Generative Adversarial Attributed Network Anomaly Detection

Zhenxing Chen
JD Digits
Beijing, China
chenzhenxing1@jd.com

Bo Liu
JD Finance America Coporation
Mountain View, USA
bo.liu2@jd.com

Meiqing Wang
JD Digits
Beijing, China
wangmeiqing@jd.com

Peng Dai
JD Finance America Coporation
Mountain View, USA
peng.dai@jd.com

Jun Lv
JD Digits
Beijing, China
lvjun@jd.com

Liefeng Bo
JD Finance America Coporation
Santa Clara, America
liefeng.bo@jd.com

ABSTRACT

Anomaly detection is a useful technique in many applications such as network security and fraud detection. Due to the insufficiency of anomaly samples as training data, it is usually formulated as an unsupervised model learning problem. In recent years there is a surge of adopting graph data structure in numerous applications. Detecting anomaly in an attributed network is more challenging than the sample based task because of the sample information representations in the form of graph nodes and edges. In this paper, we propose a generative adversarial attributed network (GAAN) anomaly detection framework. The fake graph nodes are generated by a generator module with Gaussian noise as input. An encoder module is employed to map both real and fake graph nodes into a latent space. To encode the graph structure information into the node latent representation, we compute the sample covariance matrix for real nodes and fake nodes respectively. A discriminator is trained to recognize whether two connected nodes are from the real or fake graph. With the learned encoder module output, an anomaly evaluation measurement considering the sample reconstruction error and real-sample identification confidence is employed to make prediction. We conduct extensive experiments on benchmark datasets and compare with state-of-the-art attributed graph anomaly detection methods. The superior AUC score demonstrates the effectiveness of the proposed method.

CCS CONCEPTS

• **Theory of computation → Unsupervised learning and clustering**.

KEYWORDS

GAN, anomaly detection, attributed networks

ACM Reference Format:
Zhenxing Chen, Bo Liu, Meiqing Wang, Peng Dai, Jun Lv, and Liefeng Bo. 2020. Generative Adversarial Attributed Network Anomaly Detection. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3340531.3412070

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412070

1 INTRODUCTION

Anomaly detection [3] aims to identify unexpected instances that deviate from the pattern of the majority instances. A typical attributed network carries rich information including node attributes and links describing the relationship between nodes. With the increasing application of network data structure in real-world applications such as social network, protein data analysis and financial services [12], the methodology of anomaly detection in attributed networks, i.e. identifying unusual nodes that are significantly different from others, attracts more attention. The technique have been applied in numerous use cases such as social spammer detection [9], financial fraud detection [10] and intrusion detection [4].

One of the main challenges of anomaly detection in attributed network is how to model the heterogeneous information representation including graph nodes and edges. Traditional methods such as LOF [2] and SCAN [18] only consider either node attributes or network topology structure, leading to unsatisfactory performance. To fuse attribute and network structure information synergistically, ANOMALOUS [14] proposes to fuse node attributes and network structure information by jointly learning attribute selection and anomaly detection with CUR decomposition and residual analysis. DOMINANT [5] takes advantage of the synergy between graph neural network and auto-encoder to spot anomalies by measuring the reconstruction errors of nodes from both the structure and the attribute perspectives. Both node embedding and attribute embedding are learned jointly with a structure auto-encoder and an attribute auto-encoder in AnomalyDAE [6].

In recent years, generative adversarial network (GAN) is proposed as an effective model for high-dimensional data distribution approximation [8]. GAN learns a generator **G** and a discriminator **D** with a min-max optimization framework. The generator is trained to map a randomized input to a sample from the data distribution. Simultaneously the discriminator is trained to identify whether an instance is a real sample or from the generator. This mechanism has been successfully applied to boost many model learning tasks in recommender system [16], query expansion [13] and network embedding [17]. GAN-based methods have achieved

impressive performance in anomaly detection on image and sequence data [1, 7, 19]. In this paper, we propose a generative adversarial attributed network (GAAN) based anomaly detection method. Specifically, the generator model is designed to generate fake graph nodes from Gaussian noise. A latent space node representation is obtained for both real and fake nodes by an encoder module. The discriminator is to recognize whether the given two edge connected nodes are from the real or fake graph. After the model is learned, an evaluation measurement that considers the sample reconstruction loss and real node pair identification loss is employed for anomaly detection. We test the proposed method on benchmark datasets including BlogCatalog, Flickr and ACM. The experiment comparison shows that the proposed GAAN model significantly outperforms state-of-the-art methods in term of AUC score.

2 GAAN MODEL DESIGN

We denote an undirected attributed network as $\mathcal{G} = (\mathbf{V}, \mathbf{A}, \mathbf{X})$, where $\mathbf{V} = \{v_1, ..., v_n\}$ denotes n graph nodes and $\mathbf{X} \in \mathbb{R}^{n \times m}$ denotes the node attribute matrix. The node connection of graph $\mathcal{G}$ is represented by an adjacency matrix $\mathbf{A}$, where $\mathbf{A}_{i,j} = 1$ if there is an edge between v_i and v_j, otherwise $\mathbf{A}_{i,j} = 0$. Given an attributed network $\mathcal{G}$, the anomaly detection aims to detect the nodes whose patterns differ significantly from the majority reference instances both in attribute and structure.

Inspired by the sample based GAN model for anomaly detection proposed in [19], an encoder **E** is introduced to map the raw attributes **x** to a latent representation **z** in this work. The generator **G** generates **X** from a prior Gaussian distribution. For node pair $< v_i, v_j >$ in the adjacency where $\mathbf{A}_{i,j} > 0$, the discriminator **D** distinguishes the pairs presented in latent representation are either from a generator or from the input graph.

2.1 Generator

The generator approximates the distribution of original attributes **X** from a low-dimensional prior Gaussian distribution. We propose to use a multi-layer perception (MLP) as the generator. MLP learns a layer-wise representation by a linear transformation and a non-linear mapping:

$$\mathbf{H}_{\mathbf{G}}^{(l+1)} = f(\mathbf{W}_{\mathbf{G}}^{(l)}\mathbf{H}_{\mathbf{G}}^{(l)} + \mathbf{b}_{\mathbf{G}}^{(l)}) \tag{1}$$

where $\mathbf{H}_{\mathbf{G}}^{(l)}$ is the input of the l-th perception layer, and $\mathbf{H}_{\mathbf{G}}^{(l+1)}$ is the output of this layer. The Gaussian random noise in d dimensions, where $d \ll m$, is taken as $\mathbf{H}_{\mathbf{G}}^{(0)}$. $\mathbf{W}_{\mathbf{G}}^{(l)}$ is the layer parameter matrix and $\mathbf{b}_{\mathbf{E}}^{(l)}$ is the corresponding bias. f is set to be relu activation function.

2.2 Encoder

The encoder **E** transforms the original node attributes **X** and generator output $\mathbf{X}'$ to a low-dimensional latent space, whose dimension is same with that of prior data distribution in generator. A three-layer MLP module with relu activation function is designed to learn such a latent representation:

$$\mathbf{H}_{\mathbf{E}}^{(l+1)} = f(\mathbf{W}_{\mathbf{E}}^{(l)}\mathbf{H}_{\mathbf{E}}^{(l)} + \mathbf{b}_{\mathbf{E}}^{(l)}) \tag{2}$$

where the attribute matrix **X** and generator output $\mathbf{X}'$ are taken as the inputs of first layer $\mathbf{H}_{\mathbf{E}}^{(0)}$ while the outputs are denoted as **Z** and $\mathbf{Z}'$ respectively.

2.3 Discriminator

To capture the graph structure information, we propose to estimate the adjacency matrix **A** by graph embedding. The estimation is calculated by the dot product of the embedding output and an entry-wise sigmoid function: $\hat{A} = Sigmoid(\mathbf{Z}\mathbf{Z}^T)$ or $\hat{A}' = Sigmoid(\mathbf{Z}'\mathbf{Z}'^T)$, where the embedding **Z** and $\mathbf{Z}'$ are encoded from original node attributes **X** and generator output $\mathbf{X}'$ respectively. $\hat{A}_{ij}$ or $\hat{A}'_{ij}$ indicates the probability of existing a link between node i and node j. For any node pair $< v_i, v_j >$ where $\mathbf{A}_{ij}>0$, the discriminator **D** is trained to distinguish whether the dot product of embedding is from $\hat{A}$(real) or $\hat{A}'$(fake). We minimize the cross-entropy cost of the binary classifier for the GAN model training.

2.4 Optimization

The encoder **E**, generator **G** and discriminator **D** are learned by solving the optimization problem

$$\min_{\mathbf{G}} \max_{\mathbf{D}} \mathcal{L}(\mathbf{D}, \mathbf{E}, \mathbf{G})$$

where $\mathcal{L}(\mathbf{D}, \mathbf{E}, \mathbf{G})$ is defined as

$$\begin{aligned}\mathcal{L}(\mathbf{D}, \mathbf{E}, \mathbf{G}) = \mathbb{E}_{x \sim p_{\mathbf{X}}}[\mathbb{E}_{z \sim p_{\mathbf{E}}(\cdot|x)}[log\mathbf{D}(\mathbf{Z})]] \\ +\mathbb{E}_{x' \sim p_{\mathbf{G}}}[\mathbb{E}_{z' \sim p_{\mathbf{E}}(\cdot|x')}[1 - log\mathbf{D}(\mathbf{G}(\mathbf{Z}'))]]\end{aligned} \tag{3}$$

where $p_{\mathbf{X}}(x)$ is the attribute distribution over the original data and $p_{\mathbf{G}}(x')$ is the distribution over the generator. $p_{\mathbf{E}}(z|x)$ and $p_{\mathbf{E}}(z'|x')$ are latent distributions induced by the input graph and generator respectively. During the training, the discriminator **D** tries to identify the node pairs are from the expected data distribution or generator. Meanwhile the generator **G** is trained to confuse the discriminator.

2.5 Anomaly Detection

After the model training, the anomaly score of each node v_i is defined based on a context reconstruction loss $\mathcal{L}_{\mathbf{G}}$ and a structure discriminator loss $\mathcal{L}_{\mathbf{D}}$:

$$score(v_i) = \alpha\mathcal{L}_{\mathbf{G}}(v_i) + (1 - \alpha)\mathcal{L}_{\mathbf{D}}(v_i) \tag{4}$$

where $\mathcal{L}_{\mathbf{G}}(v_i) = \left\|x_i - x'_i\right\|_2$ and $\mathcal{L}_{\mathbf{D}}(v_i)$ is defined as

$$\mathcal{L}_{\mathbf{D}}(v_i) = \sum_{j=1}^{n} \mathbf{A}_{ij} \cdot \sigma(\hat{A}_{ij}, 1) / \sum_{j=1}^{n} \mathbf{A}_{ij} \tag{5}$$

where σ is cross-entropy loss when the edge is discriminated to be from real distribution. Larger value of $\mathcal{L}_{\mathbf{G}}(v_i)$ means the node attribute x_i can not be well reconstructed by the generator, indicating the node v_i is more likely abnormal. If node v_i is linked to multiple nodes in **A**, $\mathcal{L}_{\mathbf{D}}(v_i)$ is the sum of cross-entropy losses with these nodes. $\mathcal{L}_{\mathbf{D}}(v_i)$ measures the loss with the node pairs identified as real. Therefore a larger $score(v_i)$ indicates the node v_i is more likely to be anomalous.

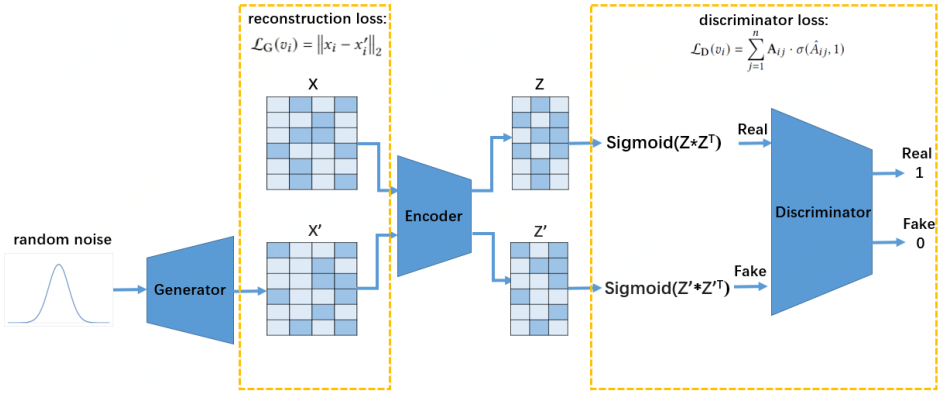


Figure 1: Framework of the proposed model. It consists of three parts: a generator (described in §2.1) recovering nodal attributes from a prior data distribution, an encoder (described in §2.2) transforming original high-level nodal attributes to low-dimension latent space and a discriminator (described in §2.3) discriminating whether an embedding pair is from generator or input graph.

Table 1: Statistics of the Used Experiment Datasets

Database	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

3 EXPERIMENTS

In this section, the effectiveness of the proposed method is evaluated on three real-world datasets. We compare the model performance with six popular methods in anomaly detection.

3.1 Datasets

We perform evaluations on the proposed method with there real-world datasets which are widely used in previous research [5, 6, 15]:

- **BlogCatalog**: BlogCatalog is a blog sharing website, in which the bloggers following each other form a social network and users together with their blogs form node attributes.
- **Flickr**: Flickr is an image hosting and sharing website, in which users following each other form a social network and the interests of users form node attributes.
- **ACM**: ACM is an academic citation network. The citation relations among different papers form a network and paper abstract forms node attribute.

There is no ground truth of anomalies in the above datasets, and we refer to the method in [5] for generating anomalies. We generate a combined set of anomalies for each dataset by perturbing the graph structure and node attributes. The statistics of these three attributed network datasets are summarized in Table 1.

3.2 Experimental Setup

We compare the proposed framework with the following baseline anomaly detection methods:

- **SCAN** [18]: SCAN clusters nodes based on a structural similarity measure and it detects outliers that have only a weak association with a particular cluster.
- **LOF** [2]: LOF assigns a degree of being an outlier to each object based on how isolated the object is with respect to the surrounding neighborhood.
- **DOMINANT** [5]: DOMINANT models the topological structure and nodal attributes simultaneously for node embedding learning with graph convolutional network. It detects anomalies by measuring the reconstruction errors of nodes from both the structure and the attribute perspectives.
- **AnomalyDAE** [6]: AnomalyDAE captures the complex interactions between network structure and node attributes with embeddings represented through a dual autoencoder. It also detects anomalies by measuring the reconstruction errors.
- **EfficientGAN** [19]: In EfficientGAN, an encoder is trained along with a generator and a discriminator. It only considers the node attributes and detects outliers by a combination of reconstruction loss and discriminator loss.

LOF only considers node attributes while SCAN only considers the graph structure. DOMINANT and AnomalyDAE take both node attributes and topological structure into consideration, and they both take the autoencoder framework. An adversarial training framework is adopted in EfficientGAN, but only node attributes are considered. Due to the imbalance between normal and abnormal samples, we adopt the criteria of the area under the ROC curve (AUC) to evaluate the quality of various anomaly detection methods. In the experiment, the parameters of DOMINANT and AnomalyDAE are fixed at the recommended values in corresponding papers. We train the encoder **E**, generator **G** and discriminator **D** with 100 iterations for each dataset. Adam [11] algorithm is utilized as optimizer with learning rate being 0.005. The latent embedding size is set as 32 for all datasets. In addition, both the encoder **E** and the first three layers of generator **G** are composed with three MLP

Table 2: AUC scores of different anomaly detection methods

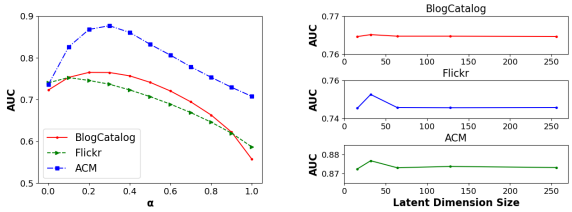
Database	BlogCatalog	Flickr	ACM
IForest	0.722	0.690	0.736
LOF	0.665	0.674	0.723
SCAN	0.484	0.313	0.384
DOMINANT	0.722	0.739	0.751
AnomalyDAE	0.727	0.721	0.778
EfficientGAN	0.719	0.708	0.640
GAAN	**0.765**	**0.753**	**0.877**

layers. The dimension of each MLP layer is 32, 64 and 128 respectively. The parameter α in the anomaly score is set at 0.2, 0.3, 0.1 for BlogCatalog, Flickr, and ACM respectively.

3.3 Result Analysis

3.3.1 Performance Evaluation. The AUC scores for anomaly detection on three datasets are summarized in Table 2. The evaluation results show the proposed method outperforms other baseline methods on all the three real-world datasets. **GAAN** outperforms traditional methods IForest, LOF and SCAN significantly because they only consider the node attributes or topological structure. DOMINANT and AnomalyDAE achieve much better performance since they take both attributes and structure into consideration under the autoencoder framework. Although EfficientGAN only utilizes nodal attributes, it adopts a GAN-based method and achives comparable results to DOMINANT and AnomalyDAE in BlogCatalog and Flickr. It indicates the power of adversarial framework. The proposed GAAN model captures both node and structure information under an adversarial framework, thus it achieves the best performances on all three datasets among all considered methods.

3.3.2 Parameter Sensitivity. In this section, we investigate the impact of trade-off parameter in anomaly score and the dimension of embedding size. α balances the contribution of reconstruction loss $\mathcal{L}_G$ and discriminator loss $\mathcal{L}_D$ on anomaly detection. The AUC scores under varied α are shown in Figure 2(a), The $\mathcal{L}_G$ merely considers the node attributes, while $\mathcal{L}_D$ considers both attributes and graph structure. It indicates $\mathcal{L}_D$ plays a more important role on anomaly detection. Therefore the AUC with $\alpha = 0$ is higher than that with $\alpha = 1$. We conduct experiments on three datasets with varying the dimension of embedding d from 16 to 256 and Figure 2(b) shows the AUC scores under varied latent embedding size. It indicates the method is not sensitive to the embedding size. The highest AUC scores are achieved when d is 32 in all three datasets.



(a) AUC scores under variad α (b) AUC scores under varied latent dimension size

Figure 2: Impact of different parameters on AUC scores

4 CONCLUSION

In this paper, we propose an adversarial attributed network anomaly detection method. In the proposed framework, a generator is trained to reconstruct node attributes while the discriminator discriminates whether the embedding pair encoded by an encoder is from original input or generator output. The anomaly score is determined by a combination of reconstruction loss and discriminator loss. Experiments on real-world datasets achieve state-of-the-art performance, demonstrating the effectiveness of the proposed method.

REFERENCES

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. 2018. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision*.

[2] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. LOF: Identifying Density-Based Local Outliers. 29, 2 (2000).

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly Detection: A Survey. *ACM Comput. Surv.* 41, 1, Article 15 (July 2009), 58 pages. https://doi.org/10.1145/1541880.1541882

[4] Pin-Yu Chen, Sutanay Choudhury, and Alfred O Hero. 2016. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing*.

[5] Kaize Ding, Jundong Li, Bhanushali Rohit, and Huan Liu. 2019. Deep Anomaly Detection on Attributed Networks. In *SIAM International Conference on Data Mining*.

[6] Haoyi Fan, Fengbin Zhang, and Zuoyong Li. 2020. AnomalyDAE: Dual autoencoder for anomaly detection on attributed networks. In *International Conference on Acoustics, Speech, and Signal Processing*.

[7] Elies Gherbi, Blaise Hanczar, Jean-Christophe Janodet, and Witold Klaudel. 2019. An Encoding Adversarial Network for Anomaly Detection. In *Asian Conference on Machine Learning*.

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *International Conference on Neural Information Processing Systems*.

[9] Xia Hu, Jiliang Tang, and Huan Liu. 2014. Online social spammer detection. In *AAAI Conference on Artificial Intelligence*.

[10] Dongxu Huang, Dejun Mu, Libin Yang, and Xiaoyan Cai. 2018. CoDetect: financial fraud detection with anomaly feature detection. (2018).

[11] Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (2014).

[12] Danai Koutra and Christos Faloutsos. 2017. Individual and collective graph mining: principles, algorithms, and applications. *Synthesis Lectures on Data Mining and Knowledge Discovery* 9, 2 (2017), 1–206.

[13] Mu-Chu Lee, Bin Gao, and Ruofei Zhang. 2018. Rare query expansion through generative adversarial networks in search advertising. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

[14] Zhen Peng, Minnan Luo, Jundong Li, Huan Liu, and Qinghua Zheng. 2018. ANOMALOUS: A Joint Modeling Approach for Anomaly Detection on Attributed Networks. In *International Joint Conference on Artificial Intelligence*.

[15] Ryan A. Rossi and Nesreen K. Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI Conference on Artificial Intelligence*.

[16] Jinhui Tang, Xiaoyu Du, Xiangnan He, Fajie Yuan, Qi Tian, and Tat-Seng Chua. 2019. Adversarial training towards robust multimedia recommender system. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[17] Hongwei Wang, Jia Wang, Jialin Wang, Miao Zhao, Weinan Zhang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2017. GraphGAN: Graph Representation Learning with Generative Adversarial Nets. *IEEE Transactions on Knowledge and Data Engineering* (2017).

[18] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas Schweiger. 2007. SCAN: A Structural Clustering Algorithm for Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 3, 824–833.

[19] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar. 2018. Efficient GAN-Based Anomaly Detection. arXiv:arXiv:1802.06222