

Reconstruction Enhanced Multi-View Contrastive Learning for Anomaly Detection on Attributed Networks

Jiaqiang Zhang^{1,2}, Senzhang Wang³ and Songcan Chen^{1,2*}

¹College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics

²MIT Key Laboratory of Pattern Analysis and Machine Intelligence

³Central South University

{zhangjq, s.chen}@nuaa.edu.cn, szwang@csu.edu.cn

Abstract

Detecting abnormal nodes from attributed networks is of great importance in many real applications, such as financial fraud detection and cyber security. This task is challenging due to both the complex interactions between the anomalous nodes with other counterparts and their inconsistency in terms of attributes. This paper proposes a self-supervised learning framework that jointly optimizes a multi-view contrastive learning-based module and an attribute reconstruction-based module to more accurately detect anomalies on attributed networks. Specifically, two contrastive learning views are firstly established, which allow the model to better encode rich local and global information related to the abnormality. Motivated by the attribute consistency principle between neighboring nodes, a masked autoencoder-based reconstruction module is also introduced to identify the nodes which have large reconstruction errors, then are regarded as anomalies. Finally, the two complementary modules are integrated for more accurately detecting the anomalous nodes. Extensive experiments conducted on five benchmark datasets show our model outperforms current state-of-the-art models.

1 Introduction

Attributed networks are ubiquitous in many real-world scenarios, such as social networks and citation networks. Recently, anomaly detection on attributed networks is of great significance in many security-related applications, such as social spam detection [Jain *et al.*, 2019], financial fraud detection [Wang *et al.*, 2019] and network intrusion detection [Shone *et al.*, 2018], and has attracted rising research interest. Though it has been extensively studied, detecting anomalies on attributed networks is still a challenging task [Ding *et al.*, 2019]. One major challenge of this problem is that the abnormal patterns of nodes are related to not only their interactions with other nodes on the topological structure, but also their inconsistency in terms of node attributes [Ding *et al.*, 2019; Peng *et al.*, 2018].

The early anomaly detection techniques, such as matrix factorization [Li *et al.*, 2017] and OC-SVM [Erfani *et al.*, 2016] have been widely used in many applications. A primary limitation of such methods is that they largely rely on feature engineering constructed by domain experts. Recently, deep learning techniques, especially Graph Neural Networks, have been widely used in various graph mining tasks (e.g. link prediction and node classification), and have achieved considerable performance gains. GNN-based techniques [Ding *et al.*, 2021] have also been adopted to anomaly detection which aim to learn the anomaly-aware node representations. Due to the high cost of obtaining anomaly samples, anomaly detection is usually performed in an unsupervised manner. [Ding *et al.*, 2019] proposed an unsupervised autoencoder-based method to detect the abnormal patterns effectively, which can capture the abnormal nodes according to the reconstruction errors. [Liu *et al.*, 2021b] for the first time proposed a contrastive learning based framework for anomaly detection on graphs. Their model takes the local contextual information as supervision signals, and learns the representative features from the node-subgraph instance pairs. The final discriminative scores are used to detect abnormal nodes. [Jin *et al.*, 2021] performed patch- and context-level contrastive learning via two GNN-based models to further improve the performance.

Although considerable research efforts have been devoted to this problem recently, we argue that the current approaches still have shortcomings due to the following reasons. First, local and global structural information are not well utilized and integrated. Existing SOTA approach [Liu *et al.*, 2021b] typically constructed contrastive instance pair based on node-subgraph to efficiently focus on the local information of a node for anomaly detection. They adopted one-layer GCN to extract the local structure information from the one-hop neighbors, while the global structure information beyond one-hop neighbors cannot be effectively captured and utilized. Previous studies show that higher-order neighboring information is also beneficial for graph mining tasks. [Ding *et al.*, 2019] proposed to use the multi-layer GCN to capture the high-order interactions for anomaly detection on graphs, but the local and global structure information is modeled uniformly and coupled together, leading to information redundancy in feature representation learning. Meanwhile, GCN can be seen as a special form of low-pass filter, which has the

*Corresponding author

effectiveness in smoothing signals [Bo *et al.*, 2021]. Stacking multiple layers of GCN will more likely smooth out the abnormal signals. Therefore, a more effective and decoupled local-global structure information extraction and integration method is needed to provide high quality structure features. Second, how to effectively fuse the node attributes with network topological structures to further boost the detection performance is not well explored. Although GCN can handle both topological structure and attributes, the learned representations are not well suitable for anomaly detection under the unsupervised learning scenario [Liu *et al.*, 2021b]. Anomalous nodes usually present inconsistency in terms of node attributes with their neighboring nodes, which provides us with additional self-supervised signals. How to fully utilize such signals and build an attribute reconstruction model based on the attribute consistency principle to further improve the anomaly detection performance is not well studied.

To alleviate the above drawbacks, we present a **Subgraph**-based multi-view self-supervised framework which contains a **Contrastive**-based module and a **Reconstruction**-based module (**Sub-CR** for convenience). Sub-CR can separate the local information from the global information instead of directly fusing them together, and alleviate the risk of smoothing out abnormal signals in GNN. Specifically, the original graph is first augmented by both graph diffusion and subgraph sampling to obtain local and global view subgraphs for each node. Then a two-view (intra- and inter-view) contrastive learning module is proposed. The intra-view contrastive learning aims to maximize the agreement between the target node and its subgraph-level representations of both views, here the agreement can be quantified as a discriminative score. The inter-view contrastive learning aims to make the discriminative score of the two views closer, which allows the model to encode both local and global information. In order to further utilize the self-supervised signals of the attribute consistency between the neighboring nodes, a masked autoencoder is also introduced to reconstruct the original attributes of the target node based on its neighboring nodes in both views. Finally, the two complementary modules are integrated for anomalous node detection. Our main contributions are summarized as follows:

- A novel self-supervised learning framework is proposed to address the problem of anomaly detection on attributed networks, which can effectively integrate contrastive learning-based and attribute reconstruction-based models.
- A multi-view learning model is proposed. Local and global information are modeled separately first and then integrated together, which can provide high quality features for detecting anomalous nodes.
- We conduct extensive experiments on five benchmark datasets for model evaluation. The results verify the effectiveness of our proposal and it outperforms existing state-of-the-art methods.

2 Related Work

2.1 Anomaly Detection on Attributed Networks

The early work [Perozzi and Akoglu, 2016] proposed a quality measure called normality, which utilized structure and at-

tributes together on attributed networks. [Li *et al.*, 2017] proposed a learning framework to describe the residual of attributes and their consistency with network structure to detect anomalies in attributed networks. With the rapid development of deep neural networks, researchers also tried to adopt deep learning techniques to detect anomalies in attributed networks. [Ding *et al.*, 2019] proposed to use GCN and autoencoder to measure the reconstruction error of nodes from the perspective of structure and attributes to identify anomalies, which can alleviate the network sparsity and data non-linearity issues. [Yu *et al.*, 2018] adopted a clustering-based method through the learned dynamic network representation to dynamically identify network anomalies based on clustering in an incremental manner. [Li *et al.*, 2019] proposed SpecAE to detect global anomalies and community anomalies via a density estimation approach. [Liu *et al.*, 2021b; Jin *et al.*, 2021] proposed a contrastive self-supervised learning framework for graph anomaly detection, effectively capturing the relationship between nodes and their subgraphs to capture the anomalies.

2.2 Graph Contrastive Learning

Recently, contrastive learning as an effective mean of self-supervised learning [Liu *et al.*, 2021a], has gained rising research interest, which aims to learn representations from supervised signals derived from the data itself. The core idea is that it considers two augmented views of the same data instance as positive sampler to be pulled closer, and all other instances are considered as the negative samplers to be pushed farther apart. Due to the great success of self-supervised learning on images [Chen *et al.*, 2020], contrastive learning methods are also adopted to various graph learning tasks. [Velickovic *et al.*, 2019] aimed to learn the node representations by maximizing the mutual information between the patch representation and the corresponding high-level graph summary. [Peng *et al.*, 2020] directly maximized the mutual information between the input and output of the graph neural encoder according to the node features and topology to improve the DGI. [Zhu *et al.*, 2020] proposed to generate two data views and then pull the representation of the same node in the two views closer, push the representation of all other nodes apart. [Zhu *et al.*, 2021] proposed a contrastive framework for unsupervised graph representation learning with adaptive data augmentation.

3 Problem Formulation

In this paper, for the convenience of presentation, the calligraphic fonts (e.g. \mathcal{G}) represent sets, bold lowercase letters and bold uppercase letters represent vectors (e.g. \mathbf{x}) and matrices (e.g. \mathbf{X}), respectively. The terminology and problem definitions are given as follows. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ denote an attributed network, where $\mathcal{V} = \{v_1, \dots, v_N\}$ represents the set of nodes, N is the number of the nodes in the graph, \mathcal{E} is the set of edges, $\mathbf{X} \in \mathbb{R}^{N \times F}$ is the attribute matrix, where $\mathbf{x}_i \in \mathbb{R}^F$ ($i = 1, \dots, N$) is the attributes for the i^{th} node, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix.

Problem Definition. Given an attributed network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ whose adjacency matrix is \mathbf{A} , we aim to learn an

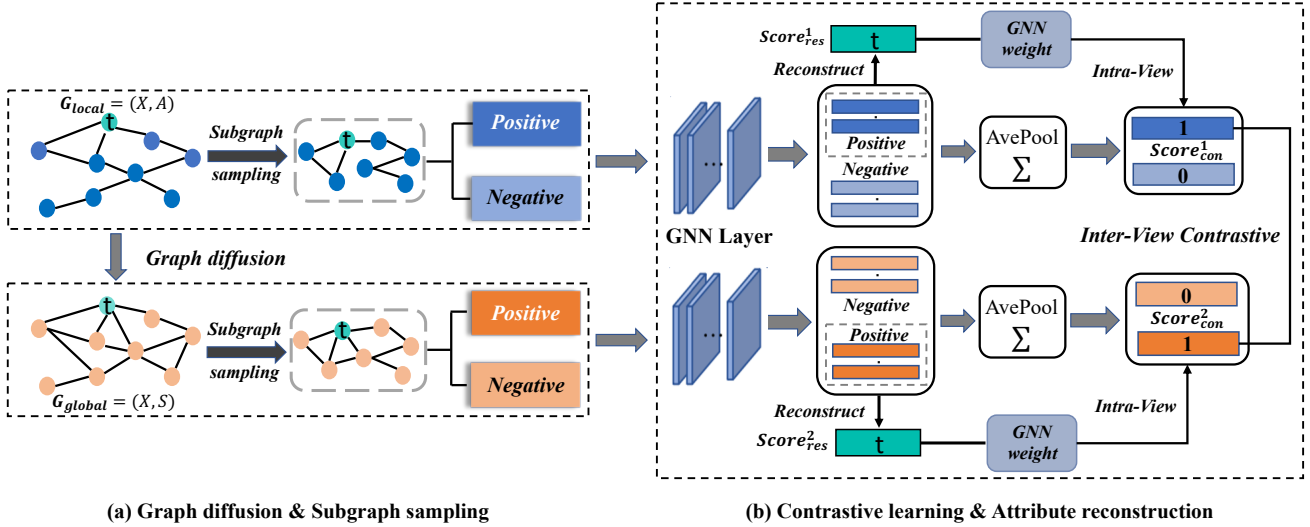


Figure 1: The framework of the proposed model.

anomaly score function $score(\cdot)$ to measure the degree of the abnormality for each node in \mathcal{G} .

4 Methodology

As shown in Figure.1, the proposed Sub-CR contains a contrastive learning-based module and an attribute reconstruction-based module. We first construct two types of data augmentations: the subgraph sampling (left upper part in Figure.1) and the graph diffusion (left lower part in Figure.1), to get local and global view subgraphs for each node. For each view, both the subgraph and the target node are input into a GNN encoder to learn the embedding as shown in the middle part in Figure.1. Then we establish the intra- and inter-view contrastive learning (right part in Figure.1) to encode local and global structure information. Meanwhile, we further introduce the masked autoencoder to identify the anomalous nodes through the attribute reconstruction errors in two views. Finally, the two complementary modules are integrated to give a final anomaly score.

4.1 Multi-View Graph Generation with Graph Diffusion and Subgraph Sampling

The effectiveness of the contrastive learning paradigm largely relies on the choice of the pretext task and the data augmentations [Liu *et al.*, 2021a]. We adopt the node-subgraph matching pattern based pretext task which is proved useful for the anomaly detection on attributed networks [Liu *et al.*, 2021b]. Meanwhile, the graph diffusion augmentation [Hassani and Khasahmadi, 2020] is introduced to get the global view of the structure, which allows each node to match with its local- and global-view subgraphs respectively. Next, we introduce how to generate the two views of graph.

Graph Diffusion. We propose to use the graph diffusion as the first augmentation to obtain the global view graph. Graph diffusion can effectively capture the global structure information [Hassani and Khasahmadi, 2020; Klicpera *et al.*, 2019].

This process is formulated as follows:

$$\mathbf{S} = \sum_{k=0}^{\infty} \theta_k \mathbf{T}^k \in \mathbb{R}^{N \times N} \quad (1)$$

where θ_k is the weighting coefficient to control the proportion of local and global structure information and $\mathbf{T} \in \mathbb{R}^{N \times N}$ denotes the generalized transition matrix to transfer the adjacency matrix. Note that $\theta_k \in [0, 1]$ and $\sum_{k=0}^{\infty} \theta_k = 1$. In this paper, Personalized PageRank (PPR) [Page *et al.*, 1999] is adopted to power the graph diffusion. Given the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, the identity matrix \mathbf{I} and its degree matrix \mathbf{D} , the transition matrix and the weight can be formulated respectively as $\mathbf{T} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ and $\theta_k = \alpha(1 - \alpha)^k$. Then the graph diffusion \mathbf{S} can be reformulated as:

$$\mathbf{S} = \alpha(\mathbf{I} - (1 - \alpha)\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2})^{-1} \quad (2)$$

where α denotes the teleport probability in a random walk. The graph diffusion \mathbf{S} is the global view of the initial graph.

Subgraph Sampling. We adopt random walk with restart (RWR) [Tong *et al.*, 2006] to sample the local subgraph. In both views, a subgraph with the size P is sampled for each node. In particular, we sample nodes and their edges from the original view to form the local-view subgraph, and then select the exact nodes and edges from the other view to form the global-view subgraph. In each view, for node v_i , its subgraph is regarded as its positive pair, and a subgraph corresponding to the other node is regarded as the negative pair.

4.2 Local & Global Graph Contrastive Learning

As mentioned before, the representation of a normal node is similar to that of the neighboring nodes in its subgraph while this does not hold for an abnormal node. Therefore, we define the intra-view contrastive learning in the local and the global views, which maximizes the agreement between the node and the corresponding subgraph-level representations. As shown

in the right part of Figure.1, for node v_i , the discriminative score of its positive pair should be close to 1, while the negative one should be close to 0. Meanwhile, the inter-view contrastive learning is defined between the two views. It aims to encourage the discriminative score that the same node between its subgraph in two views to be close.

Intra-View Contrastive Learning. This contrastiveness is defined in each view. Taking the local view in Figure.1 as an example, we obtain the corresponding positive and negative subgraphs $\mathcal{G}_i, \tilde{\mathcal{G}}_i$ for the target node v_i . Note that in order to make the obtained representation more discriminative, the attributes of the target node in the subgraph are masked. Then the subgraph are fed into a shared GCN encoder to learn a low-dimensional representation, which can be formulated as follows:

$$\mathbf{H}_i^l = \phi(\hat{\mathbf{D}}_i^{-1/2} \mathbf{A}'_i \hat{\mathbf{D}}_i^{-1/2} \mathbf{H}_i^{(l-1)} \mathbf{W}^{(l-1)}) \quad (3)$$

where $\hat{\mathbf{D}}_i \in \mathbb{R}^{N \times N}$ is the degree matrix of $\mathbf{A}'_i = \mathbf{A}_i + \mathbf{I}_N$, \mathbf{A}'_i is the subgraph adjacency matrix with self-loop, \mathbf{I}_N is the identity matrix and $\mathbf{W}^0 \in \mathbb{R}^{F \times d}$ is a learnable weight matrix. Since the target node is masked in the subgraph, we employ the weight matrix of GCN to map the features into the same embedding space. It can be formulated as:

$$\mathbf{h}_i^l = \phi(\mathbf{h}_i^{l-1} \mathbf{W}^{l-1}) \quad (4)$$

where $h_i^0 = \mathbf{x}_i$, ϕ is the activation function such as ReLU. So we obtain the representations \mathbf{H}_i and $\tilde{\mathbf{H}}_i$ for the two-view subgraphs $\mathcal{G}_i, \tilde{\mathcal{G}}_i$, \mathbf{h}_i for the target node v_i . In order to apply the node-subgraph matching pattern pretext task (the representation of node and its corresponding subgraph is more consistent), we take the average pooling function as the readout module to obtain the subgraph-level embedding vector \mathbf{e}_i :

$$\mathbf{e}_i = \text{Readout}(\mathbf{H}_i) = \sum_{k=1}^{n_i} \frac{(\mathbf{H}_i)_k}{n_i} \quad (5)$$

where n_i is the number of nodes in the subgraph \mathcal{G}_i . So the discriminative scores for the positive pair can be defined as:

$$s_i = \sigma(\mathbf{h}_i \mathbf{W}_s \mathbf{e}_i^T) \quad (6)$$

where \mathbf{W}_s is a learnable matrix. Similarly, we can calculate \tilde{s}_i for the negative pair. Hence, the local view contrastive loss for the instance pair $(v_i, \mathcal{G}_i, \tilde{\mathcal{G}}_i)$ can be formulated as:

$$\mathcal{L}_{intra}^1(v_i) = -\frac{1}{2}(\log(s_i) + \log(1 - \tilde{s}_i)) \quad (7)$$

Besides, we can similarly calculate \mathcal{L}_{intra}^2 for the global view. Finally, by combining the two losses, we have the intra-view contrastive learning objective function defined below:

$$\mathcal{L}_{intra} = \frac{1}{2N} \sum_{i=1}^N (\mathcal{L}_{intra}^1(v_i) + \mathcal{L}_{intra}^2(v_i)) \quad (8)$$

Inter-View Contrastive Learning. This contrastiveness is defined between the two views. Following the core idea in contrastive learning, we make the discriminative scores from node-subgraph pair of the two views closer. So the inter-view

contrastive loss between the local- and global-view can be formulated as:

$$\mathcal{L}_{inter} = (\|\mathbf{s}_1 - \mathbf{s}_2\|_F^2) \quad (9)$$

where $\mathbf{s}_1, \mathbf{s}_2$ are the vectors that consist of positive pair discriminative scores in the two views. By combining the intra- and the inter-view contrastive learning, the overall loss function of the multi-view contrastive learning module is:

$$\mathcal{L}_{con} = \mathcal{L}_{inter} + \mathcal{L}_{intra} \quad (10)$$

4.3 Attribute Reconstruction Based on Neighbors

The self-supervised learning method based on masked autoencoders (MAE) has been proved to be effective in many fields. For example, the BERT [Devlin *et al.*, 2018]-based pre-training model has achieved remarkable results in many NLP tasks, and similar works are also proposed in CV [He *et al.*, 2021]. The basic idea of these methods is to mask a part of a sentence or a picture and then input the unmasked part into the encoder and decoder for predicting the masked part.

Motivated by the masked autoencoder method, we propose to design an attribute reconstruction based on neighbors module to boost the performance. We adopt an asymmetric design. The encoder operates on the nodes in the subgraph that are not masked. The latent representations of these nodes are concatenated as the input of the lightweight decoder to reconstruct the masked node's raw attributes. Taking the local view as an example, for the subgraph \mathcal{G}_i of node v_i , we can get the representation of \mathcal{G}_i through the GCN encoder. The reconstruction loss of the local view can be defined as:

$$\mathcal{L}_{res}^1(v_i) = \|g(\mathbf{Z}_i) - x_i\|^2 \quad (11)$$

where $g(\cdot)$ is the multilayer perceptron, \mathbf{Z}_i is the concatenation of neighboring nodes' representations in the subgraph. Similarly, we can calculate \mathcal{L}_{res}^2 for the global view. The overall loss function of this module can be formulated as:

$$\mathcal{L}_{res} = \frac{1}{2N} \sum_{i=1}^N (\mathcal{L}_{res}^1(v_i) + \mathcal{L}_{res}^2(v_i)) \quad (12)$$

4.4 Anomaly Score Inference

To jointly train the contrastive learning module and the attribute reconstruction module, we optimize the following objective function:

$$\mathcal{L} = \mathcal{L}_{con} + \gamma \mathcal{L}_{res} \quad (13)$$

where γ is a controlling parameter which balances the importance of the two modules.

By minimizing the above objective function, we can compute the anomaly score of each node. Inspired by [Liu *et al.*, 2021b], taking the node v_i as an example, the anomaly score of the contrastive-based module can be calculated by:

$$\text{score}_{con}(v_i) = \frac{1}{2}[\text{score}_{con}^1(v_i) + \text{score}_{con}^2(v_i)] \quad (14)$$

where $\text{score}_{con}^1(v_i) = s_i^{1-} - s_i^{1+}$, $\text{score}_{con}^2(v_i) = s_i^{2-} - s_i^{2+}$, $s_i^{1+}(s_i^{2+})$ and $s_i^{1-}(s_i^{2-})$ are the local (global) view discriminative scores of positive and negative pairs by Eq.(6). The anomaly score from the reconstruction-based module is:

$$\text{score}_{res}(v_i) = \frac{1}{2} \sum_{k=1}^2 \text{score}_{res}^k(v_i) \quad (15)$$

Datasets	Nodes	Edges	Features	Anomalies
BlogCatalog	5196	171743	8189	300
Flickr	7575	239738	12407	450
Cora	2708	5429	1433	150
CiteSeer	3327	4732	3703	150
Pubmed	19717	44338	500	600

Table 1: The statistics of the datasets.

$$score_{res}^k(v_i) = \|g(\mathbf{Z}_i^k) - x_i\|_2^2, k = 1, 2 \quad (16)$$

where \mathbf{Z}_i^1 and \mathbf{Z}_i^2 are the concatenated neighboring node representations in local and global subgraphs for node v_i , respectively. The two scores are first normalized, and then integrated by the following formula:

$$score(v_i) = score_{con}(v_i) + \gamma score_{res}(v_i) \quad (17)$$

5 Experiments

5.1 Datasets

We evaluate the proposed model on five benchmark datasets that are widely used in anomaly detection on attributed networks [Liu *et al.*, 2021b; Zheng *et al.*, 2021]. These datasets include two social network datasets BlogCatalog and Flickr and three citation network datasets Cora, Citeseer, and Pubmed. Due to the shortage of ground truth anomalies, the anomalies are injected by the perturbation scheme [Ding *et al.*, 2019]. The statistics of the datasets is shown in Table 1.

5.2 Experimental Settings

Baselines and evaluation metrics. In the experiments, we compare the proposed model with seven baselines. [Perozzi and Akoglu, 2016] evaluates the attribute correlation of node to detect anomalies. Specifically, it analyzes the abnormality of each node from the ego-network point of view. [Li *et al.*, 2017] is a residual analysis-based method. It detects anomalies whose behaviors are singularly different from the majority [Peng *et al.*, 2018] proposes a novel joint framework to conduct attribute selection and anomaly detection. [Ding *et al.*, 2019] conducts the investigation on the problem of anomaly detection on attributed networks. [Velickovic *et al.*, 2019] is an approach for learning node representations. A bilinear function is provided to score abnormality. [Liu *et al.*, 2021b] is the first contrastive learning based method for graph anomaly detection. [Jin *et al.*, 2021] is a recent method which performs patch- and context-level contrastive learning for anomaly detection on attributed networks. We use ROC and AUC to measure the performance.

Implementation Details. In Sub-CR, we set the node size P of subgraph to 4 for all datasets by considering both the efficiency and performance. Following [Liu *et al.*, 2021b], we employ a one-layer GCN as the encoder, and the embedding dimension is set to 64. The model is optimized with the Adam optimizer during training. The batch size is set to 300 for all datasets. The learning rate is set to 0.001 for Cora, Citeseer, Pubmed and Flickr, set to 0.003 for BlogCatalog. We train the model 400 epochs for BlogCatalog and Flickr, and 100

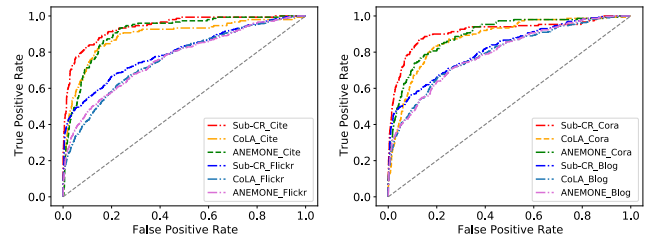


Figure 2: ROC curves on four benchmark datasets.

epochs for Cora, Citeseer and Pubmed. The parameter γ is set to 0.6 for BlogCatalog, Flickr, Cora and Citeseer and 0.4 for Pubmed. In the inference phase, we average the anomaly scores in Eq. (17) over 300 rounds to get the final anomaly score for each node.

5.3 Result and Analysis

Table 2 shows the comparison result of these methods in terms of AUC value. Due to space limitation, Figure.2 shows the ROC curves on Cora, BlogCatalog, Citeseer, and Flickr with two competitive baselines. As shown in the table, one can first see that our proposed model is superior to its counterparts on five benchmark datasets, which demonstrates the effectiveness of Sub-CR. To be specific, Sub-CR improves the performance by 3.38% and 1.61% over the best baseline ANEMONE on Flickr and Pubmed datasets, respectively. In details, the shallow methods including AMEN, Radar and ANOMALOUS perform worse than other models due to their less effectiveness in modeling the high-dimension node attributes and complex topological structures. The DOMINANT and DGI also do not show the competitive performance because they focus on modeling the whole graph structure instead of directly explore the abnormal patterns. DGI takes the node vs full-graph instance pair contrastive, and DOMINANT aims to reconstruct the whole graph structure or attributes for each node, which do not decouple local and global information for anomaly detection. CoLA and ANEMONE achieve the suboptimal performance due to the consideration of surrounding substructures. But they cannot fully catch the high-order structure information and ignore the self-supervised signal of the original attribute reconstruction error. The superior performance of Sub-CR verifies the effectiveness of intergating the multi-view contrastive-based and reconstruction-based modules, which can decouple local and global information to better capture the attribution and the topological structure of nodes for anomaly detection.

5.4 Ablation Study

We next conduct an ablation study to verify the effectiveness of each component in Sub-CR. Sub-R and Sub-C denote the model without reconstruction-based and contrastive-based module, respectively. The variants Sub-weight and Sub-global are defined as that we remove the balance weight (γ is set to 1) and the global view, respectively. The results are presented in Table 3. One can see that Sub-CR outperforms all the variants consistently on all the datasets, which demonstrates the components are all helpful to the studied task. Especially, Sub-CR outperforms Sub-C by 6.81%, 5.41%,

Methods	Blogcatalog	Flickr	Cora	Citeseer	Pubmed
AMEN ^[2016]	0.6392	0.6573	0.6266	0.6154	0.7713
Radar ^[2017]	0.7401	0.7399	0.6587	0.6709	0.6233
ANOMALOUS ^[2018]	0.7237	0.7434	0.5770	0.6307	0.7316
DOMINANT ^[2019]	0.7468	0.7442	0.8155	0.8251	0.8081
DGI ^[2019]	0.5827	0.6237	0.7511	0.8293	0.6962
CoLA ^[2021]	0.7854	0.7513	0.8779	0.8968	0.9512
ANEMONE ^[2021]	0.8067	0.7637	0.9057	0.9189	0.9548
Sub-CR	0.8141	0.7975	0.9132	0.9303	0.9709

Table 2: The AUC values comparison on five benchmark datasets.

	BlogCatalog	Flickr	Cora	CiteSeer	Pubmed
Sub-R	0.7943	0.7609	0.9002	0.9017	0.9553
Sub-C	0.7460	0.7434	0.8220	0.7892	0.8006
Sub-weight	0.8083	0.7928	0.9041	0.9275	0.9491
Sub-global	0.8090	0.7923	0.8975	0.9195	0.9625

Table 3: The AUC values of ablation study.

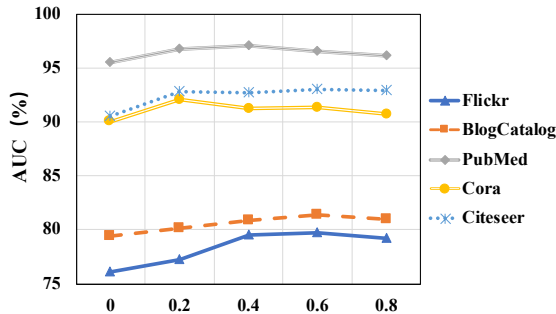


Figure 3: Performance with different γ .

9.12%, 14.11%, 17.03% on five datasets, respectively, which demonstrates the contrastive-based module is crucially important to the model. The self-supervised signal of the matching pattern on the node-subgraph pair is more effective for capturing anomalies. As for the remaining three components, Sub-R performs worse than Sub-weight and Sub-global on BlogCatalog, Flickr, and CiteSeer datasets. On the Cora, the component of the global view is more important. Sub-weight has a more significant effect on Pubmed. The reason may be that the effect of different modules on different datasets varies due to the different data characteristics.

5.5 Parameter Study

We finally conduct the model sensitivity analysis on critical hyper-parameters in Sub-CR, which are the size P of subgraph, the embedding dimension in GCN and the balance factor γ . Figure.3 shows the effect of different γ values on the model performance. One can see that the most datasets can achieve the best result and not sensitive to the balance factor when $\gamma \geq 0.4$. Figure.4 shows the AUC values under different subgraph sizes and embedding dimensions. One can observe that the best performance for Flickr and Citeseer is

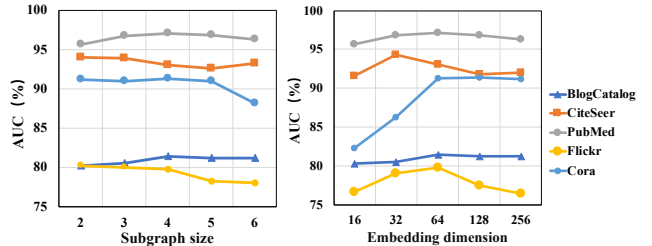


Figure 4: Performance with different subgraph sizes and embedding dimensions.

achieved at $P = 2$, and for BlogCatalog, Cora and PubMed it is achieved at $P = 4$. When the size is too large, the subgraph will contain redundant information, which will hurt the performance. A suitable embedding dimension is $d = 64$ for most datasets as shown in the figure. A too small or large embedding dimension will both degrade the model performance.

6 Conclusion

In this paper, we propose a novel multi-view self-supervised learning framework for anomaly detection on attributed networks. The proposed contrastive learning-based module consists of two carefully designed contrastive views to better capture local and global structure information related to anomaly patterns. The attribute reconstruction module adopts the representation of those neighbors based on the subgraph to rebuild the raw attributes of the target node of the two views. Finally, the two complementary modules are integrated for more effective anomaly detection. Extensive experiments conducted over five benchmark datasets demonstrate the effectiveness of our proposed model.

Acknowledgements

This work was supported by NSFC under granted No. 62076124, partially financially supported by the National Science and Technology Major Project (J2019-IV-0018-0086), NSFC (No. 62172443), and the Fundamental Research Funds for the Central Universities (No. NZ2020014).

References

- [Bo *et al.*, 2021] Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *AAAI*, 2021.
- [Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.
- [Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [Ding *et al.*, 2019] Kaize Ding, Jundong Li, Rohit Bhanushali, and Huan Liu. Deep anomaly detection on attributed networks. In *SDM*, 2019.
- [Ding *et al.*, 2021] Kaize Ding, Jundong Li, Nitin Agarwal, and Huan Liu. Inductive anomaly detection on attributed networks. In *IJCAI*, 2021.
- [Erfani *et al.*, 2016] Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121–134, 2016.
- [Hassani and Khasahmadi, 2020] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *ICML*, 2020.
- [He *et al.*, 2021] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. *arXiv preprint arXiv:2111.06377*, 2021.
- [Jain *et al.*, 2019] Gauri Jain, Manisha Sharma, and Basant Agarwal. Spam detection in social media using convolutional and long short term memory neural network. *Annals of Mathematics and Artificial Intelligence*, 85(1):21–44, 2019.
- [Jin *et al.*, 2021] Ming Jin, Yixin Liu, Yu Zheng, Lianhua Chi, Yuan-Fang Li, and Shirui Pan. Anemone: Graph anomaly detection with multi-scale contrastive learning. In *CIKM*, 2021.
- [Klicpera *et al.*, 2019] Johannes Klicpera, Stefan Weisberger, and Stephan Günnemann. Diffusion improves graph learning. *arXiv preprint arXiv:1911.05485*, 2019.
- [Li *et al.*, 2017] Jundong Li, Harsh Dani, Xia Hu, and Huan Liu. Radar: Residual analysis for anomaly detection in attributed networks. In *IJCAI*, pages 2152–2158, 2017.
- [Li *et al.*, 2019] Yuening Li, Xiao Huang, Jundong Li, Mengnan Du, and Na Zou. Specac: Spectral autoencoder for anomaly detection in attributed networks. In *CIKM*, 2019.
- [Liu *et al.*, 2021a] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 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*, 2021.
- [Page *et al.*, 1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- [Peng *et al.*, 2018] Zhen Peng, Minnan Luo, Jundong Li, Huan Liu, and Qinghua Zheng. Anomalous: A joint modeling approach for anomaly detection on attributed networks. In *IJCAI*, 2018.
- [Peng *et al.*, 2020] Zhen Peng, Wenbing Huang, Minnan Luo, Qinghua Zheng, Yu Rong, Tingyang Xu, and Junzhou Huang. Graph representation learning via graphical mutual information maximization. In *WWW*, 2020.
- [Perozzi and Akoglu, 2016] Bryan Perozzi and Leman Akoglu. Scalable anomaly ranking of attributed neighborhoods. In *SDM*, 2016.
- [Shone *et al.*, 2018] Nathan Shone, Tran Nguyen Ngoc, Vu Dinh Phai, and Qi Shi. A deep learning approach to network intrusion detection. *IEEE transactions on emerging topics in computational intelligence*, 2(1):41–50, 2018.
- [Tong *et al.*, 2006] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. Fast random walk with restart and its applications. In *ICDM*, 2006.
- [Velickovic *et al.*, 2019] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *ICLR*, 2019.
- [Wang *et al.*, 2019] Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, Shuang Yang, and Yuan Qi. A semi-supervised graph attentive network for financial fraud detection. In *ICDM*, 2019.
- [Yu *et al.*, 2018] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. Network: A flexible deep embedding approach for anomaly detection in dynamic networks. In *KDD*, 2018.
- [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*, 2021.
- [Zhu *et al.*, 2020] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.
- [Zhu *et al.*, 2021] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. In *WWW*, 2021.