hat{\psi}]}(\cdot)$ and $\hat{\Phi}_{\hat{\phi}}(\cdot)$ serves as an auxiliary task to impose implicit regularization and enhance feature representation.

For the discriminators corresponding to $\Psi_{[\phi,\psi]}(\cdot)$ and $\hat{\Psi}_{[\hat{\phi},\hat{\psi}]}(\cdot)$, a hard-sharing design mechanism is adopted, utilizing the same discriminator parameters π,as in Eq(6). This constrains the two sets of node-community representations to be discriminable within the same measure space, thereby facilitating knowledge transfer. The training objective can be expressed as:

$$\begin{aligned}\min_{\pi,\phi,\hat{\phi},\psi,\hat{\psi}} -\{&\mathbb{E}_{(v_i,C_i^+)\sim\mathcal{P}^+}[\gamma log(s_i^+) + (1-\gamma)\, log(\hat{s}_i^+)] + \\ &\mathbb{E}_{(v_i,C_i^-)\sim\mathcal{P}^-}[\gamma log(1-s_i^-) + (1-\gamma)\, log(1-\hat{s}_i^-)]\}\end{aligned} \tag{23}$$

Here, s_i^+ and $\hat{s}_i^+$ are the estimation scores computed by Eq(7) based on the community representations and target node representations obtained from $\Psi(\cdot)$ and $\hat{\Psi}(\cdot)$, respectively. Similarly, s_i^- and $\hat{s}_i^-$ are the scores for negative pairs. The parameter γ controls the involvement of the auxiliary task.

4.3 Anomaly Score Inference

The CL-based method primarily relies on the assumption that anomaly nodes exhibit higher discrimination ambiguity. Given that the discriminator is trained on two estimation tasks, both inherently possess certain anomaly inference capabilities. Therefore, ANIMA adopts a composite form of score inference:

$$score_{CL}(v_i) = \mathbb{E}_{\substack{(v_i,C_i^+)\sim\mathcal{P}^+ \\ (v_i,C_i^-)\sim\mathcal{P}^-}}[\gamma\left(s_i^- - s_i^+\right) + (1-\gamma)\left(\hat{s}_i^- - \hat{s}_i^+\right)]|_{\mathbb{P}(v_i)=1} \tag{24}$$

where γ is the hyper-parameter, like Eq(23), balances the contributions of the primary and auxiliary discrimination tasks to the composite score.

In addition to the contrastive learning-based score, ANIMA leverages the affinity matrix $\mathbf{M}$ to provide an auxiliary anomaly inference criterion. Since $\mathbf{M}$ captures the reliability of edges within the community, it can serve as an indicator of potential anomalies. Therefore, the final anomaly score in ANIMA is computed as:

$$\begin{aligned}score(v_i) = &(1-\lambda)\, score_{CL}(v_i) \\ &+ \lambda\left(\left[\mathbb{1}[\mathbf{M} < \varepsilon] \vee \mathbb{1}[\mathbf{M} < \varepsilon]^T\right]_i \odot \mathbf{a}_i\right) \cdot \mathbb{1}^T\end{aligned} \tag{25}$$

Here, λ is the hyper-parameter that balances the contributions of the contrastive learning score and the affinity matrix based anomaly indicator. And $\mathbb{1}^T$ is for conducting summation over the elements of the resulting vector to obtain a scalar anomaly score. While the contrastive learning score captures the discrimination ambiguity of nodes, the affinity matrix provides an additional layer of anomaly detection by identifying suspicious edges. This dual approach ensures that both node-level and edge-level information are leveraged effectively, improving the overall accuracy and reliability of anomaly inference.

5 EXPERIMENT

5.1 Experiment Setting

5.1.1 Datasets and Evaluation Metrics. In previous unsupervised learning studies, only artificially synthesized abnormal datasets are adopted for validation [3–6, 8, 11, 26, 27]. To ensure consistency and fairness with existing research, this study also adopt five widely recognized synthetic datasets, namely Cora, Citeseer, PubMed, ACM, and DBLP [19, 21], to validate the effectiveness of the proposed model. However, the synthetic datasets is not sufficient to accurately simulate anomaly patterns in the real world.

Table 1: Datasets Statistics and Implementation Details

Datasets	Cora	Citeseer	PubMed	ACM	DBLP	Amazon	Questions
Nodes	2708	3327	19717	16484	5484	11944	48921
Edges	5429	4732	44338	71980	8117	4398392	153540
Attributes	1433	3703	500	8337	6775	25	301
Anomalies	150	150	600	600	300	1135	1468
LR	1e-3	1e-3	1e-3	5e-4	3e-4	1e-3	1e-3
Epoch	100	100	100	400	400	100	100
γ	0.9	0.8	0.1	0.1	0.6	0.9	0.9
ε	0.1	0.1	0.05	0.01	0.05	0.25	0.3
λ	0.2	0.2	0.2	0.2	0.1	0.1	0.2

Table 2: Performance comparison for AUC. The bold and underlined values indicates the best and the under-best results, respectively.

Method	Cora	CiteSeer	PubMed	ACM	DBLP	Amazon	Questions
Radar[2017]	$0.5906_{\pm0.023}$	$0.5580_{\pm0.021}$	$0.5813_{\pm0.021}$	$0.4848_{\pm0.034}$	$0.5411_{\pm0.011}$	$0.4738_{\pm0.025}$	$0.4444_{\pm0.009}$
ANOMALOUS[2018]	$0.6279_{\pm0.019}$	$0.4336_{\pm0.017}$	$0.4624_{\pm0.011}$	$0.4967_{\pm0.026}$	$0.4508_{\pm0.018}$	$0.4461_{\pm0.022}$	$0.4749_{\pm0.004}$
DOMINANT[2019]	$0.8639_{\pm0.017}$	$0.9112_{\pm0.014}$	$0.7709_{\pm0.012}$	$0.8009_{\pm0.031}$	$0.6780_{\pm0.004}$	$0.6001_{\pm0.015}$	$0.5864_{\pm0.005}$
CoLA[2021]	$0.8910_{\pm0.009}$	$0.8982_{\pm0.018}$	$0.9532_{\pm0.009}$	$0.7957_{\pm0.022}$	$0.7291_{\pm0.024}$	$0.5899_{\pm0.029}$	$0.5542_{\pm0.003}$
ANEMONE[2021]	$0.9096_{\pm0.012}$	$0.8356_{\pm0.019}$	$0.9527_{\pm0.009}$	$0.8226_{\pm0.012}$	$0.7474_{\pm0.025}$	$0.5904_{\pm0.028}$	$0.5598_{\pm0.006}$
SL-GAD[2021]	$0.9080_{\pm0.008}$	$0.9243_{\pm0.015}$	$0.9662_{\pm0.004}$	$0.8156_{\pm0.016}$	$0.8170_{\pm0.013}$	$0.5989_{\pm0.009}$	$0.5621_{\pm0.002}$
Sub-CR[2022]	$0.9018_{\pm0.007}$	$0.9385_{\pm0.022}$	$0.9687_{\pm0.003}$	$0.8051_{\pm0.008}$	$0.8224_{\pm0.010}$	$0.6731_{\pm0.005}$	$0.5674_{\pm0.004}$
GRADATE[2023]	$0.9053_{\pm0.020}$	$0.8978_{\pm0.024}$	$0.9547_{\pm0.005}$	$0.8881_{\pm0.024}$	$0.7482_{\pm0.018}$	$0.6957_{\pm0.018}$	$0.5601_{\pm0.010}$
NLGAD[2023]	$0.9173_{\pm0.013}$	$0.9446_{\pm0.020}$	$0.9538_{\pm0.004}$	$0.8977_{\pm0.016}$	$0.7762_{\pm0.014}$	$0.7555_{\pm0.017}$	$0.5729_{\pm0.005}$
ARISE[2023]	$0.9226_{\pm0.009}$	$0.8966_{\pm0.012}$	$0.9664_{\pm0.002}$	$0.9217_{\pm0.019}$	$0.8278_{\pm0.013}$	$0.7813_{\pm0.019}$	$0.5871_{\pm0.004}$
TAM[2024]	$0.7843_{\pm0.010}$	$0.7591_{\pm0.007}$	$0.8630_{\pm0.015}$	$0.8826_{\pm0.017}$	$0.7052_{\pm0.006}$	$0.8286_{\pm0.007}$	$0.5354_{\pm0.006}$
GADAM[2024]	$0.9441_{\pm0.011}$	$0.9377_{\pm0.020}$	$0.9268_{\pm0.014}$	$0.9340_{\pm0.011}$	$0.7802_{\pm0.013}$	$0.7046_{\pm0.022}$	$0.5354_{\pm0.006}$
ANIMA	$\mathbf{0.9492}_{\pm0.012}$	$\mathbf{0.9710}_{\pm0.017}$	$\mathbf{0.9847}_{\pm0.005}$	$\mathbf{0.9442}_{\pm0.012}$	$\mathbf{0.8849}_{\pm0.007}$	$\mathbf{0.8693}_{\pm0.014}$	$\mathbf{0.6009}_{\pm0.005}$

To overcome this limitation, two datasets containing real labels, Amazon[13] and Questions[16] are introduced in this study. The detailed statistical information of the datasets is listed in Table 1. Besides, the ROC-AUC is utilized to measure the performance which is a widely-used anomaly detection metric.

5.1.2 Baselines. In this subsection, we present the datasets that were used for comparison with ANIMA. The first two models (Radar and ANOMALOUS [10, 14]) are non-deep methods, and the rest are based on graph neural networks, including an Auto-Encoder based method (DOMINANT [3]), six CL-based methods (CoLA, ANEMONE, GRADATE, NLGAD, ARISE and GADAM[2] [4–6, 8, 11]), two hybrid methods that combine Auto-Encoder and CL-based learning (SL-GAD and Sub-CR [26, 27]), and one method that proposed a novel anomaly scoring approach (TAM [17]).

5.1.3 Implementation Details. In our experiments, the one-layer TRC-Encoder is employed and in the auxiliary task, one-layer GCN is adopted as the encoder. The embedding dimension d is both set to 64. The batch size is set to 300 for all datasets. In the inference phase, To compute the expectation in Eq(8), we follow the practical approach in CoLA and use multiple rounds of sampling for the mean, specifically the number of rounds R is set to 256. Other parameters settings are displayed in Table 1.

5.2 Performance Comparison

Firstly, to comprehensively evaluate the performance of the ANIMA in attributed network anomaly detection, an exhaustive performance comparison with twelve state-of-the-art baseline methods is conducted. The comparison results are summarized in Table 2. Based on these results, we draw the following conclusions: (1) ANIMA demonstrates remarkable effectiveness, showcasing substantial improvements ranging from 1.09% to 11.26% in ROC-AUC across all datasets. Notably, it achieves state-of-the-art performance on eight datasets and remains highly competitive on 2 real-world datasets. (2) Non-deep methods perform poorly due to limited network structure modeling. Auto-Encoder methods underperform as they focus on node recovery rather than anomaly detection. (3) ANIMA excels over traditional CL-based methods by deeply understanding intrinsic community distortion flaws and devising efficient encoding methods. (4) Despite the fact that hybrid methods combine modules of Auto-Encoders and contrastive learning to form a technological complement, they still fail to adequately take into account the key issues in anomaly detection, and thus fail to outperform ANIMA in terms of performance.

5.3 Motivation Verification

In order to confirm the validity of the methodological motivation, empirical experiments are implemented in this section. Based on the previous hypothesis, ANIMA benefits from the corrective properties

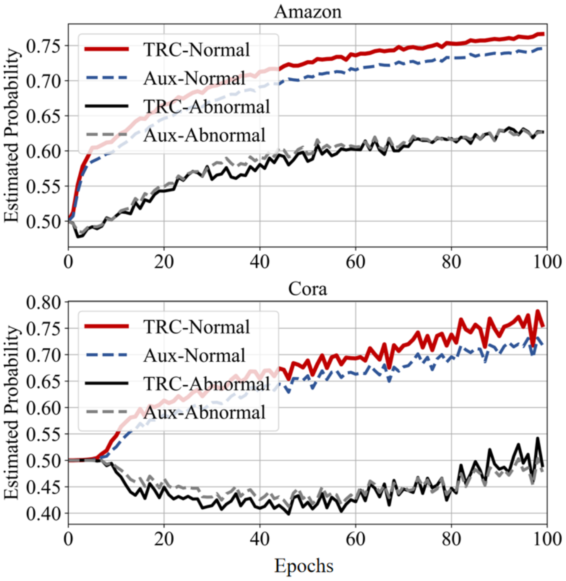

Figure 4: The trend of the estimation probability changing with the training process.

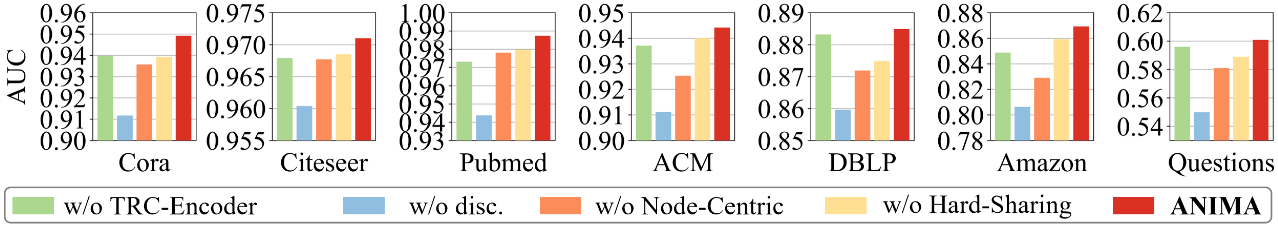

Figure 5: The AUC values of ablation study.

of TRC-Encoder for community representation. Therefore, we will explore in depth whether TRC-Encoder can really contribute to the performance improvement of ANIMA compared to other CL-based methods. To ensure representativeness, we selected one dataset from each category—synthetic and real-world for validation, namely Cora and Amazon.

5.3.1 Visualization of Estimated Probability. TRC-Encoder enhances the estimation of the probability that the pair (v_i, C_i) of a normal node belongs to the positive distribution $\mathcal{P}^+$ by curtailing the possible anomalous information in the positive community where the node is located. Thus, we demonstrate the estimation results $\mathbb{P}((v_i, C_i) \sim \mathcal{P}^+)$ of different labeled nodes under different community encoders during the training process.

As depicted in Figure 4, it can be observed that the estimation of normal nodes are generally higher than those of abnormal nodes. This phenomenon remains consistent in both encoders, thus demonstrating the effectiveness of CL-based method in distinguishing normal nodes from abnormal nodes. Further analysis indicates that for normal nodes, the estimated probability in the TRC-Encoder are markedly enhanced relative to the auxiliary encoder. In contrast, the estimated probability for abnormal nodes do not exhibit a significant disparity between the two encoders.

5.3.2 TRC-Encoder as Plug-and-Play Module. TRC-Encoder, a key component of ANIMA, is designed based on a profound understanding of CL-based methods. Therefore, it is reasonable to posit that the TRC-Encoder can be integrated into other CL-based methods as a plug-and-play module by replacing their existing community encoders. This not only enables direct performance comparison but also highlights TRC-Encoder's adaptability across different frameworks.

Table 3: The performance increase on several CL-based Methods by TRC-Encoder

	CoLA_TRC	**SL-GAD_TRC**	**GRADATE_TRC**
Cora	91.20(**2.10↑**)	92.31(**1.51↑**)	92.96(**2.43↑**)
Amazon	70.13(**11.14↑**)	73.56(**13.67↑**)	81.59(**12.02↑**)

To be specific, we selected three methods, namely CoLA, SL-GAD, and GRADATE, for our experiments. The results are presented in Table 3, where the figures in parentheses denote the improvement in scores compared with the original results. The results indicate performance improvements across all methods, demonstrating the TRC-Encoder's effectiveness. However, the gains are lower than those achieved by ANIMA, suggesting that while the TRC-Encoder is beneficial, it lacks the tailored design of ANIMA.

5.4 Ablation Study

To assess the contribution of the individual components in ANIMA, a series of ablation experiments are designed and executed in this subsection. The specific setup of the experiments is described below: (1) w/o TRC-Encoder, where the TRC-Encoder is removed to assess the impact of the corrective method on the community representation distortion. (2) w/o disc., where the role of affinity information in aided scoring in Eq(25) is assessed by removing the scores obtained by the discriminator. (3) w/o Node-Centric, the instruction matrix **M** is replaced by the cosine similarity. (4) w/o Hard-Sharing, where independent discriminators on each of the two views are adopted. The results of the ablation experiments are summarized in Figure 5. The results in Figure 5 show that removing the TRC-Encoder, discarding discriminator scores, replacing the node-centric prior, or dropping hard-sharing all sharply cut performance, confirming each component is indispensable to ANIMA's anomaly detection.

5.5 Sensitivity Analysis

Two sensitivity analysis experiments were conducted to understand the ANIMA's dependence on key parameters. The first analyzed parameters γ and λ, assessing the model's sensitivity to the balance between two encoders and the tuning of discriminative scoring and auxiliary criterion. The second focused on the truncation threshold ε. Results in Figure 6 showed that ANIMA's performance improves with increased TRC-Encoder importance on some datasets, while others require a balance between encoders. Performance was optimized when the auxiliary scoring criteria had a smaller share, highlighting the estimation's decisive role. Variations emphasized

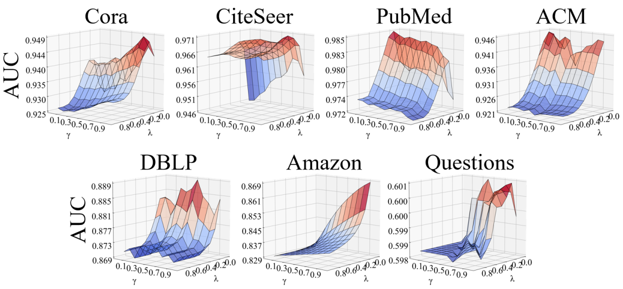

Figure 6: Performance with different γ and λ.

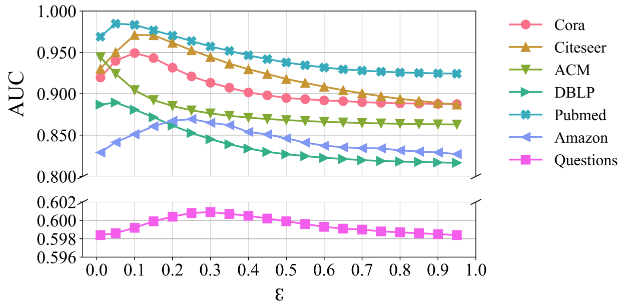

Figure 7: Performance with different ε.

the need for dataset-specific parameter selection. For ε, as showed in Figure 7, performance initially improved with a low threshold, indicating effective noise removal, but declined with higher thresholds, suggesting excessive truncation leads to loss of critical information.

6 PROOFS

6.1 Proof of Theorem 1

Proof. From the definitions in §2, we have:

$$
\begin{aligned}
&\mathbb{E}\left[\|e_i - e_i^{\text{ideal}}\|_F^2\right]\\
&= \mathbb{E}\left[\left\|\frac{1}{P}\left(\sum_{v_i\in\mathcal{V}_+}^{P-m} h_i^+ + \sum_{v_j\in\mathcal{V}_-}^{m} h_j^-\right) - \frac{1}{P-m}\sum_{v_i\in\mathcal{V}_+}^{P-m} h_i^+\right\|_F^2\right]\\
&= \frac{1}{P^2}\mathbb{E}\left[\left\|\left(-\frac{m}{P-m}\right)\sum_{v_i\in\mathcal{V}_+}^{P-m}(h_i^+ - \mu^+)\right.\right.\\
&\qquad \left.\left. + \sum_{v_j\in\mathcal{V}_-}^{m}(h_j^- - \mu^-) + m(\mu^- - \mu^+)\right\|_F^2\right]\\
&\leqslant \frac{1}{P^2}\left(\mathbb{E}\left[\left\|\left(-\frac{m}{P-m}\right)\sum_{v_i\in\mathcal{V}_+}^{P-m}(h_i^+ - \mu^+)\right\|_F^2\right]\right.\\
&\qquad \left. + \mathbb{E}\left[\left\|\sum_{v_j\in\mathcal{V}_-}^{m}(h_j^- - \mu^-)\right\|_F^2\right] + \mathbb{E}\left[\|m(\mu^- - \mu^+)\|_F^2\right]\right)\\
&= \frac{1}{P^2}\left(\left(\left(\frac{m}{P-m}\right)^2 + 1\right)\mathrm{Tr}[\Sigma] + m^2\|\mu^- - \mu^+\|_F^2\right)
\end{aligned}
$$

and then:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\|e_i - e_i^{\text{ideal}}\|_F^2}{\|e_i^{\text{worst}} - e_i^{\text{ideal}}\|_F^2}\right]\\
&\leq \frac{\left(\left(\left(\frac{m}{P-m}\right)^2 + 1\right)\mathrm{Tr}[\Sigma] + m^2\|(\mu^- - \mu^+)\|_F^2\right)}{P^2\|(\mu^- - \mu^+)\|_F^2}\\
&= \frac{1}{P^2}\left(\left(\left(\frac{m}{P-m}\right)^2 + 1\right)\frac{\mathrm{Tr}[\Sigma]}{\|(\mu^- - \mu^+)\|_F^2} + m^2\right)
\end{aligned}
$$

Applying Cauchy-Schwarz inequality for the normalized distortion, i.e., $\mathbb{E}^2\left[\tau(e_i)\right] \leq \mathbb{E}\left[\tau^2(e_i)\right] = \mathbb{E}\left[\frac{\|e_i - e_i^{\text{ideal}}\|_F^2}{\|e_i^{\text{worst}} - e_i^{\text{ideal}}\|_F^2}\right]$, then the theorem 1 is proved. □

6.2 Proof of Theorem 2

Proof. From the above setting in §2, we have:

$$
\begin{aligned}
&\mathbb{E}\left[\|e_i^r - e_i^{\text{ideal}}\|_F^2\right]\\
&= \mathbb{E}\left[\left\|\frac{(1-r)}{P-m}\sum_{v_i\in\mathcal{V}_+}^{P-m} h_i^+ + \frac{r}{m}\sum_{v_j\in\mathcal{V}_-}^{m} h_j^- - \frac{1}{P-m}\sum_{v_i\in\mathcal{V}_+}^{P-m} h_i^+\right\|_F^2\right]\\
&= r^2\mathbb{E}\left[\left\|\left(-\frac{1}{P-m}\right)\sum_{v_i\in\mathcal{V}_+}^{P-m}(h_i^+ - \mu^+)\right.\right.\\
&\qquad \left.\left. + \frac{1}{m}\sum_{v_j\in\mathcal{V}_-}^{m}(h_j^- - \mu^-) + (\mu^- - \mu^+)\right\|_F^2\right]\\
&\leq r^2\left(\mathbb{E}\left[\left\|\left(-\frac{1}{P-m}\right)\sum_{v_i\in\mathcal{V}_+}^{P-m}(h_i^+ - \mu^+)\right\|_F^2\right]\right.\\
&\qquad \left. + \mathbb{E}\left[\left\|\frac{1}{m}\sum_{v_j\in\mathcal{V}_-}^{m}(h_j^- - \mu^-)\right\|_F^2\right] + \mathbb{E}\left[\|\mu^- - \mu^+\|_F^2\right]\right)\\
&= r^2\left[\frac{m^2 + (P-m)^2}{(P-m)^2m^2}\mathrm{Tr}(\Sigma) + \|\mu^- - \mu^+\|_F^2\right]
\end{aligned}
$$

Similarly, we bound the $\mathbb{E}[\tau(e_i^r)]$ by:

$$
\mathbb{E}^2\left[\tau(e_i^r)\right] \leq r^2\left[\frac{m^2 + (P-m)^2}{(P-m)^2m^2}\frac{\mathrm{Tr}(\Sigma)}{\|\mu^- - \mu^+\|_F^2} + 1\right]
$$

then the Theorem 2 is proved. □

7 RELATED WORK

Unsupervised Anomaly Detection on Attributed Network. Initial non-deep methods for anomaly detection, such as those in [10] and [14], used matrix decomposition and residual analysis, while [1] and [15] designed metrics to measure abnormality, and [24] used clustering difficulty. However, these shallow methods cannot model complex network interactions. With deep learning, approaches like autoencoders in [3], high-order motif structures in [25], community-conditioned GCNs in [12], one-class SVM combined with GCN in [22], and deviation-based embeddings in [28] have been developed. Contrastive learning was introduced in [11], followed by extensions like [8], [4], [27], [26], [5], [6], [17], and [2]. Yet, these methods often fail to address distortions in community representations, which is the focus of our study.

8 CONCLUSION

In this paper, we proposed ANIMA, which addresses the critical challenge of community representation distortion in CL-based methods by introducing the TRC-Encoder. This design effectively suppresses the contributions of anomalous nodes, reducing distortion and enhancing anomaly detection performance. Comprehensive experiments validates the necessity and effectiveness of ANIMA. In conclusion, ANIMA provides a robust solution for anomaly detection by mitigating community representation distortion, and its innovative design and strong performance position it as a leading approach in the field.

9 GENAI USAGE DISCLOSURE

In this work, we have utilized generative AI software tools to improve the quality of our existing text. The tools were used to enhance the clarity, engagement, and correctness of the text, similar to the way a typing assistant like Grammarly is used to improve spelling, grammar, and punctuation. Specifically, we employed KIMI (k1.5 version) to refine the wording and structure of certain sections of the paper. The prompts provided to the tool were designed to generate suggestions for improving the text, and the generated suggestions were carefully reviewed and edited by the authors to ensure they aligned with our intended meaning and contribution. The amount of text generated by the tool was limited to phrases and sentences, and the final content reflects the authors' original ideas and intellectual contributions. All other components of the work, including the conception, design, analysis, and original drafting of the manuscript, were completed by the human authors without the use of generative AI tools for content creation. The authors take full responsibility for the accuracy and correctness of all material in this work, including any content that was improved with the assistance of generative AI tools.

REFERENCES

[1] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *SIGMOD*.

[2] Jingyan Chen, Guanghui Zhu, Chunfeng Yuan, and Yihua Huang. 2024. Boosting Graph Anomaly Detection with Adaptive Message Passing. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=CanomFZssu

[3] Kaize Ding, Jundong Li, Rohit Bhanushali, and Huan Liu. 2019. Deep anomaly detection on attributed networks. In *SDM*.

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

[5] Jingcan Duan, Bin Xiao, Siwei Wang, Haifang Zhou, and Xinwang Liu. 2023. ARISE: Graph Anomaly Detection on Attributed Networks via Substructure Awareness. *IEEE transactions on neural networks and learning systems* (2023).

[6] Jingcan Duan, Pei Zhang, Siwei Wang, Jingtao Hu, Hu Jin, and Jiaxin Zhang. 2023. Normality Learning-based Graph Anomaly Detection via Multi-Scale Contrastive Learning. In *ACM MM*.

[7] Zheng Gong, Guifeng Wang, Ying Sun, Qi Liu, Yuting Ning, Hui Xiong, and Jingyu Peng. 2023. Beyond Homophily: Robust Graph Anomaly Detection via Neural Sparsification. In *International Joint Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:260852064

[8] Ming Jin, Yixin Liu, Yu Zheng, Lianhua Chi, Yuan-Fang Li, and Shirui Pan. 2021. Anemone: Graph anomaly detection with multi-scale contrastive learning. In *CIKM*.

[9] Hwan Kim, Byung Suk Lee, Won-Yong Shin, and Sungsu Lim. 2022. Graph anomaly detection with graph neural networks: Current status and challenges. *IEEE Access* (2022).

[10] Jundong Li, Harsh Dani, Xia Hu, and Huan Liu. 2017. Radar: Residual analysis for anomaly detection in attributed networks.. In *IJCAI*.

[11] Yixin Liu, Zhao Li, Shirui Pan, Chen Gong, Chuan Zhou, and George Karypis. 2021. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE transactions on neural networks and learning systems* (2021).

[12] Xuexiong Luo, Jia Wu, Amin Beheshti, Jian Yang, Xiankun Zhang, Yuan Wang, and Shan Xue. 2022. Comga: Community-aware attributed graph anomaly detection. In *WSDM*.

[13] Julian John McAuley and Jure Leskovec. 2013. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2488388.2488466

[14] Zhen Peng, Minnan Luo, Jundong Li, Huan Liu, Qinghua Zheng, et al. 2018. ANOMALOUS: A Joint Modeling Approach for Anomaly Detection on Attributed Networks.. In *IJCAI*.

[15] Bryan Perozzi and Leman Akoglu. 2016. Scalable anomaly ranking of attributed neighborhoods. In *SDM*.

[16] Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. 2023. A critical look at the evaluation of GNNs under heterophily: are we really making progress? *ArXiv* abs/2302.11640 (2023). https://api.semanticscholar.org/CorpusID:257102689

[17] Hezhe Qiao and Guansong Pang. 2023. Truncated Affinity Maximization: One-class Homophily Modeling for Graph Anomaly Detection. In *Advances in Neural Information Processing Systems*.

[18] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. 2018. DeepInf: Social Influence Prediction with Deep Learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 2110–2119. https://doi.org/10.1145/3219819.3220077

[19] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine* (2008).

[20] Qiaoyu Tan, Ninghao Liu, and Xia Hu. 2019. Deep Representation Learning for Social Network Analysis. *FRONTIERS IN BIG DATA* (2019).

[21] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *SIGKDD*.

[22] Xuhong Wang, Baihong Jin, Ying Du, Ping Cui, Yingshui Tan, and Yupu Yang. 2021. One-class graph neural networks for anomaly detection in attributed networks. *Neural computing and applications* (2021).

[23] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2021. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[24] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas AJ Schweiger. 2007. Scan: a structural clustering algorithm for networks. In *SIGKDD*.

[25] Xu Yuan, Na Zhou, Shuo Yu, Huafei Huang, Zhikui Chen, and Feng Xia. 2021. Higher-order structure based anomaly detection on attributed networks. In *Big Data*.

[26] Jiaqiang Zhang, Senzhang Wang, and Songcan Chen. 2022. Reconstruction Enhanced Multi-View Contrastive Learning for Anomaly Detection on Attributed Networks. In *IJCAI*.

[27] Yu Zheng, Ming Jin, Yixin Liu, Lianhua Chi, Khoa T Phan, and Yi-Ping Phoebe Chen. 2021. Generative and contrastive self-supervised learning for graph anomaly detection. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[28] Shuang Zhou, Qiaoyu Tan, Zhiming Xu, Xiao Huang, and Fu-lai Chung. 2021. Subtractive aggregation for attributed network anomaly detection. In *CIKM*.