MVAD_HAN: A Multi-View Based Anomaly Detection Method for Heterogeneous Attributed Networks

Jing Han and Kenan Qin
School of Computer Science, Shaanxi Normal University, Xi’an, Shaanxi, 710062, China

With the frequent occurrence of network security incidents in recent years, it has become very important to detect anomalous behaviour in networks as early and accurately as possible. Anomaly detection can improve the security of complex network systems by detecting abnormal and unreliable nodes, and thus it has become a hot research direction that has attracted wide attention. At present, abstracting real complex systems into complex networks for anomaly detection is the mainstream research method. However, the existing methods still have challenges in extracting network heterogeneity information and attribute information, so we propose a multi-view based anomaly detection method for heterogeneous attributed networks, MVAD_HAN. This method can better extract the heterogeneous structural information and rich attribute information of the network to model heterogeneous attributed networks. Our method adopts an encoder-decoder architecture. First, in the encoder part, we use the Heterogeneous Graph Transformer with multiple views to learn node embeddings that fuse the heterogeneous information of the network. In the decoder part, we use an inner product decoder to reconstruct the network topology, a multilayer perceptron-based decoder to better reconstruct the network attribute information, and a linear projection to reconstruct the node type information of the network. Finally, we compute an anomaly score for each node using three reconstruction errors: network structure, attributes and node type. The higher the reconstruction error of a node, the higher the anomaly score and the higher the probability of an anomaly. Finally, anomalous nodes are identified by ranking the anomaly scores and setting a threshold. We validate the effectiveness of the proposed method on four real-world datasets. The experimental results show that this method outperforms several of the baseline methods and has a good performance in anomaly detection.

***Index Terms*—Heterogeneous Attributed Networks, Anomaly Detection, Multi-View, Network Feature Extraction, Encoder-Decoder.**

I. Introduction

In the real world, many complex systems, such as social networks, transportation networks, and citation networks, are everywhere. With the development of network technology, cybersecurity incidents are numerous. The common examples are scammers in social networks, network intruders or malware in computer networks, damaged equipment in industrial systems or malfunctions in systems. They can cause great damage to real world systems.

An economic study reported [1] that the global economic cost of online fake news reaches about $ 78 billion per year by 2020. In April 2021, a low-level hacking forum exposed data on more than 533 million Facebook users, according to news website Business Insider [2]. Among the information leaked was personal accounts, user names, locations, birthdays, phone numbers and email addresses. This carries a great deal of risk.On June 22, 2022, Northwestern Polytechnical University publicly stated that the university’s email system was subject to a cyberattack [3]. The attack posed a significant security threat to information systems and users’ critical data within the university. In 2023, Oakland, California was the target of multiple ransomware attacks [4]. The attack resulted in over 600 gigabytes of personal information of city workers and residents being compromised.

For these increasingly frequent cybersecurity events, a commonly used cybersecurity model is the PDRR (Protection, Detection, Reaction, Recovery) model. This model can describe the main architecture of the network security defense system. In this model, the most important part is the detection part. How to be able to effectively and accurately detect these abnormal behaviors in the network, to do early detection, early warning, to avoid causing more harm to us has become particularly important.

Research on anomaly detection in computer science dates back to the 1980s. Among them network anomaly detection was an important part from the beginning. In the last decade, extensive connections between real-world objects and advances in graph data mining have brought network anomaly detection to the forefront. One of the most important changes is that network anomaly detection has evolved from relying heavily on the domain knowledge of human experts to machine learning techniques that eliminate human intervention [5]. More recently it has evolved to various deep learning techniques. For networks with anomalies, the anomalous behavior in them is bound to be reflected in data such as entities and their relationships in the network. Processing and analyzing these data containing information about the network using complex network related techniques is now a mainstream method and is very effective.

Because of the diverse patterns and cumbersome data of complex networks, it is difficult to describe or abstract them with one pattern, which poses a great challenge for anomaly detection in networks. In recent years, research on anomaly detection methods for complex networks has rapidly emerged [6]. For the existing anomaly detection methods, on the one hand, they are mainly conducted for homogeneous attributed networks [7], and seldom consider the heterogeneity of the real attributed networks.However, in the real world, the types of nodes and edges in complex networks are often diverse, so

Manuscript received December 28, 2023; revised January 15, 2024. Corresponding author: Jing Han (email: han-jing@snnu.edu.cn).

network heterogeneity is a non-negligible aspect in network anomaly detection. In addition, existing methods for network attributes are represented using a single vector, ignoring the differences in attributes between different kinds of nodes. Therefore, we propose a multi-view based heterogeneous attributed network anomaly detection method, MVAD_HAN, to alleviate the above problems by taking into account the heterogeneity of the network structure along with the differences in attributes of different kinds of nodes. Specifically, the MVAD_HAN model mainly consists of a multi-view encoder and three decoders, which reconstruct the structure, attributes and node types of the heterogeneous attributed network respectively, and finally detects the anomalies existing in the network based on the reconstruction error. The main contributions of this paper are as follows:

- We propose a multi-view based anomaly detection method for heterogeneous attributed networks to better capture network information in heterogeneous information networks.
- We use encoder-decoder architecture. Multi-view based Heterogeneous Graph Transformer (HGT) is used as encoder. Inner product decoder and Multilayer Perceptron (MLP) decoder are used to reconstruct the network from three aspects: network structure, network attributes and network node types, and reconstruction error is used to detect anomalies.
- The effectiveness of our proposed MVAD_HAN method is demonstrated by conducting experiments on four real-world datasets.

The rest of the paper is organized as follows. In Section II, we present the current research in the related research area. In Section III, we define the concepts involved in this paper and the problems to be addressed. In Section IV, we specifically describe the proposed MVAD_HAN anomaly detection model on heterogeneous attributed networks. In Section V, we compare our method with other baseline methods on four real datasets and analyze the experimental results. In Section VI, our work is summarized and outlook is given.

II. Related Works

This part mainly introduces homogeneous attributed networks [7], heterogeneous attributed networks [8] in anomaly detection [9] related content. We briefly review related work in the following three areas: (1) homogeneous network anomaly detection; (2) heterogeneous network anomaly detection; and (3) multi-view representation learning.

A. Homogeneous Network Anomaly Detection

For e.g. social networks, communication networks and citation networks, in these networks, in addition to the topology of the network itself, each node itself has a series of attributes describing its own characteristics, which can be abstracted as homogeneous attributed networks [7]. Homogeneous network anomaly detection methods mainly include community discovery-based methods [10], subspace selection-based methods, residual analysis-based methods, and deep learning-based methods [11]. The LOF method [12] determines whether a point is anomalous or not by comparing the densities of each point and its neighboring points. Gao et al. proposed a community anomaly detection algorithm [13], which detects outliers in a community and thus identifies members whose outliers significantly deviate from the community. FocusCO method [14] detects anomalies in a subspace by constructing a feature subspace. The ANOMALOUS method [15] is based on CUR decomposition and residual analysis for anomaly detection in attribute networks. The Radar method [16] proposes a learning framework for anomaly detection from the perspective of residual analysis. The DOMINANT approach [17] models attributed networks by designing a deep learning model for anomaly detection. GATAE method [18] uses an attention mechanism to better learn the representation of nodes for anomaly detection.

None of the homogeneous network anomaly detection methods consider the diversity of node types in the network. Whereas the nodes in real complex systems are diverse, so treating the nodes in the network as of the same type is limiting. Therefore we choose to conduct anomaly detection studies on heterogeneous networks.

B. Heterogeneous Network Anomaly Detection

Due to the variety of node and connecting edge types in the network, we consider the heterogeneity of the network on top of the homogeneous attributed network and abstract it as a heterogeneous attribute network [8]. It plays an important role in analysing cyber security based incidents [19]. Among the heterogeneous network anomaly detection methods, One cybersecurity incident study [20] collected network characteristics of organizations from external sources and used a random forest classifier to predict intrusion events in organizations. Sarabi et al.'s risky business study [21] uses publicly available business details to predict the risk of data leakage based on a random forest approach. An innovative statistical framework [22] is proposed to model and forecast multivariate time series using sparse externally organized data. The HinAp framework [23] construct an AHIN model based on attack events and use attribute heterogeneous attentional networks and transformational learning to predict cyber-attack preferences. HinCTI [24] aims to model cyber-threat intelligence and identify threat types to alleviate the burden of heavy analytical work on the part of security analysts. The HINTI framework [25] propose a framework to model heterogeneous networks with respect to their interdependencies to quantify their correlations. Cyevent2vec [26] identify network anomalies by modeling them from an event perspective. However, it is difficult for these existing methods to advance work on them due to the confidentiality of the data used.

Most of these existing methods are studied on real datasets and the data they use are confidential. It is difficult for us to continue the experimental validation advancement work on them. In addition, this part of related work focuses more on solving the problem of network heterogeneity, and needs to be further improved in the extraction of network attribute information.

C. Multi-View Representation Learning

Multi-view representation learning is based on multi-view data to obtain a representation containing useful information and apply it to downstream machine learning tasks [27]. By exploring complementary information from multiple views, a more comprehensive representation of the data is possible. Su et al [28] proposed a multiview CNN that integrates information from multiple 2D views into a single representation based on a convolutional neural network. However, early multi-view representation learning had limitations in capturing interactions between multi-view data. With deep learning and multimodal recurrent neural networks [29], most existing approaches capture interactions between multiple views through feature alignment. A survey [30] fuses separate view-based features into a single compact representation.

Among the existing multiview related works, few works have applied it in the field of anomaly detection. However, the multi-view approach can better extract the attribute information of the nodes in the network. Therefore we draw on the multi-view approach to extract richer node information for better modeling of the network.

III. Problem Statement

In this section, we present some preparatory knowledge to help understand the proposed model. Firstly we define the multi-view heterogeneous attributed network studied in this paper. Secondly, we describe the encoder-decoder architecture used. Finally, we define the problem to be addressed in this paper.

Definition I (Multi-view Heterogeneous Attributed Networks): A multi-view heterogeneous attributed network is denoted as $G = (V, E, A, R)$, where V is the set of nodes, E is the set of edges, A is the type of nodes, and R is the type of edges. The mapping relations are $\tau : \quad V \rightarrow A$ and $\phi : E \rightarrow R$. The structure of G can be represented by the adjacency matrix $A \in \mathbb{R}^{|V|*|V|}$. The value of A_{ij} is 1 if there is a connecting edge between nodes v_i and v_j, and 0 otherwise. The edge e from s to t can be expressed as a triple (s, e, t). The corresponding can be denoted as $(\tau(s), \phi(e), \tau(t))$.

For each node v_i in G there is a fixed-length attribute vector x_i corresponding to it. Different types of nodes correspond to different dimensions of attribute vectors and have different attributed matrices X_a. In addition, the node attribute matrix can be represented by k different features spaces, that is $X_a = \left[X_a^{(1)}, X_a^{(2)}, \ldots, X_a^{(k)}\right]$, which divides X_a into k views.

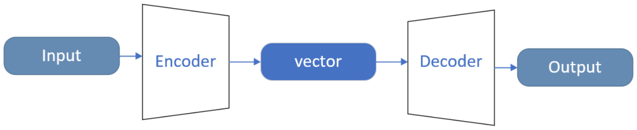

Figure 1. An Encoder-Decoder Architecture

Definition II (Encoder - Decoder): The Encoder - Decoder architecture is shown in Figure 1. It is important to note that encoder-decoder is a model architecture and does not refer to a specific algorithm. In this framework, different algorithms can be employed as encoder and decoder to solve various tasks. The encoding process consists of an encoder that converts the input into a fixed dimensional dense vector. The decoding process then converts that vector into an output. This is an end-to-end learning algorithm.

Problem (Anomaly Detection): For a given multi-view heterogeneous attributed network $G = (V, E, A, R)$, the goal of anomaly detection is to discover nodes that are significantly different from the majority node pattern [5]. In our approach, we use an encoder-decoder structure, where the features of the network are encoded by an encoder. And then the structure, attributes and node types of the network are reconstructed by a decoder. Here, we use the reconstruction error for anomaly detection. In detail, the reconstruction error is used for each node in G to compute an anomaly score and finally a threshold K is set to select the anomalous node.

IV. The Proposed MVAD_HAN Method

The overall framework of MVAD_HAN model is shown in Figure 2, which includes three main parts: feature extraction encoder based on HGT, decoder based on MLP, and anomaly score calculation and anomaly detection. The specific functions of each part are as follows:

- *Encoder:* We adopt a multi-view model based on HGT to learn different node representations from different views. Aggregation using the attention mechanism is used to obtain embedded representations that are more expressive of the characteristic attributes of the nodes of the heterogeneous attributed network.
- *Decoder:* We use inner product decoder to reconstruct the topology of the network, MLP-based decoder to reconstruct the attribute matrix of the network, and linear mapping to reconstruct the type information of the nodes in heterogeneous networks.
- *Anomaly Detection:* We can calculate the anomaly score based on the reconstruction errors of each of the three decoders, then get the overall anomaly score, and then rank the anomalies.The larger the error in the reconstruction process, the more likely the instance is an anomaly.

A. Multi-view Heterogeneous Attributed Network Encoder

Compared to homogeneous networks, heterogeneous networks have diversity of nodes. Existing modeling of heterogeneous networks is mainly through the establishment of meta-paths. However, different datasets require different meta-paths, which leads to a lot of effort in data processing and meta-path selection. Hu et al. [31] proposed Heterogeneous Graph Transformer (HGT) to model heterogeneous graphs. The method does not require manual construction of meta-paths and can generate specialized representations for different types of nodes and edges. Therefore, we use HGT as an encoder for feature extraction of heterogeneous attributed networks.

For (s, e, t), we use s $\in N(t)$ to denote all neighboring nodes of t. For this (s,e,t), we have three basic operations

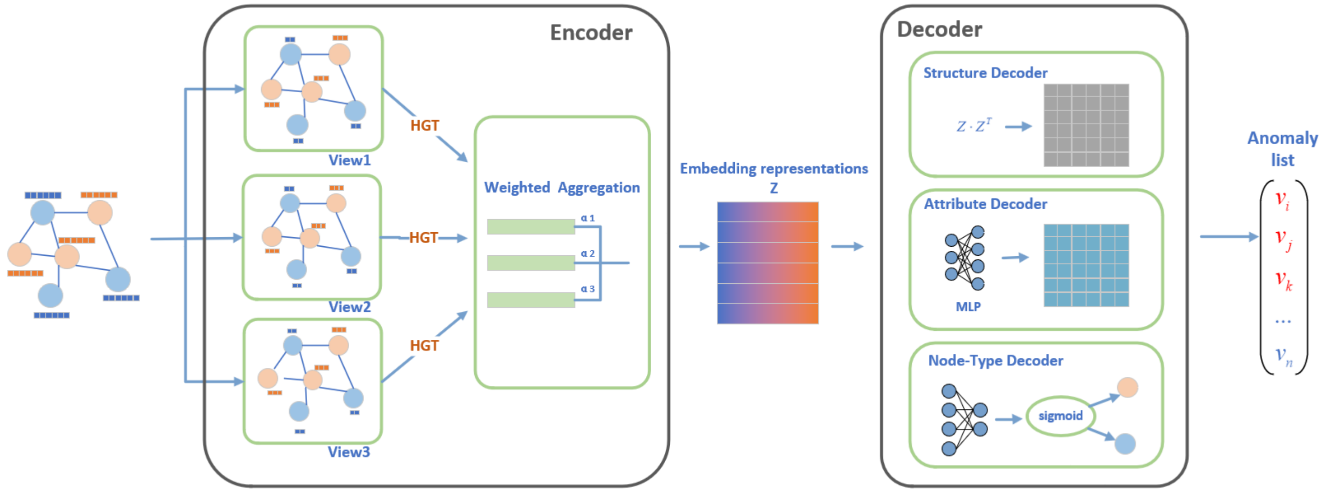

Figure 2. The MVAD_HAN Model Framework

which are Attention, Message and Aggregate. where Attention is used to evaluate the importance of each source node. message denotes the extraction of information from the source node s. Aggregate denotes the use of attention coefficients as weights to aggregate the information from the neighbors.

For the target node t, we can represent the vector of this node as:

$$\tilde{H}^{(l)}[t]=\sum_{\forall s\in N(t)}(\mathrm{Attention}(s,e,t)\cdot\mathrm{Message}(s,e,t)) \quad (1)$$

where $\mathrm{Attention}(s,e,t)$ can be denoted as:

$$\mathrm{Attention}(s,e,t)=H^{(l-1)}[s]W^{Att}_{\phi(e)}H^{(l-1)}[t]\mu(s,e,t) \quad (2)$$

and $\mathrm{Message}(s,e,t)$ can be denoted as:

$$\mathrm{Message}(s,e,t)=H^{(1-1)}[s]W^{Mes}_{\phi(e)} \quad (3)$$

where $W^{Att}_{\phi(e)}$ is an edge-based weight matrix for edge type φ(e). $W^{Mes}_{\phi(e)}$ is a weight matrix containing edge dependencies. And μ is an adaptive scaling tensor on attention.

Then, we get the embedding vector of the target node t:

$$H^{(l)}[t]=RELU\left(\text{ Linear }_{\tau(t)}\tilde{H}^{(l)}[t]\right)+H^{(l-1)}[t] \quad (4)$$

where Linear $_{\tau(t)}$ denotes a linear mapping that maps the embedding vector of the target node t to an attribute dimension of that node type.

For the multiple types of nodes included in the heterogeneous attributed network, we input one view of each node of each type into the HGT to be encoded separately. Thus, the target node t has a total of K HGT models, where:

$$K=\prod_{\forall s\in N(t)\cup\{t\}}k_{\tau(s)} \quad (5)$$

Each of the K outputs of HGT is an embedding vector of a view of the target node t. Here we first map the attributes of each view into the same space. That is, for (s,e,t), the input of s to the i-th view and the j-th view of t of the HGT can be described as:

$$H^{(0)}\left[t_i\right]=H-\text{ Linear }{}^{i}_{\tau(t)}\left(X^{i}_{\tau(t)}\right) \quad (6)$$

$$H^{(0)}\left[s_j\right]=H-\text{ Linear }{}^{i}_{\tau(t)}\left(X^{j}_{\tau(s)}\right) \quad (7)$$

where H - Linear $r^{i}_{\tau(t)}$ denotes a linear mapping. It is indexed by the type of node t. Each type of node has a unique linear mapping to compress the dimensions of each view attribute. t_i denotes the i-th view of node t, and s_j denotes the j-th view of node s.

For the HGT in layer l, we can obtain the embedding representation of t_i as follows:

$$H^{(l,j)}\left[t_i\right]=\mathrm{RELU}\left(\text{ Linear }_{\tau(t)}\tilde{H}^{(l)}\left[t_i\right]\right)+H^{(l-1,j)}\left[t_i\right] \quad (8)$$

which $\tilde{H}^{(l,j)}\left[t_i\right]$ can be calculated by:

$$\tilde{H}^{(l,j)}\left[t_i\right]=\sum_{\forall s\in N(t)}(\mathrm{Attention}(s,e,t)\cdot\mathrm{Message}(s,e,t)) \quad (9)$$

For a node t, the embedding information for each view of that node comes from the view embedding information of its neighboring nodes. Here we use a linear projection $A-Linear$ to map all the K embeddings of each view of t into a given dimension. This embedding can be represented as follows:

$$Z^{(i,j)}_{\tau(t)}=A-\mathrm{Linear}\left(H^{(l,j)}\left[t_i\right]\right) \quad (10)$$

where $Z^{(i,j)}_{t(t)}$ is the potential vector representation of the i-th view of the node t. For t_i and s_j, we can denote $Z^{(i,j)}_{\tau(t)}$ as $Z^{(k)}_{\tau(t)}$, where k takes the value range $[1,K]$.

In order to be able to better learn the importance of each view and get node embedding vectors that contain more comprehensive information about the network, we next use the attention mechanism to aggregate the embeddings of these k views. The potential vector representation of node t can be expressed as:

$$\begin{aligned} Z_{\tau(t)} = & \quad \text{Aggregate}\left(Z_{\tau(t)}^{(k)}\right) \\ = & \sum_{\forall k \in [1,K]} W_v^k \times Z_{\tau(t)}^{(k)} \end{aligned} \tag{11}$$

where W_v^k is a learnable weight vector and k refers to the importance of the k-th view combination.

B. Decoder

In order to obtain a more comprehensive characterization of the heterogeneous attribute network, the structure decoder, attribute decoder and node type decoder are used here to reconstruct the network matrix.

1) Structure Decoder

We reconstruct the topology of the network through a structural decoder. For the node embedding representation obtained via the encoder, we want to find a way to learn the similarity of each node in the hidden vector to generate the output adjacency matrix. The inner product computes the cosine similarity of two vectors, which is useful when we want a distance metric that is invariant to the size of the vectors. Therefore, we choose the inner product decoder to learn the similarity of each node to reconstruct our adjacency matrix.

The network structure is represented by the network's adjacency matrix. Therefore, in the structure decoder, we reconstruct the structure of the original network using an inner product decoder with node embeddings as input:

$$\hat{A} = S\left(Z_{\tau(s)} \cdot \left(Z_{\tau(t)}\right)^T\right) \tag{12}$$

where S is a Sigmoid function and $\hat{A}$ denotes the reconstructed adjacency matrix.

2) Attribute Decoder

We reconstruct the attribute information of the network through an attribute decoder. Here we use MLP as a decoder. It consists of multiple layers and can be used to model nonlinear dependencies. It is robust to nonlinear transformations through activation functions. The goal of the decoder is to remap the compressed embedding vectors into a reconstructed output with the same dimensions as the original input data.

In the attribute decoder, we take the learned node embeddings as input and reconstruct the attribute information of each view through an MLP with l layers:

$$\hat{X}^{(l)} = \text{RELU}\left(Z_{r(t)}^{(l-1)} W^{(l)} + b^{(l)}\right) \tag{13}$$

where l denotes the number of layers of the multilayer perceptron. $Z_{\tau(t)}^{(l-1)}$ denotes the input of layer l, and the input of layer 0 is $Z_{\tau(t)}$, there is $Z^{(0)} = Z_{\tau(t)}$. $W^{(l)}$ and $b^{(l)}$ denote the trainable weights and biases of layer l.

3) Node Type Decoder

Yang et al. [32] used a node-type decoder to reconstruct the type information of nodes in heterogeneous networks to better extract potential representations of the network. Here we also use a One-hot encoder to encode the type information of nodes in the network. The node type information of node t can be encoded as:

$$T[t] = \text{Encoder}(\tau(t)) \tag{14}$$

where $T[v] \in \mathbb{R}^{|V|} \times \mathbb{R}^{|A|}$ is an ont-hot vector and only the position corresponding to the node type of node t is 1, all others are 0.

In the node type decoder, we take as input the potential vector representation of a node. Since the linear mapping method is simple and computationally fast, we use a linear projection to reconstruct it into node vectors. An attention mechanism is used to reflect the importance of each node type to the overall node type information.

$$\hat{T}[v] = \text{S}\left(W_T \cdot T - \text{ Linear }_{\tau(t)}\left(Z_{\tau(t)}\right)\right) \tag{15}$$

where W_T is a learnable weight vector.

C. Loss Function

We use encoders and decoders to reconstruct the structure matrix, attribute matrix, and node type matrix of the network, and employ the reconstruction error as a loss function to train the model. The loss function can be expressed as follows:

$$L = \alpha L_A + \beta L_X + (1 - \alpha - \beta) L_T \tag{16}$$

where α and β are hyperparameters that maintain the balance between structure reconstruction, attribute reconstruction and node type reconstruction. L_A, L_X and L_T denote the reconstruction errors of the structure matrix, attribute matrix and node type matrix, respectively.

$$L_A = \sum_{\forall \dot{\varphi}(e) \in \mathbb{R}} \left\|(\hat{A} - A) \odot \theta_1\right\|_F^2 \tag{17}$$

$$L_X = \sum_{\forall \tau(t) \in A} \left\|(\hat{X} - X) \odot \theta_2\right\|_F^2 \tag{18}$$

$$L_T = \sum_{\forall \tau(v) \in V} \left\|(\hat{T} - T) \odot \theta_3\right\|_F^2 \tag{19}$$

Due to the fact that some edges or attributes are missing in the real world, we set the penalty parameter to impose more penalties on the reconstruction errors of non-zero elements. Where θ_1, θ_2 and θ_3 are penalty parameters that improve the effect of reconstruction. $\odot$ is the hadamard product.

Thus the total loss function L can be expressed as:

$$\begin{aligned} L \quad = \quad & \alpha \sum_{\forall \phi(e) \in \mathbb{R}} \left\|(\hat{A}[v] - A[v]) \odot \theta_1\right\|_F^2 \\ & + \beta \sum_{\forall r(t) \in A} \left\|(\hat{X} - X) \odot \theta_2\right\|_F^2 \\ & + (1 - \alpha - \beta) \sum_{\forall r(v) \in V} \left\|(\hat{T} - T) \odot \theta_3\right\|_F^2 \end{aligned} \tag{20}$$

D. Anomaly Detection

For anomaly detection, we use the reconstruction error as the anomaly score. The reconstruction error usually indicates the degree of anomaly of a node.

For a node v in the network, if the decoder is able to reconstruct its original network information approximately, then it indicates that the probability of the node's anomaly is low. Conversely, if the decoder is not able to reconstruct the network information well, this indicates that the node's pattern deviates from most of the other nodes and the probability of anomaly is high. Therefore, here we use structural reconstruction error, attribute reconstruction error and node type reconstruction error to calculate the anomaly score for each node:

$$\begin{aligned}\mathrm{score}(v) &= \alpha\left\|(\hat{A}[v]-A[v])\odot\theta_1\right\|_F^2 \\ &+\beta\left\|(\hat{X}[v]-X[v])\odot\theta_2\right\|_F^2 \\ &+(1-\alpha-\beta)\left\|(\hat{T}[v]-T[v])\odot\theta_3\right\|_F^2\end{aligned} \tag{21}$$

We map the prediction scores to between 0 and 1 by normalizing the ordering of outliers based on the anomaly scores of the nodes. where α and β are hyperparameters that maintain a balance between structural reconstruction, attribute reconstruction, and node type reconstruction. θ_1, θ_2 and θ_3 are penalty parameters that impose more penalties on the reconstruction errors of non-zero elements.

From the existing studies, it is known that because the information representation of an abnormal node does not conform to the pattern of most nodes, it is more difficult to reconstruct compared to normal nodes. Therefore, the larger the node reconstruction error, the higher the probability of anomalies.

V. Experiments

In order to demonstrate the validity and accuracy of our proposed MVAD_HAN model, we conducted the following experiments. This section focuses on the baseline methodology for comparison, the evaluation metrics for performance evaluation, the performance of the proposed MVAD_HAN on multiple real-world datasets, and the analysis of important parameters in our model.

The framework used for the experiments is the gpu version of pytorch 1.7.1. we set the number of iterations for the four datasets to 100, 100, 80, 80. the Adam algorithm [33] was used for optimisation, and the learning rate was set to 0.01. the number of MLP layers was set to 3.

A. Datasets

We evaluate the performance of MVAD_HAN and validate its effectiveness based on four real-world datasets. The four datasets we used are GossipCop, IMDB, CoAID and PolitiFact datasets. The above datasets are heterogeneous attribute networks without ground truth anomaly labels, and Yang et al. [32] manually injects anomalies in the datasets including structural and attribute anomalies based on existing methods. The specific details of the dataset are shown in Table I.

Table I. Dataset Description.

Datasets	Nodes	Attributes	Edges
GossipCop	N:22140 S:2027	N:1536 S:768	19213
IMDB	M:4278 A:5257	M:3066 A:3066	12908
CoAID	N:5457 S:199	N:1536 S:768	5434
PolitiFact	N:1054 S:285	N:1536 S:768	787

B. Baseline Methods

To demonstrate the overall performance of our proposed method MVAD_HAN , we compare it with five state-of-the-art baseline methods. The details are given below:

- *VGAE[34]:* Kipf and Welling proposed Variational Graph Auto-Encoder, a graph-based auto-encoder, where the encoder is used to obtain vector representations of the nodes, and then the decoder uses the vector representations to reconstruct the graph structure. The method can be used for anomaly detection.
- *DOMINANT[17]:* The problem of anomaly detection on attributed networks is investigated using a GCN-based deep model considering both attribute information and network structure information.
- *DONE[35]:* A self-encoder based deep architecture is proposed, which uses an unsupervised approach to minimize the effect of network embedding of outliers, and detects anomalies by calculating anomaly scores.
- *ALARM[36]:* Peng et al. took into account the characteristics of user preferences and explored node attributes from a multi-attribute view to improve the performance of anomalous node detection.
- *AHEAD[32]:* Considering the network heterogeneous type, an encoder-decoder architecture is used on the heterogeneous attributed network to obtain the anomaly score of each node for anomaly detection by reconstruction.

C. Experiment Results and Analysis

This part is about the presentation and analysis of the experimental results. We mainly describe the three aspects of evaluation indexes, experimental results and parameter analyses.

1) Evaluation Metrics

In the field of machine learning, the AUC value is often used to evaluate the training effect of a binary classification model. In this paper, we also use the evaluation metric AUC to evaluate the performance of anomaly detection methods. There are four cases for real and detected anomalies: (1) True Positive (TP) : judged as positive and actually positive; (2) False Positive (FP) : judged as positive and actually negative; (3) True Negative (TN) : judged as negative and actually negative; (4) False Negative (FN) : judged negative, but actually positive.

The true positive rate (TPR) and the false positive rate (FPR) of the abnormal test were defined as:

$$TPR = \frac{TP}{TP+FN}, \quad FPR = \frac{FP}{TN+FP}$$

The AUC value is the area under the curve with FPR and TPR as the horizontal and vertical coordinates, respectively. the higher the AUC value, the better the performance of this abnormality detection method.

2) Experimental Results

We compared the proposed MVAD_HAN method with the baseline method, and its AUC scores for anomaly detection are shown in Table II.

Table II. AUC Values On The Four Datasets.

Methods	GossipCop	Politifact	IMDB	CoAID
GCANE	0.4706	0.4496	0.9056	0.6224
DOMINANT	0.5001	0.5586	0.8835	0.8297
DONE	0.5481	0.4776	0.9063	0.6773
ALARM	0.4844	0.5472	0.4605	0.2130
AHEAD	0.5716	0.6287	0.9139	0.8890
MVAD_HAN	**0.6359**	**0.6822**	**0.9163**	**0.9288**

The baseline methods we have selected are all methods that use the self-encoder architecture for network anomaly detection so that we can have a more intuitive comparison result.

From Table II, we can see that our proposed MVAD_HAN method gives better results than the other baseline methods. Because the VGAE, DOMINANT and DONE methods only consider the structural information and attribute information of the network, and do not use the attention mechanism and do not consider the heterogeneity of the network, they are less effective.The ALARM method works poorly when the number of network nodes and connecting edges increases massively, which shows that the method only works better for some datasets and does not have a very good generality. The AHEAD method considers the heterogeneity of the network, but its structure of each decoder is relatively simple. In contrast, our MVAD_HAN method uses the multilayer perceptron to better reconstruct the features of the network.

We use the reconstruction error as the loss function, and we can conclude from Table II. Our method can better extract the network features with effectiveness and feasibility, and it works better than most of the existing methods.

3) Parameter Analysis

In our experiments, we further investigated the effect of each parameter on the experimental results in both Politifact and CoAID datasets. Since either too large or too small learning rate we used during the experiments will affect the method, we finally chose a learning rate of 0.01 after several experimental tests. Figure 3 and Figure 4 show the experimental results when the learning rate is 0.01.

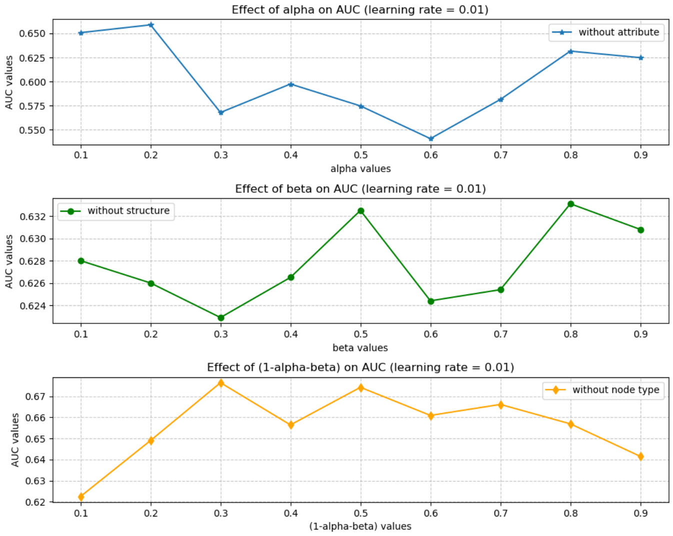

Figure 3. AUC Values For Different Parameters On The Politifact Dataset

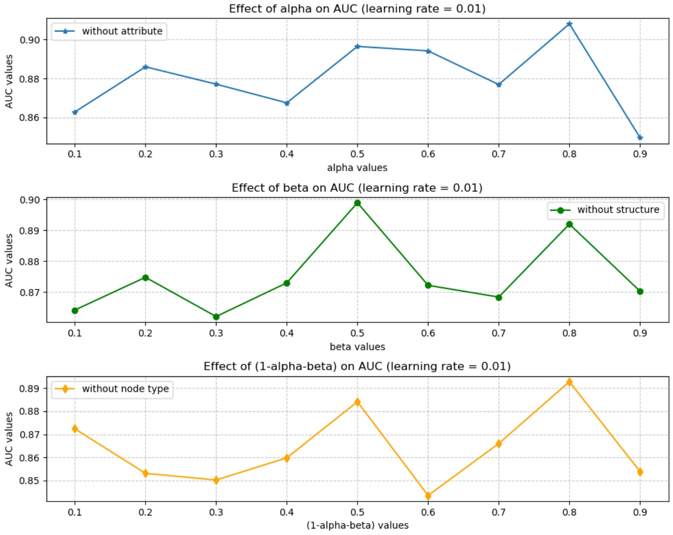

Figure 4. AUC Values For Different Parameters On The CoAID Dataset

The three line graphs in Figure 3 and Figure 4 show the effect of our proposed method when we ignore the network structure, ignore the node attributes and ignore the network node types, respectively. We can see that the method effectiveness decreases no matter which feature we ignore. This illustrates that the topology of the network, the attribute features of the network, and the heterogeneity of the network are all indispensable parts of the network when it comes to feature extraction. This also further illustrates the effectiveness of our method.

VI. Conclusion

In order to better solve the heterogeneous feature extraction problem of networks in heterogeneous attributed network anomaly detection, this paper proposes a multi-view based anomaly detection method MVAD_HAN for heterogeneous attributed networks. The encoder uses multi-view based HGT for feature extraction of heterogeneous attributed networks, which allows us to better extract the attribute and heterogeneity information of the network. In decoder uses MLP structure for better reconstruction of attribute information of the network. Since some edges or attributes are missing in the real world, we set the reconstruction error penalty parameter to impose more penalties on the reconstruction errors of non-zero elements, making the anomalous nodes easier to detect. Our anomaly detection method is validated by experiments on four datasets.

Heterogeneity is an indispensable property for realistic networks. Our method can better extract information about various aspects of the network without using manually designed meta-paths, and is therefore generalizable and practical. Realistic scenarios are often characterized by a large number of interactions and rich information, and thus can all be naturally modeled using our method. Currently, heterogeneous attributed network research has been gradually combined with practice, and gradually applied to the fields of e-commerce, security and medicine, which has a wide range of application prospects. In addition, since the actual network is often dynamically changeable on the basis of heterogeneity, for example, new users and new products are constantly generated in shopping websites. Therefore, in future work, we also need to consider extending the static network anomaly detection work to dynamic anomaly detection by fusing network time series information.

References

[1] S. Ngadiron, A. Abd Aziz, and S. S. Mohamed, "The spread of covid-19 fake news on social media and its impact among malaysians," *MULTI-DISCIPLINARY APPROACHES IN SOCIAL SCIENCES, ISLAMIC & TECHNOLOGY (ICMASIT 2020)*, vol. 13, p. 222, 2020.

[2] A. Talwar, A. Chaudhary, and A. Kumar, "Encryption policies of social media apps and its effect on user's privacy," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1–4, IEEE, 2022.

[3] H. Xiao, H. Wei, Q. Liao, Q. Ye, C. Cao, and Y. Zhong, "Exploring the gamification of cybersecurity education in higher education institutions: An analytical study," in *SHS Web of Conferences*, vol. 166, p. 01036, EDP Sciences, 2023.

[4] S. Hussain, M. Musa, T. Neeshat, R. Batool, O. Ahmed, F. Zaffar, A. Gehani, A. Poggio, and M. K. Yadav, "Towards reproducible ransomware analysis," in *Proceedings of the 16th Cyber Security Experimentation and Test Workshop*, pp. 1–9, 2023.

[5] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using lstm networks," *Computers in Industry*, vol. 131, p. 103498, 2021.

[6] J. Su, Y. Dong, M. Yan, J. Qian, and Y. Xin, "Research progress of anomaly detection for complex networks," *Control Decis*, vol. 36, pp. 1293–1310, 2021.

[7] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data mining and knowledge discovery*, vol. 29, pp. 626–688, 2015.

[8] C. Shi, R. Wang, and W. X, "Survey on heterogeneous information networks analysis and applications," *Journal of Software*, vol. 33, no. 2, pp. 598–621, 2021.

[9] Z. Li, X. Jin, C. Zhuang, and Z. Sun, "Overview on graph based anomaly detection," *Journal of Software*, vol. 32, no. 1, pp. 167–193, 2020.

[10] J. Wang and I. C. Paschalidis, "Botnet detection based on anomaly and community detection," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 2, pp. 392–404, 2016.

[11] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, pp. 1–36, 2021.

[12] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93–104, 2000.

[13] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, "On community outliers and their efficient detection in information networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 813–822, 2010.

[14] B. Perozzi, L. Akoglu, P. Iglesias Sánchez, and E. Müller, "Focused clustering and outlier detection in large attributed graphs," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1346–1355, 2014.

[15] Z. Peng, M. Luo, J. Li, H. Liu, Q. Zheng, *et al.*, "Anomalous: A joint modeling approach for anomaly detection on attributed networks.," in *IJCAI*, pp. 3513–3519, 2018.

[16] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks.," in *IJCAI*, vol. 17, pp. 2152–2158, 2017.

[17] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *Proceedings of the 2019 SIAM International Conference on Data Mining*, pp. 594–602, SIAM, 2019.

[18] Z. You, X. Gan, L. Fu, and Z. Wang, "Gatae: Graph attention-based anomaly detection on attributed networks," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 389–394, IEEE, 2020.

[19] D. Sun, Z. Wu, Y. Wang, Q. Lv, and B. Hu, "Cyber profiles based risk prediction of application systems for effective access control," in *2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–7, IEEE, 2019.

[20] Y. Liu, A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, M. Bailey, and M. Liu, "Cloudy with a chance of breach: Forecasting cyber security incidents," in *24th USENIX Security Symposium (USENIX Security 15)*, pp. 1009–1024, 2015.

[21] A. Sarabi, P. Naghizadeh, Y. Liu, and M. Liu, "Risky business: Fine-grained data breach prediction using business profiles," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 15–28, 2016.

[22] Z. Fang, M. Xu, S. Xu, and T. Hu, "A framework for predicting data breach risk: Leveraging dependence to cope with sparsity," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2186–2201, 2021.

[23] J. Zhao, X. Liu, Q. Yan, B. Li, M. Shao, H. Peng, and L. Sun, "Automatically predicting cyber attack preference with attributed heterogeneous attention networks and transductive learning," *computers & security*, vol. 102, p. 102152, 2021.

[24] Y. Gao, X. Li, H. Peng, B. Fang, and S. Y. Philip, "Hincti: A cyber threat intelligence modeling and identification system based on heterogeneous information network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 708–722, 2020.

[25] J. Zhao, Q. Yan, X. Liu, B. Li, and G. Zuo, "Cyber threat intelligence modeling based on heterogeneous graph convolutional network," in *23rd international symposium on research in attacks, intrusions and defenses (RAID 2020)*, pp. 241–256, 2020.

[26] X. Ma, L. Wang, Q. Lv, Y. Wang, Q. Zhang, and J. Jiang, "Cyevent2vec: Attributed heterogeneous information network based event embedding framework for cyber security events analysis," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 01–08, IEEE, 2022.

[27] Y. Han, L. Qiao, J. Zheng, Z. Kan, L. Feng, Y. Gao, Y. Tang, Q. Zhai, D. Li, and X. Liao, "Multi-view interaction learning for few-shot relation classification," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 649–658, 2021.

[28] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE international conference on computer vision*, pp. 945–953, 2015.

[29] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille, "Deep captioning with multimodal recurrent neural networks (m-rnn)," *arXiv preprint arXiv:1412.6632*, 2014.

[30] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE transactions on knowledge and data engineering*, vol. 31, no. 10, pp. 1863–1883, 2018.

[31] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proceedings of the web conference 2020*, pp. 2704–2710, 2020.

[32] S. Yang, B. Zhang, S. Feng, Z. Tan, Q. Zheng, J. Zhou, and M. Luo, "Ahead: A triple attention based heterogeneous graph anomaly detection approach," in *Chinese Intelligent Automation Conference*, pp. 542–552, Springer, 2023.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *arXiv preprint arXiv:1611.07308*, 2016.

[35] S. Bandyopadhyay, L. N, S. V. Vivek, and M. N. Murty, "Outlier resistant unsupervised deep architectures for attributed network embedding," in *Proceedings of the 13th international conference on web search and data mining*, pp. 25–33, 2020.

[36] Z. Peng, M. Luo, J. Li, L. Xue, and Q. Zheng, "A deep multi-view framework for anomaly detection on attributed networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 6, pp. 2539–2552, 2020.

Jing Han received her B.Sc degree in Computer Science from Shaanxi Normal University, Xian,China,in 2021. She is currently pursuing the Master degree in Computer Science with Shaanxi Normal University, Xian, China. Her research interests include representation learning and anomaly detection of complex networks.

Kenan Qin received her B.Sc degree in Computer Science from Shaanxi Normal University, Xian,China, in 2020. She received her master's degree in Computer Science from Shaanxi Normal University in Xi'an, China, in 2023. Her research interests include anomaly detection in complex network systems.