Deep Anomaly Detection on Attributed Networks

Kaize Ding[*] Jundong Li[*] Rohit Bhanushali[*] Huan Liu[*]

Abstract

Attributed networks are ubiquitous and form a critical component of modern information infrastructure, where additional node attributes complement the raw network structure in knowledge discovery. Recently, detecting anomalous nodes on attributed networks has attracted an increasing amount of research attention, with broad applications in various high-impact domains, such as cybersecurity, finance, and healthcare. Most of the existing attempts, however, tackle the problem with shallow learning mechanisms by ego-network or community analysis, or through subspace selection. Undoubtedly, these models cannot fully address the computational challenges on attributed networks. For example, they often suffer from the network sparsity and data nonlinearity issues, and fail to capture the complex interactions between different information modalities, thus negatively impact the performance of anomaly detection. To tackle the aforementioned problems, in this paper, we study the anomaly detection problem on attributed networks by developing a novel deep model. In particular, our proposed deep model: (1) explicitly models the topological structure and nodal attributes seamlessly for node embedding learning with the prevalent graph convolutional network (GCN); and (2) is customized to address the anomaly detection problem by virtue of deep autoencoder that leverages the learned embeddings to reconstruct the original data. The synergy between GCN and autoencoder enables us to spot anomalies by measuring the reconstruction errors of nodes from both the structure and the attribute perspectives. Extensive experiments on real-world attributed network datasets demonstrate the efficacy of our proposed algorithm.

Keywords: Anomaly Detection; Attributed Networks; Graph Convolutional Network; Deep Autoencoder

1 Introduction

Attributed networks provide a potent tool to handle the data heterogeneity that we are often confronted with in vast amounts of information systems. Apart from traditional plain networks in which only node-to-node interactions are observed, attributed networks also encode a rich set of features for each node [2, 13, 18]. They are increasingly used to model a wide range of complex systems, such as social media networks, critical infrastructure networks, and gene regulatory networks [2, 26]. For example, in social networks, users not only are connected with each other by performing various social activities but also are affiliated with rich profile information; in critical infrastructure networks, different power stations form grids, and are also associated with additional attribute information (e.g., electricity capacity); in gene regulatory networks, genes interact with each other to control specific cell functions in addition to the rich gene sequence expressions. Studies from social science have shown that data often exhibits correlation among the attributes of connected individuals [20, 29], and such insights are helpful in distilling actionable knowledge from such networks.

Detecting anomalies from data (e.g., attribute-value data, networked data) is a vital research problem of pressing societal concerns, with significant implications in many security-related applications, ranging from social spam detection, financial fraud detection to network intrusion detection [1]. Due to the strong modeling power of attributed networks in unifying information of different modalities, there is a surge of research interests in detecting anomalous nodes whose patterns deviate significantly from other majority nodes on attributed networks. Generally, the abnormality of nodes on attributed networks is not only determined by their mutual interactions with others (w.r.t.*topological structure*), but also is measured by their content dissonance (w.r.t. *nodal attributes*).

Due to the prohibitive cost for accessing the ground truth anomalies, existing efforts are mostly unsupervised. Among them, one family of methods study the problem at the *mesoscope* with ego-network [24] or community analysis [10] and then identify anomalies by measuring the abnormality of ego-networks or comparing the current node with other nodes within the same community. Another family of methods heavily rely on subspace selection and attempt to find anomalies in a node feature subspace [28, 27, 21, 25]. Recently, residual analysis has emerged as another way to find anomalous nodes [17, 23], where anomalies are defined as the nodes that cannot be approximated from others. Despite their empirical success, the following challenges remain for anomaly detection on attributed networks: (1) *Network sparsity* - the network structure could be very sparse on real-world attributed networks; thus ego-network or community analysis is difficult to perform as they highly depend on the observed node interactions. (2) *Data nonlinearity* - the node interactions and nodal attributes are highly non-linear in nature while existing subspace selection based anomaly detectors mainly

[*] Computer Science and Engineering, Arizona State University, Tempe, AZ, USA. {kding9, jundongl, rbhanush, huan.liu}@asu.edu

model the attributed networks with linear mechanisms. (3) *Complex modality interactions* - attributed networks are notoriously difficult to tackle due to the bewildering combination of two information sources, which necessitates a unified feature space to capture their complex interactions for anomaly detection.

To address the challenges above, we propose to model the attributed networks with graph convolutional network (GCN) [16]. GCN, which takes the topological structure and nodal attributes as input, is able to learn discriminative node embeddings by stacking multiple layers of linear units and non-linear activation functions. Even though GCN emerges to be a principled tool to model attributed networks and achieves the state-of-the-art performance in the semi-supervised node classification task, it remains unclear how its power can be shifted to the anomaly detection problem. To bridge the gap, we propose a novel graph convolutional autoencoder framework called DOMINANT (Deep Anomaly Detection on Attributed Networks) to support anomaly detection on attributed networks. Specifically, DOMINANT first compresses the input attributed network to succinct low-dimensional embedding representations using graph convolutional network as an encoder function; then we aim to reconstruct both the topological structure and nodal attributes with corresponding decoder functions. The reconstruction errors of nodes following the encoder and decoder phases are then leveraged for spotting anomalous nodes on attributed networks. The main contributions of this paper are as follow:

- We systematically analyze the limitations of existing shallow anomaly detection methods and show the significance of developing a novel deep architectured anomaly detector on attributed networks.
- We develop a principled graph convolutional autoencoder DOMINANT which seamlessly models the attributed network and conducts anomaly detection in a joint framework. In particular, the proposed model can spot anomalies by analyzing the reconstruction errors of nodes from both the structure and the attribute perspectives.
- We evaluate our proposed model on various attributed networks from different domains. Empirical experimental results demonstrate the superior performance of our proposed framework.

The remaining of the paper is organized as follows. We formally introduce the problem definition in Section 2. In Section 3, we present the details of the proposed deep anomaly detection model. Experimental evaluations on multiple real-world datasets are shown in Section 4. Section 5 reviews the related work and Section 6 concludes the whole paper.

2 Problem Definition

Following the commonly used notations, in this paper, we use calligraphic fonts to denote sets (e.g., $\mathcal{V}$), bold lowercase letters (e.g., $\mathbf{x}$) to denote vectors and bold uppercase letters for matrices (e.g., $\mathbf{X}$). The i^{th} row of a matrix $\mathbf{X}$ is denoted by $\mathbf{x}_i$ and the $(i,j)^{th}$ element of matrix $\mathbf{X}$ is denoted by $\mathbf{X}_{i,j}$. Besides, we represent the identity matrix as $\mathbf{I}$, and the transpose of a matrix $\mathbf{X}$ is represented as $\mathbf{X}^{\mathrm{T}}$. The ℓ_2-norm of a vector is denoted by $||\cdot||_2$. The Frobenius norm of a matrix is represented by $||\cdot||_F$. Accordingly, we define the attributed network as follows:

DEFINITION 1. ***Attributed Networks**: An attributed network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ consists of: (1) the set of nodes $\mathcal{V} = \{v_1, v_2, ..., v_n\}$, where $|\mathcal{V}| = n$; (2) the set of edges $\mathcal{E}$, where $|\mathcal{E}| = m$; and (3) the node attributes $\mathbf{X} \in \mathbb{R}^{n \times d}$, where the i^{th} row vector $\mathbf{x}_i \in \mathbb{R}^d$ $(i = 1, ..., n)$ is the attribute[1] information for the i^{th} node.*

The topological structure of attributed network $\mathcal{G}$ can be represented by an adjacency matrix $\mathbf{A}$, where $\mathbf{A}_{i,j} = 1$ if there is a link between node v_i and node v_j. Otherwise, $\mathbf{A}_{i,j} = 0$. We follow the setting of [17] to obtain the adjacency matrix $\mathbf{A} = max(\mathbf{A}, \mathbf{A}^{\mathrm{T}})$ for directed networks. To make the results more interpretable, we formulate the task of anomaly detection on attributed networks as a ranking problem:

PROBLEM 1. ***Anomaly Ranking on Attributed Networks**: Given an attributed network $\mathcal{G}$, with the adjacency matrix $\mathbf{A}$ and attribute information matrix $\mathbf{X}$ of n node instances, the task is to rank all the nodes according to the degree of abnormality, such that the nodes that differ singularly from the majority reference nodes should be ranked on high positions.*

Next, we will introduce our proposed deep framework which models network topological structure and nodal attributes coherently for detecting anomalies on attributed networks in details.

3 The Proposed Model

In this section, we present the proposed framework of DOMINANT in details. The architecture of the deep model is illustrated in Figure 1. As can be observed, the fundamental building block of DOMINANT is the deep autoencoder [11] and it consists of three essential components: (i) *attributed network encoder* - which models network structure and nodal attributes seamlessly in a joint framework for node embedding representation learning with GCN; (ii) *structure reconstruction*

[1] In this paper, we use attribute and feature interchangeably.

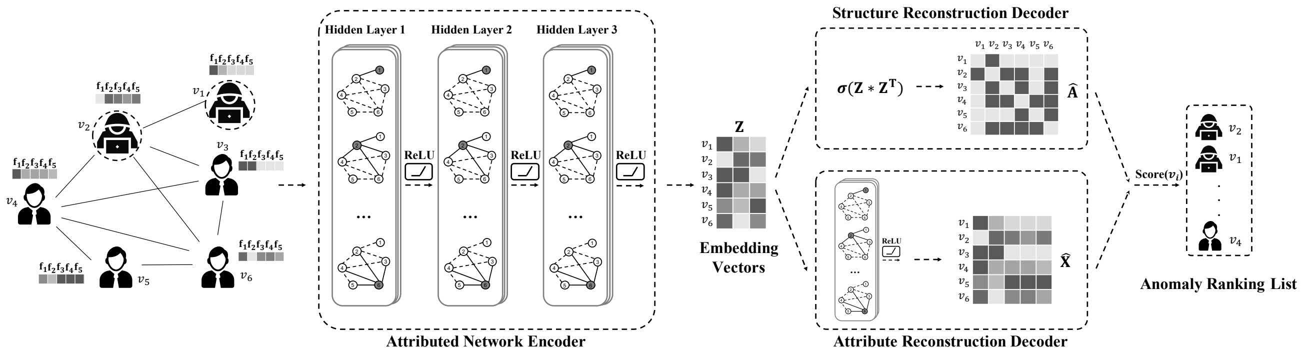


Figure 1: The overall framework of our proposed Dominant for deep anomaly detection on attributed networks.

decoder - which aims to reconstruct the original network topology with the learned node embeddings; and (iii) *attribute reconstruction decoder* - which attempts to reconstruct the observed nodal attributes with the obtained node embeddings. Afterwards, the reconstruction errors of nodes are then leveraged to flag anomalies on attributed networks.

3.1 Preliminary - Deep Autoencoder As suggested by [32, 37, 17], the disparity between the original data and the estimated data (i.e., reconstruction errors) is a strong indicator to show the abnormality of instances in a dataset. Specifically, the data instances with large reconstruction errors are more likely to be considered as anomalies, since their patterns deviate significantly from the majority and cannot be accurately reconstructed from the observed data. Among various reconstruction based anomaly detection methods, deep autoencoder achieves state-of-the-art performance. Deep autoencoder is a type of deep neural network that is used to learn latent representations of data in an unsupervised manner by stacking multiple layers of encoding and decoding functions together. It has achieved promising learning performance in various domains, such as computer vision, natural language processing, and speech recognition [11].

Given an input dataset $\mathbf{X}$, the encoder $Enc(\cdot)$ is first applied to map the data into a latent low-dimensional feature space, and then the decoder $Dec(\cdot)$ tries to recover the original data based on the latent representations. The learning process can be described as minimizing a cost function described below:

$$\min \mathbb{E}[dist(\mathbf{X}, Dec(Enc(\mathbf{X}))], \tag{3.1}$$

where $dist(\cdot,\cdot)$ is a predefined distance metric. In practice, we often choose the ℓ_2-norm distance to measure the reconstruction errors. It also should be noted that deep autoencoder is able to capture the highly nonlinear information from high-dimensional input by applying multiple layers of linear units and nonlinear activation functions in the encoder and decoder phases, which is advantageous compared to conventional shallow learning models. Subsequently, in this study, we propose to solve the problem of anomaly detection on attributed networks in a deep autoencoder architecture.

3.2 Attributed Network Encoder As a rich network representation, attributed networks encode not only the network structure but also abundant nodal attributes. However, conventional deep autoencoders can only handle *i.i.d.* attribute-value data [37, 35], which cannot be directly applied to our scenario. How to design an effective encoder to capture the underlying properties of attributed networks remains a daunting task as we need to address the three challenges (i.e., *network sparsity*, *data nonlinearity*, and *complex modality interactions*) simultaneously. To this end, we propose a new type of attributed network encoder inspired by the graph convolutional network (GCN) model [16]. Specifically, GCN considers the high-order node proximity when learning the embedding representations, thus it mitigates the network sparsity issue beyond the observed links among nodes. Meanwhile, through multiple layers of nonlinear transformations, it captures the nonlinearity of data and the complex interactions of two information modalities on attributed networks.

Mathematically, GCN extends the operation of convolution to networked data in the spectral domain and learns a layer-wise new latent representation by a spectral convolution function:

$$\mathbf{H}^{(l+1)} = f(\mathbf{H}^{(l)}, \mathbf{A}|\mathbf{W}^{(l)}), \tag{3.2}$$

where $\mathbf{H}^{(l)}$ is the input for the convolution layer l, and $\mathbf{H}^{(l+1)}$ is the output after the convolution layer. We take the attribute matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ as the input of first layer, which is equivalent to $\mathbf{H}^{(0)}$. $\mathbf{W}^{(l)}$ is a layer-specific trainable weight matrix we need to learn in the neural network. Each layer of the graph convolutional network can be expressed with the function $f(\mathbf{H}^{(l)}, \mathbf{A}|\mathbf{W}^{(l)})$ as follows:

$$f(\mathbf{H}^{(l)}, \mathbf{A}|\mathbf{W}^{(l)}) = \sigma(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}), \tag{3.3}$$

where $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ and $\widetilde{\mathbf{D}}$ is the diagonal matrix of $\widetilde{\mathbf{A}}$ with the diagonal element as $\widetilde{\mathbf{D}}_{i,i} = \sum_j \widetilde{\mathbf{A}}_{i,j}$. Thus we can directly compute $\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}$ as a preprocessing step. Note that $\sigma(\cdot)$ is a non-linear activation function, such as $Relu(x) = max(0, x)$. It is worth noting that the filter or feature map parameters $\mathbf{W}^l$ are shared for all nodes on the attributed network. Given the attribute matrix $\mathbf{X}$ as input, the k^{th}-hop neighborhood of each node can be effectively captured by successively stacking a number of k convolutional layers. Therefore, the embeddings $\mathbf{Z}$ not only encode the attribute information of each node but also involve the k^{th}-order node proximity information. In this work, we propose to use three convolutional layers for constructing the attributed network encoder, but it should be noted that more layers can also be stacked for building a deeper network. The attributed network encoder can be formulated as:

$$\mathbf{H}^{(1)} = f_{Relu}(\mathbf{X}, \mathbf{A}|\mathbf{W}^{(0)}) \tag{3.4}$$

$$\mathbf{H}^{(2)} = f_{Relu}(\mathbf{H}^{(1)}, \mathbf{A}|\mathbf{W}^{(1)}) \tag{3.5}$$

$$\mathbf{Z} = \mathbf{H}^{(3)} = f_{Relu}(\mathbf{H}^{(2)}, \mathbf{A}|\mathbf{W}^{(2)}). \tag{3.6}$$

Here, $\mathbf{W}^{(0)} \in \mathbb{R}^{n \times h_1}$ is an input-to-hidden layer with h_1 feature maps. Similarly, $\mathbf{W}^{(1)} \in \mathbb{R}^{h_1 \times h_2}$ and $\mathbf{W}^{(3)} \in \mathbb{R}^{h_2 \times h_3}$ are two hidden-to-hidden weight matrices. After applying three layers of convolution, the input attributed network can be transferred to the h_3-dimensional latent representations $\mathbf{Z}$, which can capture the high non-linearity in the topological network structure and nodal attributes.

3.3 Structure Reconstruction Decoder In this subsection, we will discuss how to reconstruct the original network structure with the learned latent representations $\mathbf{Z}$, which is from the aforementioned encoder module. Let $\widehat{\mathbf{A}}$ denote the estimated adjacency matrix, then the structure reconstruction error $\mathbf{R}_S = \mathbf{A} - \widehat{\mathbf{A}}$ can be exploited to determine structural anomalies on the network. Specifically, for a certain node, if its structure information can be approximated through the structure reconstruction decoder, thus it is of low probability to be anomalous. On the opposite side, if the connectivity patterns cannot be well reconstructed, it implies that its structure information does not conform to the patterns of majority normal nodes. Therefore, a larger norm of $\mathbf{R}_S(i,:)$ indicates that the i^{th} node on the attributed network has a higher probability of being an anomaly from the network structure aspect. Specifically, the decoder takes the latent representations as input and predicts whether there is a link between each pair of two nodes:

$$p(\widehat{\mathbf{A}}_{i,j} = 1|\mathbf{z}_i, \mathbf{z}_j) = sigmoid(\mathbf{z}_i, \mathbf{z}_j^{\mathrm{T}}). \tag{3.7}$$

Accordingly, we train a link prediction layer based on the output of attributed network encoder $\mathbf{Z}$, which can be presented as follows:

$$\widehat{\mathbf{A}} = sigmoid(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}). \tag{3.8}$$

3.4 Attribute Reconstruction Decoder Similarly, to compute the reconstruction errors of nodal attributes, we propose an attribute reconstruction decoder that approximates the nodal attributes information from the encoded latent representations $\mathbf{Z}$. Specifically, the attribute reconstruction decoder leverages another graph convolutional layer to predict the original nodal attributes as follows:

$$\widehat{\mathbf{X}} = f_{Relu}(\mathbf{Z}, \mathbf{A}|\mathbf{W}^{(3)}). \tag{3.9}$$

With the computed reconstruction errors $\mathbf{R}_A = \mathbf{X} - \widehat{\mathbf{X}}$, we can spot anomalies on the attributed networks from the attribute perspective.

3.5 Anomaly Detection Until now, we have introduced how to reconstruct the topological network structure, and nodal attributes using structure reconstruction decoder and attribute reconstruction decoder, respectively. To jointly learn the reconstruction errors, the objective function of our proposed deep graph convolutional autoencoder can be formulated as:

$$\begin{aligned}\mathcal{L} &= (1-\alpha)\mathbf{R}_S + \alpha\mathbf{R}_A \\ &= (1-\alpha)||\mathbf{A} - \widehat{\mathbf{A}}||_F^2, +\alpha||\mathbf{X} - \widehat{\mathbf{X}}||_F^2,\end{aligned} \tag{3.10}$$

where α is an important controlling parameter which balances the impacts of structure reconstruction and attribute reconstruction.

By minimizing the above objective function, our proposed deep graph convolutional autoencoder can iteratively approximate the input attributed network based on the encoded latent representations until the objective function converges. The final reconstruction errors are then employed to assess the abnormality of nodes. Note that the weight matrices of the deep graph convolutional autoencoder are trained using gradient descent on the objective function. After a certain number of iterations, we can compute the anomaly score of each node $\mathbf{v}_i$ according to:

$$score(\mathbf{v}_i) = (1-\alpha)||\mathbf{a} - \widehat{\mathbf{a}}_i||_2 + \alpha||\mathbf{x}_i - \widehat{\mathbf{x}}_i||_2. \tag{3.11}$$

	BlogCatalog	Flickr	ACM
# nodes	5,196	7,575	16,484
# edges	171,743	239,738	71,980
# attributes	8,189	12,047	8,337
# anomalies	300	450	600

Table 1: Details of the three datasets

Specifically, instances with larger scores are more likely to be considered as anomalies; thus we can compute the ranking of anomalies according to the corresponding anomaly scores.

3.6 Complexity Analysis Graph convolutional network is a computationally efficient model whose complexity is linear to the number of edges on the network. For a particular layer, the convolution operation is $\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}\mathbf{W}$, and its complexity is $\mathcal{O}(mdh)$ [16] as $\widetilde{\mathbf{A}}\mathbf{X}$ can be efficiently implemented using sparse-dense matrix multiplications, where m is the number of non-zero elements in matrix $\mathbf{A}$ and d is the number feature dimensions on the attributed network, and h is the number of feature maps of the weight matrix. In addition to the convolutional layers, there is another link prediction layer in our model to reconstruct the original topological structure; thus the overall complexity is $\mathcal{O}(mdH + n^2)$, where H is the summation of all feature maps across different layers.

4 Experiments

In this section, we perform empirical evaluations on real-world attributed networks to verify the effectiveness of the proposed Dominant framework.

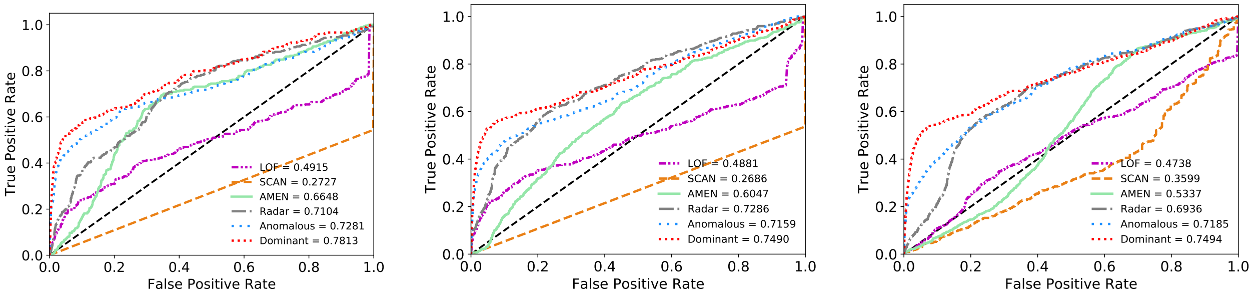
4.1 Datasets In order to have a comprehensive evaluation, we adopt three real-world attributed network datasets that have been widely used in previous research [19, 14, 8] in our experiments:

- **BlogCatalog**: BlogCatalog is a blog sharing website. The bloggers in blogcatalog can follow each other forming a social network. Users are associated with a list of tags to describe themselves and their blogs, which are regarded as node attributes.
- **Flickr**: Flickr is an image hosting and sharing website. Similar to BlogCatalog, users can follow each other and form a social network. Node attributes of users are defined by their specified tags that reflect their interests.
- **ACM**: ACM is another attributed network from academic field. It is a citation network where each paper is regarded as a node on the network, and the links are the citation relations among different papers. The attributes of each paper are generated from the paper abstract.

As there is no ground truth of anomalies in the above datasets, thus we need to inject anomalies into the attributed networks for our empirical evaluation. In particular, we refer to two anomaly injection methods that has been used in previous research [8, 31] to generate a combined set of anomalies for each dataset by perturbing topological structure and nodal attributes, respectively. On one hand, to perturb the topological structure of an attributed network, we adopt the method introduced by [8] to generate some small cliques. The intuition behind this method is that in many real-world scenarios, small clique is a typical anomalous substructure in which a small set of nodes are much more closely linked to each other than average [30]. Therefore, after we specify the clique size as m, we randomly select m nodes from the network and then make those nodes fully connected, and then all the m nodes in the clique are regarded as anomalies. We iteratively repeat this process until a number of n cliques are generated and the total number of structral anomalies is $m \times n$. In our experiments, we fix the clique size m to 15 and set n to 10, 15 and 20 for BlogCatalog, Flickr and ACM, respectively. In addition to the injection of structural anomalies, we adopt another attribute perturbation schema introduced by [31] to generate anomalies from attribute perspective. To guarantee an equal number of anomalies from structural perspective and attribute perspective will be injected into the attributed network, we first randomly select another $m \times n$ nodes as the attribute perturbation candidates. For each selected node i, we randomly pick another k nodes from the data and select the node j whose attributes deviate the most from node i among the k nodes by maximizing the Euclidean distance $||x_i - x_j||_2$. Afterwards, we then change the attributes x_i of node i to x_j. In our experiments, we set the value of k to 50. The details of these three attributed network datasets are shown in Table 1.

4.2 Experimental Settings In this section, we introduce the detailed experimental settings, including the compared baseline methods and evaluation metrics.
Compared Methods. We compare the proposed Dominant framework with the following popular anomaly detection methods:

- **LOF** [4] detects anomalies at the contextual level and only considers nodal attributes.
- **SCAN** [34] is a structure based detection method which detects anomalies at the structural level.



(a) BlogCatalog (b) Flickr (c) ACM

Figure 2: ROC curves and AUC scores of all methods on different datasets.

- **AMEN** [24] uses both attribute and network structure information to detect anomalous neighborhoods. Specifically, it analyzes the abnormality of each node from the ego-network point of view.

- **Radar** [17] is the state-of-the-art unsupervised anomaly detection framework for attributed networks. It detects anomalies whose behaviors are singularly different from the majority by characterizing the residuals of attribute information and its coherence with network information.

- **ANOMALOUS** [23] performs joint anomaly detection and attribute selection to detect anomalies on attributed networks based on the CUR decomposition and residual analysis.

Evaluation Metrics In the experiments, two evaluation metrics are used to measure the performance of different anomaly detection algorithms:

- **ROC-AUC**: As a widely used evaluation metric in previous anomaly detection methods [17, 23], the ROC curve is a plot of true positive rate (an anomaly is recognized as an anomaly) against false positive rate (a normal node is recognized as an anomaly) according to the ground truth and the detection results. AUC value is the area under the ROC curve, representing the probability that a randomly chosen abnormal node is ranked higher than a normal node. If the AUC value approaches 1, the method is of high quality.

- **Precision@**K: As each anomaly detection method outputs a ranking list according to the anomalous scores of different nodes, we use Precision@K to measure the proportion of true anomalies that a specific detection method discovered in its top K ranked nodes.

- **Recall@**K: This metric measures the proportion of true anomalies that a specific detection method discovered in the total number of ground truth anomalies.

Parameter Settings In the experiments on different datasets, we propose to optimize the loss function with Adam [15] algorithm and train the proposed model for 300 epochs for the performance evaluation. We set the learning rate to 0.005. In addition, the attributed network encoder is built with three convolutional layers (64-neuron, 32-neuron and 16-neuron, respectively). For the other baselines, we retain to the settings described in the corresponding papers.

4.3 Experimental Results In the experiments, we evaluate the performance of our proposed model Dominant by comparing it with the aforementioned baselines. We first present the experimental results in terms of ROC-AUC on the three datasets in Figure 2. Then we present the results w.r.t. Precision@K and Recall@K for other methods on all the attributed networks in Table 2. Note that we do not include the results of SCAN and AMEN in Table 2 as they are clustering based methods that cannot provide a precise ranking list for all the nodes. From the evaluation results, we make the following observations:

- The proposed deep model Dominant outperforms other baseline methods on all the three attributed networks. It verifies the effectiveness of performing anomaly detection on attributed networks by deep architecture.

- LOF and SCAN cannot achieve satisfying results in our experiments as they merely consider the nodal attributes or topological structure. Even though AMEN is designed for anomaly detection on attributed networks, it centers around finding

Precision@K												
	BlogCatalog				Flickr				ACM			
K	50	100	200	300	50	100	200	300	50	100	200	300
LOF	0.300	0.220	0.180	0.183	0.420	0.380	0.270	0.237	0.060	0.060	0.045	0.037
Radar	0.660	0.670	0.550	0.416	0.740	0.700	0.635	0.503	0.560	0.580	0.520	0.430
Anomalous	0.640	0.650	0.515	0.417	**0.790**	0.710	0.650	0.510	0.600	0.570	0.510	0.410
Dominant	**0.760**	**0.710**	**0.590**	**0.470**	0.770	**0.730**	**0.685**	**0.593**	**0.620**	**0.590**	**0.540**	**0.497**

Recall@K												
	BlogCatalog				Flickr				ACM			
K	50	100	200	300	50	100	200	300	50	100	200	300
LOF	0.050	0.073	0.120	0.183	0.047	0.084	0.120	0.158	0.005	0.010	0.015	0.018
Radar	0.110	0.223	0.367	0.416	0.082	0.156	0.282	0.336	0.047	0.097	0.173	0.215
Anomalous	0.107	0.217	0.343	0.417	**0.087**	0.158	0.289	0.340	0.050	0.095	0.170	0.205
Dominant	**0.127**	**0.237**	**0.393**	**0.470**	0.084	**0.162**	**0.304**	**0.396**	**0.052**	**0.098**	**0.180**	**0.248**

Table 2: Performance of different anomaly detection methods w.r.t. precision@K and recall@K.

anomalous neighborhoods rather than nodes, which also result in relatively poor performance.

- The residual analysis based models (Radar and Anomalous) are superior to the conventional anomaly detection methods (LOF, SCAN and AMEN). However, these models are still limited by their shallow mechanisms to handle the network sparsity, data nonlinearity, and complex modality interactions issues.
- Dominant shows a stronger ability to rank anomalies on higher positions according to the results of precision@K and recall@K. It can achieve better detection performance when the objective is to find more true anomalies within the ranking list of limited length.

4.4 Parameter Analysis Next, we investigate the impact of the controlling parameter α in our proposed Dominant framework and report the performance variance results in Figure 3. Here we present the AUC values on the three attributed network datasets. The controlling parameter α balances the impact of attribute reconstruction errors and structure reconstruction errors on model training and anomaly scores computation. In two extreme cases, Dominant will only consider the structure reconstruction errors when α is set to 1 while merely consider the attribute reconstruction errors when α is set 0. The results indicate that it is necessary to find a balance between the structure reconstruction errors and attribute reconstruction errors for achieving a better performance. The reasonable choice of α is around 0.4 to 0.7 for BlogCatalog and Flickr datasets, and 0.5 to 0.8 for ACM dataset.

5 Related Work

In this section, we briefly review related work in two aspects: (1) anomaly detection on attributed networks; and (2) deep learning on network data.

5.1 Anomaly Detection on Attributed Networks As attributed networks are increasingly used to model a wide range of complex systems, the studies of anomaly detection on attributed networks have attracted a lot of attention. Generally, the existing methodologies can be divided into three categories: the first category of anomaly detection methods aims to spot anomalies with community or ego-network analysis. For instance, CODA attempts to simultaneously find communities as well as spot community anomalies within a unified probabilistic model [10]. AMEN [24] considers the ego-network information for each node and discovers anomalous neighborhoods on attributed networks. Besides that, another family of methods is focused on spotting abnormal nodes in a node feature subspace [28, 27, 21]. For example, GOutRank [21] conducts anomaly ranking on attributed networks based on subspace cluster analysis. ConSub [28] takes subspace selection algorithm as a pre-processing step before anomaly detection. FocusCO [25] focuses on community anomalies on a predefined subspace from user preferences. In addition to the methods mentioned above, residual analysis has emerged as another common way to measure the abnormality of nodes on attributed networks. In particular, Radar [17] characterizes the residuals of attribute information and its coherence with network information for anomaly detection. ANOMALOUS [23] further incorporates CUR decomposition into the residual analysis to alleviate the adverse impacts of noisy features for anomaly detection. Despite their fruitful progress, these models are limited by their shallow mechanisms and are incapable of han-

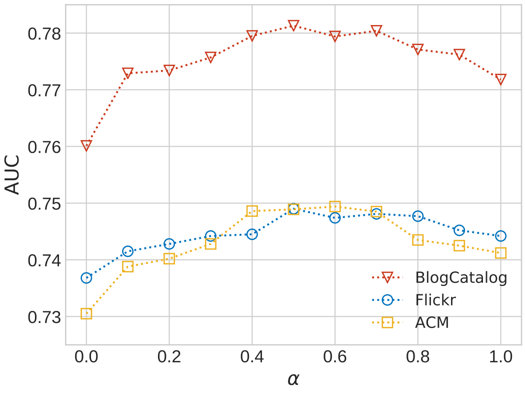


Figure 3: Impact of different α w.r.t. AUC values.

dling the critical issues of attributed networks, such as network sparsity, data nonlinearity and complex modality interactions among different information sources.

5.2 Deep Learning on Networked Data With the growing research interests on deep learning, tremendous efforts have been devoted to developing deep neural networks on networked data for various learning tasks [6, 33, 22, 5, 36, 3]. As one of the first attempts, HNE [6] develops a heterogeneous deep model to embed heterogeneous network data into a unified latent feature space. Afterward, a surge of deep autoencoder based models [33, 5, 9] have been proposed for network representation learning, and render state-of-the-art performance by their strong capability in capturing highly non-linear properties of data. Among them, SDNE [33] exploits the first and second order node proximity by extending the traditional autoencoder framework. TriDNR [22] captures the inherent correlations between structure, node content and label information via a tri-party autoencoder architecture. Meanwhile, recent research advances on graph convolutional network (GCN) [16, 7, 12] demonstrate superior learning performance by considering neighbors of nodes that are multiple hops away. In particular, GCN [16] takes the structure and attribute information as input, and extends the operation of convolution on network data in the spectral domain for embedding representation learning. GraphSAGE[12] enables inductive representation learning on graph structured data by learning a function that generates embeddings by sampling and aggregating features from a nodes local neighborhood. Nevertheless, all these methods focus on learning embedded representations of nodes, it is still not clear how to perform anomaly detection on top of the deep neural networks. Even though the recently proposed NetWalk [35] combines network representation learning and anomaly detection in a joint framework, it is proposed to solve the problem of anomaly detection on dynamic networks, which cannot be directly applied to our attributed network scenario.

6 Conclusion

In this paper, we make the first investigation on the research problem of anomaly detection on attributed networks by developing a carefully designed deep learning model. Specifically, we address the limitations of existing methods and model the attributed networks with graph convolutional network (GCN). As GCN handles the high-order node interactions with multiple layers of nonlinear transformations, it alleviates the network sparsity issue and can capture the nonlinearity of data as well as the complex interactions between two sources of information on attributed networks. To further enable the detection of anomalous nodes, we introduce a deep autoencoder framework to reconstruct the original attributed network with the learned node embeddings from GCN. The reconstruction errors of nodes are then employed to flag anomalies. The experimental results demonstrate the superiority of the proposed deep model over the state-of-the-art methods. Future work can be focused on two aspects: first we will investigate if the proposed deep model is vulnerable to data poisoning attacks as intelligent attackers can inject malicious samples to avoid the anomalies being detected; second, we will study how to develop robust anomaly detectors in the presence of adversarial attacks.

7 Acknowledgement

This material is based upon work supported by, or in part by, the Nation Science Foundation (NSF) grant 1614576, and the Office of Naval Research (ONR) grant N00014-16-1-2257.

References

[1] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: A survey. *DMKD*, 29(3):626–688, 2015.

[2] Leman Akoglu, Hanghang Tong, Brendan Meeder, and Christos Faloutsos. Pics: Parameter-free identification of cohesive subgroups in large attributed graphs. In *SDM*, pages 439–450, 2012.

[3] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[4] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based

local outliers. In *ACM Sigmod Record*, volume 29, pages 93–104, 2000.

[5] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In *AAAI*, pages 1145–1152, 2016.

[6] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In *KDD*, pages 119–128, 2015.

[7] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*, pages 3844–3852, 2016.

[8] Kaize Ding, Jundong Li, and Huan Liu. Interactive anomaly detection on attributed networks. In *WSDM*, 2019.

[9] Hongchang Gao and Heng Huang. Deep attributed network embedding. In *IJCAI*, pages 3364–3370, 2018.

[10] Jing Gao, Feng Liang, Wei Fan, Chi Wang, Yizhou Sun, and Jiawei Han. On community outliers and their efficient detection in information networks. In *KDD*, pages 813–822, 2010.

[11] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. 2016.

[12] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NIPS*, pages 1024–1034, 2017.

[13] Xiao Huang, Jundong Li, and Xia Hu. Label informed attributed network embedding. In *WSDM*, pages 731–739. ACM, 2017.

[14] Xiao Huang, Qingquan Song, Jundong Li, and Xia Hu. Exploring expert cognition for attributed network embedding. In *WSDM*, 2018.

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[16] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2016.

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

[18] Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. Attributed network embedding for learning in a dynamic environment. In *CIKM*, pages 387–396, 2017.

[19] Jundong Li, Xia Hu, Jiliang Tang, and Huan Liu. Unsupervised streaming feature selection in social media. In *CIKM*, pages 1041–1050, 2015.

[20] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[21] Emmanuel Muller, Patricia Iglesias Sánchez, Yvonne Mulle, and Klemens Bohm. Ranking outlier nodes in subspaces of attributed graphs. In *ICDE Workshop*, pages 216–222, 2013.

[22] Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. Tri-party deep network representation. *Network*, 11(9):12, 2016.

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

[24] Bryan Perozzi and Leman Akoglu. Scalable anomaly ranking of attributed neighborhoods. In *SDM*, pages 207–215, 2016.

[25] Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sánchez, and Emmanuel Müller. Focused clustering and outlier detection in large attributed graphs. In *KDD*, pages 1346–1355, 2014.

[26] Joseph J Pfeiffer III, Sebastian Moreno, Timothy La Fond, Jennifer Neville, and Brian Gallagher. Attributed graph models: Modeling network structure with correlated attributes. In *WWW*, pages 831–842, 2014.

[27] Patricia Iglesias Sánchez, Emmanuel Müller, Oretta Irmler, and Klemens Böhm. Local context selection for outlier ranking in graphs with multiple numeric node attributes. In *SSDBM*, page 16, 2014.

[28] Patricia Iglesias Sánchez, Emmanuel Muller, Fabian Laforet, Fabian Keller, and Klemens Bohm. Statistical selection of congruent subspaces for mining attributed graphs. In *ICDM*, pages 647–656, 2013.

[29] Cosma Rohilla Shalizi and Andrew C Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, 2011.

[30] David B Skillicorn. Detecting anomalies in graphs. In *ISI*, pages 209–216, 2007.

[31] Xiuyao Song, Mingxi Wu, Christopher Jermaine, and Sanjay Ranka. Conditional anomaly detection. *TKDE*, 19(5):631–645, 2007.

[32] Hanghang Tong and Ching-Yung Lin. Non-negative residual matrix factorization with application to graph anomaly detection. In *SDM*, pages 143–153, 2011.

[33] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In *KDD*, pages 1225–1234, 2016.

[34] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas AJ Schweiger. Scan: A structural clustering algorithm for networks. In *KDD*, pages 824–833, 2007.

[35] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks. In *KDD*, pages 2672–2681, 2018.

[36] Zhen Zhang, Hongxia Yang, Jiajun Bu, Sheng Zhou, Pinggang Yu, Jianwei Zhang, Martin Ester, and Can Wang. Anrl: Attributed network representation learning via deep neural networks. In *IJCAI*, pages 3155–3161, 2018.

[37] Chong Zhou and Randy C Paffenroth. Anomaly detection with robust deep autoencoders. In *KDD*, pages 665–674, 2017.