Contrastive Attributed Network Anomaly Detection with Data Augmentation

Zhiming Xu[1], Xiao Huang[2], Yue Zhao[3], Yushun Dong[1], and Jundong Li[1(✉)]

[1] University of Virginia, Charlottesville, USA
{zx2rw,yd6eb,jundong}@virginia.edu
[2] The Hong Kong Polytechnic University, Hung Hom, Hong Kong
xiao.huang@polyu.edu.hk
[3] Carnegie Mellon University, Pittsburgh, USA
zhaoy@cmu.edu

Abstract. Attributed networks are a type of graph structured data used in many real-world scenarios. Detecting anomalies on attributed networks has a wide spectrum of applications such as spammer detection and fraud detection. Although this research area draws increasing attention in the last few years, previous works are mostly unsupervised because of expensive costs of labeling ground truth anomalies. Many recent studies have shown different types of anomalies are often mixed together on attributed networks and such invaluable human knowledge could provide complementary insights in advancing anomaly detection on attributed networks. To this end, we study the novel problem of modeling and integrating human knowledge of different anomaly types for attributed network anomaly detection. Specifically, we first model prior human knowledge through a novel data augmentation strategy. We then integrate the modeled knowledge in a Siamese graph neural network encoder through a well-designed contrastive loss. In the end, we train a decoder to reconstruct the original networks from the node representations learned by the encoder, and rank nodes according to its reconstruction error as the anomaly metric. Experiments on five real-world datasets demonstrate that the proposed framework outperforms the state-of-the-art anomaly detection algorithms.

Keywords: Anomaly detection · Graph neural networks · Self-supervised learning

1 Introduction

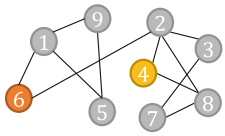
Attributed networks are a kind of graph structured data, which exists ubiquitously in many real-world scenarios, such as social networks, biological networks, and financial transaction networks [1,22]. Over the past few decades, many research efforts have been devoted to performing different learning tasks on attributed networks. Anomaly detection is one such task, which in the context of attributed networks aims to identify nodes with significantly different

J. Gama et al. (Eds.): PAKDD 2022, LNAI 13281, pp. 444–457, 2022.
https://doi.org/10.1007/978-3-031-05936-0_35

patterns from other nodes in terms of their attributes, communities, etc. [1,28]. It has become a critical research area that has broad applications in various real-world scenarios [4], such as spammer detection [1] and fraud detection [3].

Extensive progress has been made towards anomaly detection on attributed networks over the past few years [8–10,19,20,25,30,31]. Generally speaking, existing anomaly detection approaches can be mainly divided into two mainstreams, namely Non-deep Learning (Non-DL) methods and Deep Learning (DL) methods. Non-DL methods typically rely on various types of heuristic anomaly measurements [30,31,34,35] or employ matrix decomposition techniques [19,20,29] to detect anomalies while DL methods often resort to Graph Neural Networks (GNNs) for the detection of anomalies [8,10,21]. It should be noted that DL methods have shown superior performance over traditional Non-DL methods [19,25,30,31] due to the strong capability of GNNs for learning node representations. Specifically, DL methods usually follow an encoder-decoder learning scheme, where the encoder takes the given attributed network as input, while the decoder reconstructs the graph structure and node attributes and compares the reconstructed data with the original input for anomaly detection [8–10]. However, despite the superior performance, these approaches mainly detect anomalies in an unsupervised manner due to the expensive labeling cost of ground truth anomalies. Many recent studies have shown that there often exist mixed types of anomalies on attributed networks, w.r.t. graph structure and node attributes [19,44].

For example, we present two typical anomaly types, namely attribute anomaly and structure anomaly in Fig. 1. There are a community of CA software engineers and a community of MA salesperson in this network. For attribute anomaly, the attribute value of node 4 is significantly different from others, thus it is suspicious to be an attribute anomaly; for structure anomaly, node 6 belongs to the CA software engineer community by its attributes, however, it also connects to a remotely related community of MA salesperson, rendering it structurally abnormal. Beyond the anomalies in the above example, more types of commonly encountered anomalies, e.g., community anomalies, have also been identified and summarized by existing works [1,20]. As a summary, these studies equipped us with rich prior human knowledge of different anomaly types. In fact, many learning related problems have witnessed a significant performance improvement when human knowledge is considered [33,36,42]. Motivated by such success, in this paper, we study an important research problem: *whether the prior human knowledge of different anomaly types could be harnessed to advance anomaly detection on attributed networks.*



Node	Location	Occupation
1, 5, 6, 9	CA	Software Engineer
2, 3, 7, 8	MA	Salesperson
4	AL	Bank Clerk

Fig. 1. A toy example of attribute anomaly and structure anomaly on an attributed network.

Although leveraging prior human knowledge of different anomaly types could be potentially helpful for attributed network anomaly detection, how to properly model and utilize such knowledge remains a daunting task mainly because of the following two challenges: (1) *Knowledge Modeling Challenge.* How to properly model the prior human knowledge of different types of anomalies on attributed networks is the first challenge that needs to be tackled. The major problem here is that such knowledge only encodes human understanding of possible anomalous patterns on attributed networks, thus it does not have a concrete form and cannot be directly leveraged. While many existing studies proposed to model human knowledge as an invaluable data resource in addition to the original input data [18,33,42], it still remains unclear how to model human knowledge into concrete data resource that can be directly utilized in our case. (2) *Knowledge Integration Challenge.* The second challenge centers around integrating the prior human knowledge of anomaly types on attributed networks seamlessly into the detection model. Traditionally, many existing works regard human knowledge as an explicit supervision signal and integrate it into learning models by designing a specific loss term [6,26,36]. However, in our problem, existing human knowledge of anomaly types is not exhaustive, and an effective knowledge integration mechanism needs to be flexible enough to accommodate the available knowledge rather than design a flawed loss term informed only by partial observation.

To tackle the above challenges, in this paper, we propose CONTRASTIVE ANOMALY DETECTION (CONAD), a principled contrastive anomaly detection framework on attributed networks. CONAD is capable of identifying anomalous nodes on attributed networks by leveraging the prior human knowledge of different anomaly types. First, to tackle the knowledge modeling challenge, we propose a novel data augmentation strategy which explicitly models and formalizes the prior human knowledge of different anomaly types as contrastive samples (i.e., nodes whose patterns deviate significantly from existing nodes on the input attributed network) on the augmented attributed network. Second, to address the knowledge integration challenge, we propose to tightly integrate the contrastive samples on the augmented attributed network into the anomaly detection model with a well-designed contrastive loss. Methodologically, we first propose to generate an augmented attributed network to model known anomaly types. A Siamese GNN is employed as the encoder function to map both the input attributed network and the augmented attributed network into an embedding space. After that, a contrastive loss is designed based upon the normal nodes on the input attributed network and contrastive samples on the augmented attributed network, through which the human knowledge of different anomaly types can be well harnessed. The proposed contrastive loss is jointly considered with a graph reconstruction loss for end-to-end model training. During the detection phase, the suspicious score of each node is measured by the magnitude of the reconstruction error, which serves as the metric to identify anomalies, i.e., a larger error indicates the node has a higher chance of being abnormal.

The main contributions of this paper can be summarized as follows: (1) **Problem Formulation.** We study a novel problem of modeling and leveraging prior human knowledge of different anomaly types for anomaly detection on attributed networks. (2) **Algorithmic Design.** We propose a principled framework that models prior human knowledge of different anomaly types as contrastive samples in the augmented attributed network; and integrates the contrastive samples into the anomaly detection model with a well-designed contrastive loss. (3) **Experimental Evaluations.** We perform comprehensive experimental evaluations on real-world datasets to demonstrate the superiority of the proposed contrastive attributed network anomaly detection framework.

2 Problem Definition

Notations. We use bold uppercase letters (e.g. $\mathbf{A}$), bold lowercase letters (e.g. $\mathbf{x}$), and regular lowercase letters (e.g. a) to denote matrices, vectors, and scalars, respectively. Besides, for a matrix $\mathbf{A}$, we represent its (i, j)-th entry as $\mathbf{A}_{ij}$. Similarly, for a vector $\mathbf{y}$, its i-th element is denoted by y_i.

Let $\mathcal{G} = \{\mathbf{A}, \mathbf{X}\}$ be an input attributed network, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{X} \in \mathbb{R}^{n \times d}$ denote the adjacency matrix and attribute matrix, respectively. The problem of *anomaly detection on attributed networks* aims to assign a suspicion score to each node that quantifies how likely it is to be abnormal. To utilize prior human knowledge of anomaly types in this process, we assume there is an additional human knowledge input ξ that consists of typical types of anomalies studied in previous works and observed in real-world scenarios [1,8,22], e.g., attribute and structure anomalies shown in Fig. 1 before. With the additional knowledge ξ, we hence formulate the following research problem.

Definition 1 ***Modeling and Leveraging Prior Human Knowledge of Anomaly Types for Attributed Network Anomaly Detection.*** *Given an attributed network $\mathcal{G} = \{\mathbf{A}, \mathbf{X}\}$, prior human knowledge ξ of anomaly types, our goal is to model and formalize the abstract human knowledge ξ into concrete data ($\mathrm{M}(\xi)$ denoted as $\mathrm{M}(\xi)$), and then integrate it into a principled detection model f that is capable of encoding both $\mathrm{M}(\xi)$ and $\mathcal{G}$ and ultimately detect anomalies in $\mathcal{G}$.*

3 The Proposed Framework

In this section, we introduce the proposed framework CONAD. It consists of three major components as shown in Fig. 2, namely, knowledge modeling module, knowledge integration module, and anomaly detection module. The overview of each module is listed below followed by detailed descriptions.

Knowledge Modeling Module. Given the prior human knowledge ξ of different anomaly types, we first use a novel data augmentation strategy to model and formalize it as concrete contrastive samples. We achieve this by introducing each known anomaly type encoded in ξ to the input attributed network $\mathcal{G}$ and generate the augmented attributed network $\mathcal{G}_{ano}$ accordingly.

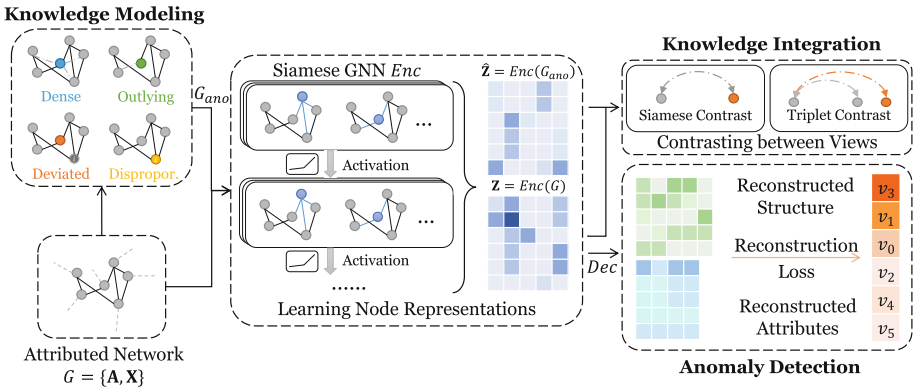


Fig. 2. Overview of CONAD. The lower-left box is the input attributed network for anomaly detection. The upper-left box shows the knowledge modeling module. Dense, Outlying, Deviated, and Disproportionate correspond to the prior human knowledge of anomaly types. The middle is the encoder built on Siamese GNN to learn node representations. The upper-right box presents two contrast strategies to integrate the prior human knowledge modeled in $\mathcal{G}_{ano}$. The lower-right part is the decoder that reconstructs both the structure and attributes of the input attributed network, which detects anomalies with the reconstruction error.

Knowledge Integration Module. After modeling prior human knowledge ξ, we feed both $\mathcal{G}$ and each $\mathcal{G}_{ano}$ into a graph encoding architecture in which a Siamese GNN acts as the encoder to learn representations of nodes. By using a Siamese network, both graphs will be encoded into the same latent space, making it possible to contrast between the node representations of $\mathcal{G}$ and $\mathcal{G}_{ano}$. After the encoding phase, to tightly integrate the human knowledge in $\mathcal{G}_{ano}$, we propose a well-designed contrastive loss. Specifically, the contrastive loss will guide the encoder to represent normal nodes on the input attributed network and contrastive samples on the augmented attributed network differently. Consequently, anomaly patterns of the augmented nodes can be captured.

Anomaly Detection Module. With the learned node representations, we aim to reconstruct the graph structure and node attributes of the input attributed network $\mathcal{G}$ with a decoder. The reconstruction errors produced by the reconstruction phase are leveraged as suspicion scores in detecting anomalies on $\mathcal{G}$.

3.1 Knowledge Modeling Module

We introduce the data augmentation strategy used to model the prior human knowledge of different anomaly types on attributed networks in this subsection. We consider four different types of anomalies on attributed networks (from both the structure side and the attribute side), and introduce a certain amount of anomalies belonging to each anomaly type to the input attributed network $\mathcal{G}$ to form an augmented attributed network $\mathcal{G}_{ano}$. Each of these four augmented anomaly types is illustrated in Fig. 3.

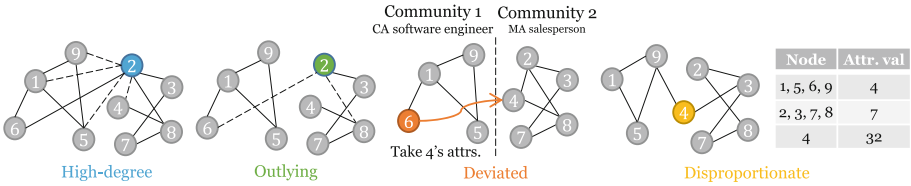


Fig. 3. An illustration of four different types of anomalies on attributed networks based on prior human knowledge.

Structure – high-degree. In social networks, spammers often follow and interact with excessively numerous users [14]. To simulate this anomaly type, we choose a certain amount of nodes with average degrees, and then connect them to many other random nodes. The chosen nodes thus have an unusually high degree and are considered structurally abnormal in the attributed network.

Structure – outlying. Another abnormal account type in social or e-commerce networks is created in large quantities to spam certain posts [38]. They behave like regular users but few users will follow them, and thus they do not belong to any communities, thus different from the majority of the whole network and deemed structurally abnormal. We simulate anomaly type by choosing a certain amount of nodes and drop most of their edges on the input attributed network.

Attribute – deviated. A common attribute anomaly on attributed network is a node with deviated attribute values from its neighbors [31]. In other words, the attribute value of this node could be rather different from others in the same community. To model this anomaly type, we first choose certain center nodes. For each center node, we randomly sample a number of other nodes from the entire network. We then calculate the similarity between the attribute vectors of this center node and the others, and then assign the attribute vector of the least similar one to the center node. In fact, through such generation process, we are introducing community anomalies to the input attributed network.

Attribute – disproportionate. In e-commerce websites, dishonest sellers might want to promote their products by setting unreasonably low prices or achieve high sale volumes by recruiting dishonest buyers [11]. Both of these sale frauds will result in unusually small or large numbers in certain node attributes. We hence largely scale up or scale down the values of certain node attributes with a preset probability to simulate this anomaly type of disproportionate numerical values in certain node attributes.

After applying the four augmentation strategies above, we obtain an augmented attributed network $\mathcal{G}_{ano}$, referred to as *anomalous view*. In the anomalous view $\mathcal{G}_{ano}$, we have a label vector $\mathbf{y}$, where $y_i = 1$ denotes that node i corresponds to one of those four known anomaly types, and $y_i = 0$ otherwise.

3.2 Knowledge Integration Module

Now, we integrate the modeled human knowledge of anomaly types into the detection model through two essential components: (1) learning node representations; and (2) contrasting between different views.

Learning Node Representations. We first discuss how to encode both $\mathcal{G}$ and $\mathcal{G}_{ano}$. In particular, we employ a Siamese GNN architecture as an encoder to learn embeddings for nodes in both $\mathcal{G}$ and $\mathcal{G}_{ano}$. Generally, various GNNs can be leveraged to learn node representations from attributed networks [40] based on the information aggregation mechanism: $\mathbf{h}_i^{(l+1)} = \text{AGG}(\{\mathbf{h}_i^{(l)}\} \cup \{\mathbf{h}_j^{(l)} : j \in \mathcal{N}_i\})$ where $\mathbf{h}_i^{(l)}$ denotes the representation of node i in the l-th layer, and $\mathbf{h}_i^{(0)}$ is the input attribute of node i. $\mathcal{N}_i$ is the set of all neighbors of node i. $\text{AGG}(\cdot)$ is an aggregation function that can be implemented by mean pooling, max pooling, and many other operations [24]. In this paper, we specify the information aggregation based on the self-attention mechanism in Graph Attention Networks (GAT) [39]. The reason is that GAT is able to account for different neighbors' contributions to the central node via assigning appropriate corresponding importance weights. Thus it is capable of capturing complicated relations in attributed networks. Each GAT layer follows the information propagation scheme of $\mathbf{h}_i^{(l+1)} = \sigma(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)})$, where $\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{e_{ij}}{\sum_{k \in \mathcal{N}_i} e_{ik}}$ and $e_{ij} = \sigma\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]\right)$. Here α_{ij} is the attention weight between node i and j. $\mathbf{W}^{(l)}$ is a learnable parameter matrix for the l-th layer. $\mathbf{W}$ and $\mathbf{a}$ are learnable parameters that are shared by all GAT layers for learning the attention weights. $\|$ denotes the concatenation operation. In practice, we stack multiple GAT layers to form the encoder *Enc* for node representation learning.

Contrasting Between Views. To fully harness the power of human knowledge in $\mathcal{G}_{ano}$, we propose to make a contrast between $\mathcal{G}_{ano}$ and the given attributed network (normal view) $\mathcal{G}$. We expect the anomalous patterns on the attributed network can be well characterized through such contrastive process. Since the augmented anomalous nodes become different from both themselves and their neighbors, we consider two different contrast strategies in this paper, and we name them as Siamese contrast and Triplet contrast. The former contrast strategy is performed by comparing the embedding representation of each abnormal node in the anomalous view and its counterpart in the normal view. The latter contrastive strategy is performed for a connected node pair (i, j) where j is considered abnormal in the anomalous view and i remains intact. It is called "triplet" because three representations, i.e., the representation of i in the normal view and the representations of j in both the normal and anomalous views are involved. These two contrast strategies are described in detail below.

Strategy 1: Siamese Contrast. Suppose *Enc* encodes $\mathcal{G}$ and $\mathcal{G}_{ano}$ through stacked GAT layers into the final representations $\mathbf{Z}$ and $\hat{\mathbf{Z}}$. Siamese contrast is performed

between $\mathbf{z}_i$ and $\hat{\mathbf{z}}_i$, i.e., the representations of each node i in the normal view and the anomalous view. The loss function of Siamese contrast is defined as follows:

$$\mathcal{L}^{\text{sc}} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{I}_{y_i=0}\cdot d\left(\mathbf{z}_i, \hat{\mathbf{z}}_i\right) + \mathrm{I}_{y_i=1}\cdot \max\left\{0, m - d\left(\mathbf{z}_i, \hat{\mathbf{z}}_i\right)\right\}) \tag{1}$$

where I is the indicator function of the condition in its subscript. When applying the Siamese contrastive loss, if $\mathbf{y}_i = 1$, i.e., node i is considered abnormal in G_{ano}, the distance between its representation in the normal and the anomalous view, $d(\mathbf{z}_i, \hat{\mathbf{z}}_i)$ will be maximized with a margin no smaller than m. If $y_i = 0$, i.e., node i is not considered abnormal in $\mathcal{G}_{ano}$, then $d(\mathbf{z}_i, \hat{\mathbf{z}}_i)$ will be minimized.

Strategy 2: Triplet Contrast. In addition to the above strategy, we further propose Triplet contrast that works on a triplet of node representations. Specifically, we consider each connected node pair (i, j) where j is an augmented anomaly in $\mathcal{G}_{ano}$ while i remains intact. The triplet of representations consists of three representations $\mathbf{z}_i$, $\mathbf{z}_j$, and $\hat{\mathbf{z}}_j$, and the loss function is defined as:

$$\mathcal{L}^{\text{tc}} = \sum_{\substack{\forall \mathbf{A}_{ij}=1,\\ y_i=0,\ y_j=1}} \max\left\{0, m - \left(d\left(\mathbf{z}_i, \hat{\mathbf{z}}_j\right) - d\left(\mathbf{z}_i, \mathbf{z}_j\right)\right)\right\}. \tag{2}$$

Through minimizing this loss function, our model will increase the gap between two distances with a margin no smaller than m. Here $d(\mathbf{z}_i, \mathbf{z}_j)$ is the distance between the representations of i and its neighbor j in the normal view, and $d(\mathbf{z}_i, \hat{\mathbf{z}}_j)$ is the distance between the representation of node i in the normal view and that of its neighbor j in the anomalous view. Therefore, CONAD can enforce an augmented anomaly to be far away from its neighbors, and thus the human knowledge regarding this anomaly type can be harnessed.

3.3 Anomaly Detection Module

Besides learning from $\mathcal{G}_{ano}$ which models prior human knowledge of anomaly types, CONAD also needs to learn from the input attributed network $\mathcal{G}$ to detect anomalies in it. Towards this objective, we aim to reconstruct the graph structure and node attributes based on the learned node representations in normal view $\mathbf{Z}$. It has been proved in previous works [7,8,19] that reconstructing structures and attributes helps the model to learn the normal patterns of the input attributed networks, and since anomalies cannot be well reconstructed, they will therefore be detected. Specifically, our model uses a decoder function *Dec* on the encoder output $\mathbf{Z}$. *Dec* consists of a GAT layer to reconstruct the adjacency and attribute matrix from $\mathbf{Z}$. Frobenius norm of the difference between the input and the reconstructed matrix, i.e., reconstruction error, serves as the loss function:

$$\hat{\mathbf{A}} = \sigma\left(\mathbf{Z}\cdot\mathbf{Z}^{\top}\right), \hat{\mathbf{X}} = \text{GATLayer}\left(\mathbf{A},\ \mathbf{Z}\right). \tag{3}$$

$$\mathcal{L}^{\text{recon}} = \lambda\left\|\mathbf{A} - \hat{\mathbf{A}}\right\|_F + (1-\lambda)\cdot\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F. \tag{4}$$

Here, $\sigma(\cdot)$ is a non-linear activation function, e.g., ReLU [27]. $(\cdot)^{\top}$ and $\|\cdot\|_2$ are the transpose and Frobenius norm on matrices. λ is a weighting factor to balance the scales of the two reconstruction errors on the structure and attributes.

3.4 Summary

We summarize the whole process of our proposed model CONAD in this subsection. Our input is an attributed network $\mathcal{G} = \{\mathbf{A}, \mathbf{X}\}$ and prior human knowledge of anomaly types ξ. We first model ξ through a novel data augmentation strategy described in Sect. 3.1. We then have two attributed networks $\mathcal{G}$ and $\mathcal{G}_{ano}$. A Siamese GNN encoder *Enc* is used to learn from prior human knowledge modeled in $\mathcal{G}_{ano}$ by contrasting between node representations in $\mathcal{G}$ and $\mathcal{G}_{ano}$ with the contrastive loss defined in Eq. (1) or Eq. (2). During this process, *Enc* learns to distinguish normal and abnormal representations in the latent space, and thus integrates the prior human knowledge. The node representations of $\mathcal{G}$ are further fed into an anomaly detection module described in Eq. (3) to learn the normal patterns in $\mathcal{G}$ with the reconstruction loss $\mathcal{L}^{\text{recon}}$. Hence CONAD learns from both knowledge ξ and attributed network $\mathcal{G}$ with $\mathcal{L}^{\text{cl}}$ and $\mathcal{L}^{\text{recon}}$, respectively. The total loss of CONAD becomes the summation of the contrastive and reconstruction loss (η is also a weighting factor to balance the two loss terms).

$$\mathcal{L}^{\text{CONAD}} = \eta \cdot \mathcal{L}^{\text{cl}} + (1-\eta)\mathcal{L}^{\text{recon}},\ \mathcal{L}^{\text{cl}} \in \left\{\mathcal{L}^{\text{sc}}, \mathcal{L}^{\text{tc}}\right\}. \tag{5}$$

4 Experiments

4.1 Datasets

Five different real-world datasets, namely, Flickr [15], Amazon [35], Enron [25], Facebook [23], and Twitter [23], are used to evaluate the anomaly detection performance of CONAD. (1) *Flickr* dataset contains user following and follower relations on the eponymous photo-sharing website. There are 7,575 nodes (600 ground truth anomalies) and 23,938 edges in the entire network, and we follow the same settings as [8,10,21] to obtain ground truth anomalies. (2) *Amazon & Enron.* These two datasets contain ground truth anomalies. The Amazon dataset represents co-purchase relations between items. The anomalies here consist of erroneous categories or prices. There are 1,418 nodes (28 ground truth anomalies) and 3,695 edges. Enron is a corporate email network. The anomalies are employees who involve in the accounting fraud in this company. There are 13,533 nodes (5 ground truth anomalies) and 176,987 edges in total. (3) *Facebook & Twitter.* We also use social networks in Facebook and Twitter, where users form relations with others and share their "circles" of friends. We obtain ground truth anomalies by introducing nodes that connect to randomly selected circles or have abnormal attributes like [8]. There are 4,039 nodes (400 ground truth anomalies) and 88,234 edges in the Facebook dataset, and we use 4,865 nodes (500 ground truth anomalies) and 66,772 edges in the Twitter dataset[1].

[1] The anomaly labels in Flickr, Facebook, and Twitter datasets result from manual injection, and the injection rule coincides with two of our data augmentations.

4.2 Experimental Settings

We compare our proposed framework with the following four popular baseline methods, including LOF [2], DOMINANT [8], AEGIS [7], and AnomalyDAE [10]. Among them, the latter three are the state-of-the-art methods that employ GNNs and a comparison with them can validate the superiority of our proposed framework which harnesses the power of human knowledge.

For our proposed framework CONAD, the encoder *Enc* is initialized with two layers of GAT, where the hidden sizes are 128 and 64, respectively. For the reconstruction part, an additional GAT layer is applied for attribute reconstruction, while dot product and sigmoid activation are applied for structure reconstruction. Two attention heads and LeakyReLU [41] activation are used for all GAT layers. The margin m is set to 0.5 for both Siamese and Triplet losses, and the model is denoted by CONAD-S and CONAD-T corresponding to the specific contrastive loss used, i.e., **S**iamese and **T**riplet. Euclidean distance is used as the distance function $d(\cdot, \cdot)$. The ratio of augmented anomalies r is 10% for smaller networks, i.e., Amazon, Flickr, Facebook, and Twitter, and 20% for the larger one, i.e., Enron. The weighting factors λ and η are set to 0.9 and 0.7, respectively. We train the model with Adam [17]. The area under ROC (AUC) serves as the evaluation metric of anomaly detection performance.

Table 1. Anomaly detection performance (AUC scores) comparison. CONAD consistently performs the best across all three datasets (higher is better).

Dataset	Amazon	Enron	Flickr	Facebook	Twitter
LOF	0.510	0.581	0.661	0.522	0.511
DOMINANT	0.592	0.716	0.749	0.554	0.571
AEGIS	0.556	0.602	0.765	0.659	0.645
AnomalyDAE	0.610	0.552	0.694	0.741	0.688
CONAD-S	**0.635**	**0.731**	**0.782**	0.612	0.670
CONAD-T	0.620	**0.731**	0.759	**0.863**	**0.742**

4.3 Anomaly Detection Performance Comparison

Table 1 shows the anomaly detection performance of CONAD and baselines, where CONAD outperforms all others in all of the five real-world datasets used. Specifically, GNN-based models generally perform better than LOF, which does not consider structure information. By modeling and integrating prior human knowledge, CONAD achieves better performance than the other three GNN-based unsupervised anomaly detection models. Besides, for networks with explicit communities, i.e., Facebook and Twitter, CONAD-T, which contrasts between each pair of neighbors, performs better than CONAD-S, which only contrasts between the representations of each individual node in the normal and anomalous views.

4.4 Ablation Study

In this subsection, we conduct further experiments to study the improvements brought by each module individually in the proposed framework CONAD on Amazon dataset. The results are shown in Table 2, and similar observations can also be found in other datasets. We first study the influence of the types of contrasting between views, i.e., Siamese contrast and Triplet contrast. The performance of CONAD-T with Triplet contrast is slightly worse than CONAD-S with Siamese contrast. However, the performance on Facebook and Twitter datasets shown in the previous subsection demonstrates the opposite. We speculate that it is because the co-purchase relation in Amazon datasets does not have explicit communities, contrary to the friendship relation in the two social networks. Therefore, contrasting between neighbors is not very helpful. We then study how the amount of prior human knowledge modeled affects the performance of CONAD. Towards this goal, we change the data augmentation strategy in 3.1, where we solely model human knowledge of structure (w/o attribute anomalies) or attribute (w/o structure anomalies) anomalies. The performance of CONAD decrease with either of these two types removed, showing that the more knowledge of anomaly types is given, the more CONAD can harness it to facilitate anomaly detection. We also investigate the effectiveness of the knowledge integration module. Concretely, we remove this module which contrasts between normal and anomalous views entirely. The resulting model becomes almost identical to DOMINANT, and the corresponding performance drops drastically, which demonstrates that integrating prior human knowledge is crucial in the superior performance of CONAD. Lastly, we study the influence of the reconstruction. We remove the decoder used to reconstruct the structure and attributes of the input attributed network, and apply LOF instead to the nodes representations learned by the encoder. The performance shows that LOF fails to detect anomalies from only those node representations. It proves that the decoder and reconstruction also contribute a lot to anomaly detection.

Table 2. Ablation study on the Amazon dataset.

Variants of CONAD	AUC score
CONAD-S	**0.635**
CONAD-T	0.620
w/o attribute anomalies	0.621
w/o structure anomalies	0.628
w/o contrasting between views	0.592
w/o reconstruction	0.510

4.5 Robustness of CONAD W.r.t. Different Ratios of Anomalies

At last, we study the robustness of CONAD on Flickr dataset where the ground truth anomalies can be easily tuned. We omit the results on other datasets due to the observation of similar patterns. We vary the ratios of ground truth anomalies in Flickr among 2.5%, 5%, 7.5%, and 10% of the total number of nodes, and find that CONAD maintains steady performances with AUC scores of 0.760, 0.772, 0.781, and 0.778. It demonstrates that CONAD is very robust in detecting anomalies in attributed networks when the ratio of anomalies present varies.

5 Related Works

5.1 Attributed Network Anomaly Detection

Attributed networks are a kind of graph structured data that exist ubiquitously in many real-world scenarios. Detecting anomalies in attributed networks is of vital importance for anti-fraud, anti-money laundering, and other safety-critical applications [1,22]. Therefore, attributed network anomaly detection has attracted an increasingly amount of research attentions in recent years. Existing approaches can be broadly categorized as traditional machine learning (Non-DL) methods and deep learning (DL) methods. Non-DL methods are often developed based on certain heuristic anomaly metrics, e.g., ConSub [35], FocusCO [31], and AMEN [30], or matrix decomposition techniques, e.g., Radar [19], ANOMALOUS [29], and ALAD [20]. More recently, many DL anomaly detection methods have been proposed, which often resort to GNNs due to their superior representation learning capability. Typical methods along this line include DOMINANT [8], AEGIS [7], and AnomalyDAE [10]. Our proposed CONAD differs from the methods introduced above as the above methods are mainly unsupervised while ours explicitly models the human knowledge of different anomaly types on attributed networks and tightly incorporate such knowledge into the detection model.

5.2 Contrastive Learning

Supervised learning achieves great success in numerous machine learning areas, but one major disadvantage of it is that a large amount of labeled data is required to train a descent model. To ease the reliance on labeled data, contrastive learning (CL) has gained popularity as a novel self-supervised learning (SSL) paradigm. It often utilizes data augmentation techniques to obtain different views of the data, and leverages InfoMax principle [13] to maximize the similarity between pairs of positive views while minimize pairs of negative views. With contrastive learning, SSL models [5,16,37] achieve comparable performance in image classification against their supervised counterparts. CL frameworks also enjoys successes in graph representation learning [12,32,43] where techniques designed specifically for graph structured data, such as random walk and graph diffusion, can be used to generative positive views.

6 Conclusions

In this paper, we propose CONAD, a contrastive learning framework capable of leveraging human knowledge to detect anomalies on attributed networks. Specifically, we first model human knowledge of real-world anomalies through a data augmentation approach. We then train a Siamese graph neural network with a contrastive loss to encode both the modeled knowledge and the original attributed networks. Finally, we use reconstruction loss to obtain anomaly scores. Experiments on several datasets with different nature and characteristics

show detection performance improvements compared to state-of-the-art models. Furthermore, we analyze the benefit brought about by each part in CONAD and show its robustness w.r.t. different anomaly ratios on the attributed network.

Acknowledgements. Yushun Dong and Jundong Li are partially supported by the National Science Foundation (NSF) under grants #2006844.

References

1. Akoglu, L., Tong, H., Koutra, D.: Graph based anomaly detection and description: a survey. Data Min. Knowl. Disc. **29**(3), 626–688 (2014). https://doi.org/10.1007/s10618-014-0365-y
2. Breunig, M.M., Kriegel, H., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: SIGMOD (2000)
3. Cao, S., Yang, X., Chen, C., Zhou, J., Li, X., Qi, Y.: Titant: online real-time transaction fraud detection in ant financial. Proc. VLDB Endow. **12**(12), 2082–2093 (2019)
4. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 15:1–15:58 (2009)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. In: ICML (2020)
6. Deng, C., Ji, X., Rainey, C., Zhang, J., Lu, W.: Integrating machine learning with human knowledge. iScience **23**(11), 101656 (2020)
7. Ding, K., Li, J., Agarwal, N., Liu, H.: Inductive anomaly detection on attributed networks. In: IJCAI (2020)
8. Ding, K., Li, J., Bhanushali, R., Liu, H.: Deep anomaly detection on attributed networks. In: SDM (2019)
9. Ding, K., Li, J., Liu, H.: Interactive anomaly detection on attributed networks. In: WSDM (2019)
10. Fan, H., Zhang, F., Li, Z.: Anomalydae: dual autoencoder for anomaly detection on attributed networks. In: ICASSP (2020)
11. Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., Ghosh, R.: Exploiting burstiness in reviews for review spammer detection. In: ICWSM (2013)
12. Hassani, K., Ahmadi, A.H.K.: Contrastive multi-view representation learning on graphs. In: ICML (2020)
13. Hjelm, R.D., et al.: Learning deep representations by mutual information estimation and maximization. In: ICLR (2019)
14. Hu, X., Tang, J., Zhang, Y., Liu, H.: Social spammer detection in microblogging. In: IJCAI (2013)
15. Huang, X., Li, J., Hu, X.: Label informed attributed network embedding. In: WSDM 2017
16. Khosla, P., et al.: Supervised contrastive learning (2020)
17. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2015)
18. Ladickỳ, L., Jeong, S., Solenthaler, B., Pollefeys, M., Gross, M.: Data-driven fluid simulations using regression forests. ACM Trans. Graph. (TOG) **34**(6), 1–9 (2015)
19. Li, J., Dani, H., Hu, X., Liu, H.: Radar: residual analysis for anomaly detection in attributed networks. In: IJCAI (2017)
20. Liu, N., Huang, X., Hu, X.: Accelerated local anomaly detection via resolving attributed networks. In: IJCAI (2017)

21. Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., Karypis, G.: Anomaly detection on attributed networks via contrastive self-supervised learning. CoRR abs/2103.00113 (2021)
22. Ma, X., et al.: A comprehensive survey on graph anomaly detection with deep learning (2021)
23. McAuley, J.J., Leskovec, J.: Learning to discover social circles in ego networks. In: NeurIPS (2012)
24. Mesquita, D.P.P., Jr., A.H.S., Kaski, S.: Rethinking pooling in graph neural networks. In: NeurIPS (2020)
25. Müller, E., Sánchez, P.I., Mülle, Y., Böhm, K.: Ranking outlier nodes in subspaces of attributed graphs. In: ICDE Workshop (2013)
26. Muralidhar, N., Islam, M.R., Marwah, M., Karpatne, A., Ramakrishnan, N.: Incorporating prior domain knowledge into deep neural networks. In: IEEE Big Data (2018)
27. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML (2010)
28. Pang, G., Shen, C., Cao, L., van den Hengel, A.: Deep learning for anomaly detection: a review. CoRR abs/2007.02500 (2020)
29. Peng, Z., Luo, M., Li, J., Liu, H., Zheng, Q.: Anomalous: a joint modeling approach for anomaly detection on attributed networks. In: IJCAI (2018)
30. Perozzi, B., Akoglu, L.: Scalable anomaly ranking of attributed neighborhoods. In: SDM (2016)
31. Perozzi, B., Akoglu, L., Sánchez, P.I., Müller, E.: Focused clustering and outlier detection in large attributed graphs. In: KDD (2014)
32. Qiu, J., et al.: GCC: graph contrastive coding for graph neural network pre-training. In: KDD (2020)
33. von Rueden, L., et al.: Informed machine learning-a taxonomy and survey of integrating knowledge into learning systems. arXiv preprint arXiv:1903.12394 (2019)
34. Sánchez, P.I., Müller, E., Irmler, O., Böhm, K.: Local context selection for outlier ranking in graphs with multiple numeric node attributes. In: SSDBM (2014)
35. Sánchez, P.I., Müller, E., Laforet, F., Keller, F., Böhm, K.: Statistical selection of congruent subspaces for mining attributed graphs. In: ICDM (2013)
36. Stewart, R., Ermon, S.: Label-free supervision of neural networks with physics and domain knowledge. In: AAAI (2017)
37. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: ECCV (2020)
38. Varol, O., Ferrara, E., Davis, C., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization. In: ICWSM (2017)
39. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: ICLR (2018)
40. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. IEEE Trans. Neural Networks Learn. Syst. **32**(1), 4–24 (2021)
41. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. CoRR abs/1505.00853 (2015)
42. Xu, J.G., Zhao, Y., Chen, J., Han, C.: A structure learning algorithm for bayesian network using prior knowledge. J. Comput. Sci. Technol. **30**(4), 713–724 (2015)
43. You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y.: Graph contrastive learning with augmentations. In: NeurIPS (2020)
44. Zhu, M., Zhu, H.: Mixedad: a scalable algorithm for detecting mixed anomalies in attributed graphs. In: AAAI (2020)