Hybrid-Order Anomaly Detection on Attributed Networks

Ling Huang, *Member, IEEE*, Ye Zhu, Yuefang Gao, Tuo Liu, Chao Chang, Caixing Liu, Yong Tang, and Chang-Dong Wang, *Member, IEEE*

Abstract—Anomaly detection on attributed networks has received an increasing amount of attention in recent years. Despite the success, most of the existing methods only focus on detecting the abnormal nodes while fail to detect the abnormal subgraphs. In this paper, we define a new problem of hybrid-order anomaly detection on attributed networks, which aims to detect both of the abnormal nodes and subgraphs. To this end, a new deep learning model called Hybrid-Order Graph Attention Network (HO-GAT) is developed, which is able to simultaneously detect the abnormal nodes and motif instances in an attributed network. In order to model the mutual influence between nodes and motif instances, the learning procedures of the node representation and the motif instance representation are integrated into a unified graph attention network with a novel hybrid-order self-attention mechanism. After learning the node representation and the motif instance representation, two decoders are respectively designed to reconstruct the attribute information of the nodes and motif instances, and the hybrid-order topological structure among nodes and motif instances. And finally, the reconstruction errors are utilized as the abnormal score of nodes and motif instances respectively. Extensive experiments conducted on real-world datasets have confirmed the effectiveness of the HO-GAT method.

Index Terms—Anomaly detection, attributed network, hybrid-order, motif, autoencoder, graph attention network

1 Introduction

Anomaly detection on attributed networks is a very important network analysis task. The goal is to find the anomalies that deviate significantly from the majority of the network in terms of some proximities, e.g., topological structure or attribute proximity. It has many applications such as web spam detection, system fraud detection, network intrusion detection, and representation learning [1], [2], [3], [4]. Many efforts have been made on developing anomaly detection methods for attributed networks, which can be roughly categorized into the traditional shallow learning based methods [5], [6], [7], [8], [9] and deep learning based methods [10], [11], [12]. The shallow learning based methods usually adopt the strategies such as residual analysis, local context analysis and density analysis for anomaly detection. Compared with the shallow learning based methods, the deep learning based methods are able to capture the nonlinear relation between nodes and their neighbors by means of learning latent representations in deep space, and therefore generate relatively better anomaly detection results.

Despite the success, most of the existing methods only focus on the anomaly detection at the level of the individual nodes. That is, they are limited to detect the abnormal nodes in terms of topological structure or attribute proximity while fail to detect the abnormal subgraphs. However, it is very important to additionally detect the abnormal subgraphs that significantly deviate from the other subgraphs in terms of the topological structure or attribute proximity [13], [14], [15]. For instance, in the social commerce platform, a group of users may generate false product reviews or fake ratings and clicks so as to boost the profit of the target items or achieve some malicious goal. Detecting the abnormal subgraphs formed by the groups of spam users can significantly improve the service level of the social commerce platform. Different from detecting the abnormal nodes, detecting the abnormal subgraphs involves additionally learning the relation between nodes and subgraphs, and the relation among subgraphs, leading to a more challenging problem. Although a few attempts have been made on detecting the abnormal subgraphs [13], [14], they require query-based supervision information, failing to detect the

- *Ling Huang is with the College of Mathematics and Informatics, Guangdong Provincial Key Laboratory of Public Finance and Taxation with Big Data Application, South China Agricultural University, Guangzhou 510642, China. E-mail: huanglinghl@hotmail.com.*
- *Ye Zhu, Yuefang Gao, and Caixing Liu are with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China. E-mail: zhuye@stu.scau.edu.cn, {gaoyuefang, liu}@scau.edu.cn.*
- *Tuo Liu and Chang-Dong Wang are with the School of Computer Science and Engineering, Guangdong Province Key Laboratory of Computational Science, Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Sun Yat-sen University, Guangzhou 510275, China. E-mail: 616669577@qq.com, changdongwang@hotmail.com.*
- *Chao Chang and Yong Tang are with the School of Computer Science, South China Normal University, Guangzhou 510631, China. E-mail: 1024737268@qq.com, ytang@m.scnu.edu.cn.*

Manuscript received 30 April 2021; revised 1 August 2021; accepted 1 October 2021. Date of publication 6 October 2021; date of current version 8 November 2023.
This work was supported in part by NSFC under Grants 62106079, 61876193, U1811263, and 61772211, in part by the Natural Science Foundation of Guangdong Province under Grant 2020A1515110337, in part by the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under Grant 2020B1212060032, in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2019B020219001, and in part by the Open Foundation of Guangdong Provincial Key Laboratory of Public Finance and Taxation with Big Data Application.
(Corresponding authors: Yuefang Gao and Yong Tang.)
Recommended for acceptance by J. Li, L. He, H. Peng, P. Cui, C. C. Aggarwal, and P. S. Yu.
Digital Object Identifier no. 10.1109/TKDE.2021.3117842

abnormal subgraphs in an unsupervised manner. In addition, they cannot simultaneously detect the abnormal nodes and subgraphs [16].

To address the above issues, in this paper, we for the first time define a new problem of hybrid-order anomaly detection on attributed networks. In the newly defined problem, the goal is to detect both of the abnormal nodes and subgraphs that deviate significantly from the other nodes and subgraphs respectively in terms of the topological structure or attribute proximity. That is, in addition to detect the conventionally defined *structure-abnormal nodes* and *attribute-abnormal nodes*, it is also required to simultaneously detect the newly defined *structure-abnormal subgraphs* and *attribute-abnormal subgraphs*. As shown in the literature, detecting *structure-abnormal nodes* and *attribute-abnormal nodes* on attributed networks is a nontrivial problem due to the heterogeneous but related information encoded in the topological structure and node attributes [17]. Therefore, additionally detecting *structure-abnormal subgraphs* and *attribute-abnormal subgraphs* makes the anomaly detection on attributed networks more challenging because the abnormal scores of both nodes and subgraphs mutually influence each other in terms of topological structure or attribute proximity. Furthermore, unlike the case of individual nodes, it is unclear which types of abnormal subgraphs should be detected.

To this end, we propose a novel deep learning model called Hybrid-Order Graph Attention Network (HO-GAT), which is able to simultaneously detect the structure/attribute-abnormal nodes and the structure/attribute-abnormal subgraphs in an attributed network. In particular, the target type of the subgraphs is mainly focused on the widely studied higher-order structure, namely motif, which is defined to be the dense subgraph occurring in complex networks at numbers that are significantly higher than those in randomized networks preserving the same degree of nodes [18]. This is because motif is the building block of the complex network [19], and therefore it is rational to detect the abnormal instances of the building blocks. That is, the proposed HO-GAT model aims to simultaneously detect the structure/attribute-abnormal nodes and the structure/attribute-abnormal motif instances. In order to take into account the mutual influence between nodes and motif instances, the learning procedures of the lower dimensional node representation and the motif instance representation are integrated into a unified graph attention network, and a *hybrid-order self-attention* mechanism is designed to simultaneously capture the importance: 1) of a node to another node; 2) of a node to a motif instance; 3) of a motif instance to a node; and 4) of a motif instance to another motif instance. After learning the node representation and the motif instance representation, two decoders are respectively designed to reconstruct the attribute information of the original nodes and motif instances, and the hybrid-order topological structure among nodes and motif instances. Finally, the reconstruction errors of the node topological structure/attribute information and the motif instance topological structure/attribute information are utilized as the abnormal score of nodes and motif instances respectively.

The main contributions of this paper are summarized as follows.

1) This paper for the first time defines a new problem of hybrid-order anomaly detection on attributed networks, which aims to detect both of the abnormal nodes and subgraphs that deviate significantly from the other nodes and subgraphs respectively in terms of the topological structure or attribute proximity.
2) We propose a novel deep learning model called Hybrid-Order Graph Attention Network (HO-GAT), which is able to simultaneously detect the structure/attribute-abnormal nodes and the structure/attribute-abnormal subgraphs.
3) Extensive experiments are conducted on several real-world datasets to confirm the effectiveness of the proposed model.

The rest of this paper is organized as follows. In Section 2, the related work will be briefly reviewed. In Section 3, the notations and problem formulations will be introduced. In Section 4, the proposed model will be described in detail, including motif augmented attributed network construction, hybrid-order attributed network encoder, hybrid-order attributed network decoder, loss function and anomaly detection. In Section 5, extensive experiments will be conducted to confirm the effectiveness of the proposed model. Finally, this paper will be concluded in Section 6.

2 Related Work

In [5], Li *et al.* propose to detect anomalies by means of residual analysis on attributed networks, which takes into account both of the residuals of attribute information and its coherence with network information. In [20], Liu *et al.* propose to measure the normality score of a node within its local context, which is able to be accelerated by a parallel algorithm in the case of spare attributed networks. In [21], Peng at al. propose a joint modeling method for anomaly detection on attributed networks which jointly conducts attribute selection and anomaly detection based on CUR decomposition and residual analysis. In [6], Gutiérrez-Gómez *et al.* propose a multi-scale anomaly detection method called MADAN, which is able to discover abnormal nodes and their context at different scales on attributed networks. From the perspective of utilizing human knowledge, Ding *et al.* [7] develop an interactive anomaly detection method for attributed network, which is able to exploit the existing known types of anomalies and explore the unknown types of anomalies by means of the contextual bandit based framework.

Different from the aforementioned work for static attributed network, some efforts have been made on detecting anomalies on dynamic attributed networks. For detecting anomalies on dynamic and multi-attributed networks, Teng *et al.* [8] propose a multi-view time-series hypersphere learning method that utilizes multi-view learning and support vector domain description. In [9], Gahrooei and Paynabar study the problem of anomaly detection on dynamic attributed networks, where the anomalies are discovered from both of the spatial and temporal perspectives. In [22], Xue *et al.* propose a residual analysis based method for dynamic anomaly detection on time-evolving attributed networks.

Deep learning has also been adopted in anomaly detection on attributed networks. In [10], Ding *et al.* propose a Graph Convolutional Network (GCN) based anomaly detection method called Dominant, which is able to handle with the network sparsity and data nonlinearity on attributed networks. In [11], Li *et al.* also propose a spectral convolution and deconvolution based anomaly detection framework, called SpecAE, by projecting an attributed network into a tailored space via GCN for measuring both global and community level anomalies. Different from Dominant that uses the reconstruction error directly for ranking anomalies, SpecAE utilizes density estimation for detecting anomalies. In order to address the inconsistency issues, namely context inconsistency, feature inconsistency and relation inconsistency, Liu *et al.* [23] propose a new Graph Neural Network (GNN) framework called GraphConsis. In [24], Dou *et al.* propose another GNN model named CARE-GNN, which is able to address the two camouflages, namely feature camouflage and relation camouflage. In [25], Chen *et al.* propose a Generative Adversarial Attributed Network (GAAN) anomaly detection framework, the basic idea of which lies in generating fake graph nodes. Following [10], Peng *et al.* [12] propose a multi-view framework for anomaly detection on attributed networks, where a two-layer GCN is adopted for encoding the feature vector in each view of the attributes separately and two different aggregators are respectively designed for integrating the learned features from different views. And then a structure decoder and an attribute decoder are respectively designed for reconstructing the original multi-view attributed network, the reconstruction errors of which are used to calculate the abnormal score of each node. However, there is still a lack of mechanisms for distinguishing topological normality and attribute normality.

Some efforts have been made on detecting group or subgraph anomalies. In [13], [14], some query-based subgraph and clique anomaly detection methods are developed for Heterogeneous Information Networks (HIN). However, they are limited to HIN only and require query information as input. In [26], Yu *et al.* propose to detect the abnormal groups on dynamic social media data via designing a hierarchical Bayes model named group latent anomaly detection. However, the model is limited to the social activity data rather than the usual attributed network. For detecting the abnormal neighborhoods on attributed networks, in [16], Perozzi and Akoglu propose a quality measure called normality, which is the integration of modularity and weighted attributed similarity. Based on the normality measure, an abnormal entity neighborhood detection method called Anomaly Mining of Entity Neighborhoods (AMEN) is developed. Despite of the capability of detecting the abnormal subgraphs, AMEN only takes into account the node-wise context rather than the context of the other subgraphs. In addition, the aforementioned approaches cannot simultaneously discover abnormal nodes and subgraphs.

3 Notations and Problem Formulation

For clarity, Table 1 summarizes the main notations used in this paper. The input is an undirected unweighted attributed network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}$. In this notation, $\mathcal{V}$ is a node set consisting of n nodes, $\mathcal{E}$ is an edge set consisting of m edges, and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ is a node attribute vector set with each vector $\mathbf{f}_i \in \mathbb{R}^{d\times 1}, \forall i = 1, 2, \ldots, n$ denoting the original attribute vector of the ith node. And d is the dimension of the original attribute vectors. If there exists an edge interconnecting a pair of nodes v_i and v_j, then $(v_i, v_j) \in \mathcal{E}$. For simplicity, an adjacency matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is used to denote the interconnection of the n nodes, i.e., $\mathbf{A}_{i,j} = 1$ if $(v_i, v_j) \in \mathcal{E}$, and $\mathbf{A}_{i,j} = 0$ otherwise. Therefore, the input attributed network can be also denoted as $\mathcal{G} = \{\mathcal{V}, \mathbf{A}, \mathcal{F}\}$.

TABLE 1
The Main Notations Used in This Paper

Notations	Descriptions
$\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}$	Undirected unweighted attributed network
$\mathcal{V}$	Node set consisting of n nodes
$\mathcal{E}$	Edge set consisting of m edges
$\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$	Node attribute vector set
$\mathbf{f}_i \in \mathbb{R}^{d\times 1}$	Original attribute vector of the ith node
d	Dimension of the original attribute vectors
$\mathbf{A} \in \mathbb{R}^{n\times n}$	Adjacency matrix of the original network
$\mathbf{M} = \{\mathcal{V}_\mathbf{M}, \mathcal{E}_\mathbf{M}\}$	Motif with node set $\mathcal{V}_\mathbf{M}$ of p nodes and edge set $\mathcal{E}_\mathbf{M}$ of q edges
$\mathbf{I} = \{\mathcal{V}_\mathbf{I}, \mathcal{E}_\mathbf{I}\}$	Motif instance
$\mathcal{M} = \{\mathbf{I}_1, \ldots, \mathbf{I}_{\bar{n}}\}$	Motif instance set consisting of $\bar{n}$ motif instances
$\hat{\mathcal{G}} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}, \hat{\mathcal{F}}\}$	Motif-augmented attributed network
$\hat{\mathcal{V}}$	Motif-augmented node set consisting of $n + \bar{n}$ nodes
$\hat{\mathbf{A}} \in \mathbb{R}^{(n+\bar{n})\times(n+\bar{n})}$	Motif-augmented adjacency matrix
$\hat{\mathcal{F}} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n, \bar{\mathbf{f}}_1, \ldots, \bar{\mathbf{f}}_{\bar{n}}\}$	Motif-augmented node attribute vector set
$\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+\bar{n}}\}$	Input of the graph attention layer
$\{\mathbf{y}_1, \ldots, \mathbf{y}_n, \mathbf{y}_{n+1}, \ldots, \mathbf{y}_{n+\bar{n}}\}$	Output of the graph attention layer
$\{\mathbf{h}_1, \ldots, \mathbf{h}_n, \mathbf{h}_{n+1}, \ldots, \mathbf{h}_{n+\bar{n}}\}$	Motif-augmented node latent representation vector set
l	Dimension of the latent representation vectors of motif-augmented nodes
$\hat{\mathcal{G}}^{Rec} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}^{Rec}, \hat{\mathcal{F}}^{Rec}\}$	Reconstructed motif-augmented attributed network
$\hat{\mathbf{A}}^{Rec} \in \mathbb{R}^{(n+\bar{n})\times(n+\bar{n})}$	Reconstructed motif-augmented adjacency matrix
$\hat{\mathcal{F}}^{Rec} = \{\mathbf{f}_1^{Rec}, \ldots, \mathbf{f}_n^{Rec}, \bar{\mathbf{f}}_1^{Rec}, \ldots, \bar{\mathbf{f}}_{\bar{n}}^{Rec}\}$	Reconstructed motif-augmented node attribute vector set

The existing study of anomaly detection on attributed networks aims to discover the abnormal nodes that significantly deviate from the other nodes in terms of the topological structure or attribute proximity. There is still a lack of methods for additionally detecting the abnormal subgraphs, which is a very important problem as shown in Introduction.

To this end, in this paper, we define a new problem of hybrid-order anomaly detection on attributed networks as follows.

Definition 1. *The problem of hybrid-order anomaly detection on attributed networks aims to simultaneously detect both of the abnormal nodes and subgraphs that deviate significantly from the other nodes and subgraphs respectively in terms of the topological structure or attribute proximity.*

That is, in addition to detect the conventionally defined *structure-abnormal nodes* and *attribute-abnormal nodes*, it is also required to simultaneously detect the newly defined *structure-abnormal subgraphs* and *attribute-abnormal subgraphs*. They are respectively defined as follows.

Definition 2 (Structure-Abnormal Nodes). *Structure-abnormal nodes refer to the nodes that deviate significantly from the rest of the network including all the important subgraphs and the other nodes in terms of the topological structure.*

Definition 3 (Attribute-Abnormal Nodes). *Attribute-abnormal nodes refer to the nodes that deviate significantly from the rest of the network including all the important subgraphs and the other nodes in terms of the attribute proximity.*

Definition 4 (Structure-Abnormal Subgraphs). *Structure-abnormal subgraphs refer to the subgraphs that deviate significantly from the rest of the network including all the nodes and all the other important subgraphs in terms of the topological structure.*

Definition 5 (Attribute-Abnormal Subgraphs). *Attribute-abnormal subgraphs refer to the subgraphs that deviate significantly from the rest of the network including all the nodes and all the other important subgraphs in terms of the attribute proximity.*

According to the above definitions, there are two issues associated with the new problem. The first issue is that it is unclear which types of abnormal subgraphs should be detected. Although some query-based subgraph and clique anomaly detection methods have been developed for Heterogeneous Information Hetworks (HIN) [13], [14], they are limited to HIN only and require query information as input. And in [26], a hierarchical Bayes model can detect the abnormal groups on dynamic social media data. However, it is limited to the social activity data rather than the usual attributed network. There is still a lack of studies on detecting the most representative subgraphs in the usual attributed networks. The second issue is that the abnormal scores of both nodes and subgraphs mutually influence each other in terms of topological structure or attribute proximity. For instance, some neighbor nodes may influence the abnormal score of one subgraph, and vice versa. However, how to model the mutual influence between the nodes and subgraphs remains unknown. In [16], a quality measure is proposed for measuring whether one subgraph (*a.k.a.* neighborhood) is normal. However, it only takes into account the node-wise context rather than the context of the other subgraphs.

4 The Proposed Model

To address the above two issues, in this paper, we propose a novel deep learning model called Hybrid-Order Graph Attention Network (HO-GAT), which is able to simultaneously detect the structure/attribute-abnormal nodes and the structure/attribute-abnormal subgraphs. The main framework of the proposed HO-GAT model is illustrated in Fig. 1.

4.1 Motif and Motif-Augmented Attributed Network Construction

4.1.1 Motif

For the first issue, it is required to detect the most representative subgraphs on attributed networks. Therefore, in this paper, we focus on the widely studied higher-order structure, namely motif, which is the building block of the complex network [19], [27]. It is defined as follows.

Definition 6 (Motif [28]). *Motif is a densely connected subgraph in complex networks with the occurrence number being notably larger than that in a set of randomized networks preserving the same node degrees. The notation $\mathbf{M} = \{\mathcal{V}_{\mathbf{M}}, \mathcal{E}_{\mathbf{M}}\}$ is used to denote motif with $\mathcal{V}_{\mathbf{M}}$ and $\mathcal{E}_{\mathbf{M}}$ being the node set of p nodes and the edge set of q edges respectively. And q is between $p-1$ and $\frac{p(p-1)}{2}$.*

Following the literature [28], [29], [30], the widely adopted Z-score is used to identify the motif, i.e., the densely connected subgraph with the largest Z-score. Please refer to [28] for the detailed definition of Z-score. According to the experiments, the most widely discovered motif is the triangle motif consisting of 3 nodes and 3 edges, although the quadrilateral motif consisting of 4 nodes and 4 edges has also been studied [29]. In this work, following the conventional setting [31], [32], [33], we focus on the 3-node triangle motif, but our technique can be easily extended to other motifs as well.

Definition 7 (Motif Instance). *The motif instance $\mathbf{I} = \{\mathcal{V}_{\mathbf{I}}, \mathcal{E}_{\mathbf{I}}\}$ is a specific occurrence of motif $\mathbf{M}$ in the network.*

For simplicity of notations, a triple of nodes (numerical index ordered) v_1, v_2 and v_3 is denoted as $\{v_1, v_2, v_3\}$, and if there exists a motif instance containing the three nodes, say the jth motif instance, then $\{v_1, v_2, v_3\} = \mathcal{V}_{\mathbf{I}_j}$.

After applying some existing motif search approaches, e.g., [30], [34], [35], [36], [37], [38], [39], [40], [41], a motif instance set can be obtained, which is defined as follows.

Definition 8 (Motif Instance Set). *The motif instance set is defined to be a set $\mathcal{M} = \{\mathbf{I}_1, \ldots, \mathbf{I}_{\bar{n}}\}$ consisting of all the motif instances occurring in the network, i.e., $\mathbf{I}_i, \forall i = 1, \ldots, \bar{n}$ is a motif instance.*

4.1.2 Motif-Augmented Attributed Network Construction

After searching the motif instance set $\mathcal{M}$, a motif-augmented attributed network is constructed by regarding both of the original nodes and the motif instances as the motif-augmented nodes, their interconnection structure as the linkage structure, and their attribute information as the motif-augmented node attribute information. It is denoted as $\hat{\mathcal{G}} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}, \hat{\mathcal{F}}\}$, where $\hat{\mathcal{V}}$, $\hat{\mathbf{A}}$ and $\hat{\mathcal{F}}$ denote the motif-augmented node set, the motif-augmented adjacency matrix and the motif-augmented node attribute vector set respectively, which are explained in detail as follows.

The motif-augmented node set $\hat{\mathcal{V}}$ consists of $n + \bar{n}$ nodes, i.e., the first n nodes are the nodes in the original network

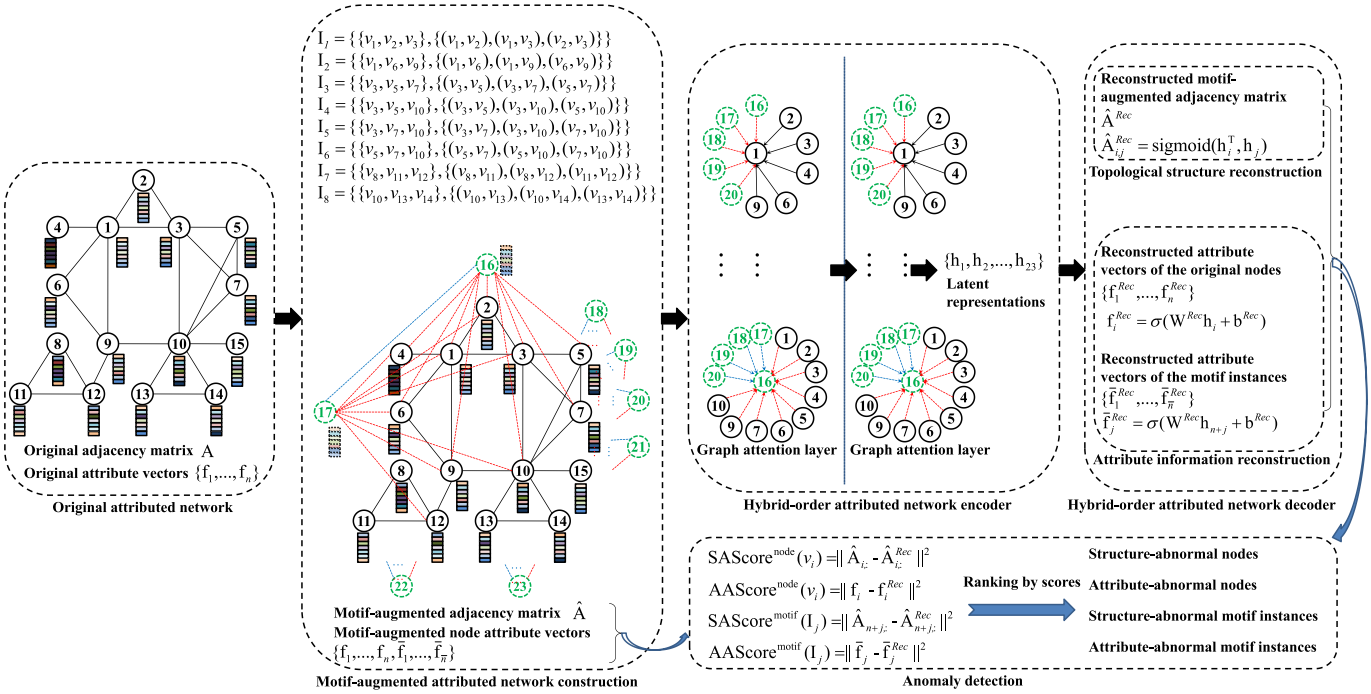


Fig. 1. The overview of the proposed HO-GAT model. In the motif-augmented attributed network construction, the green dashed circles represent the virtual nodes corresponding to the motif instances, the blue dashed lines represent the connections between the virtual nodes, and the red dashed lines represent the connections between the original nodes and the virtual nodes. To avoid clutter, for the virtual nodes, only the connections and the attribute vectors of the first two virtual nodes are plotted. In the hybrid-order attributed network encoder, the dashed arrows represent the corresponding attention based impact of one node to another.

and the rest $\bar{n}$ nodes are the virtual nodes representing the $\bar{n}$ motif instances.

The motif-augmented adjacency matrix $\hat{\mathbf{A}} \in \mathbb{R}^{(n+\bar{n})\times(n+\bar{n})}$ represents the interconnection of the $n+\bar{n}$ motif-augmented nodes as explained below.

- $\forall i = 1, 2, \ldots, n, j = 1, 2, \ldots, n, \hat{\mathbf{A}}_{i,j} = \mathbf{A}_{i,j}$, i.e., the original adjacency matrix is used to represent the interconnection of the n original nodes.
- $\forall i = 1, 2, \ldots, n, j = 1, 2, \ldots, \bar{n}$,

$$\hat{\mathbf{A}}_{i,n+j} = \hat{\mathbf{A}}_{n+j,i} = \begin{cases} 1 & \text{if } v_i \in \mathcal{V}_{\mathbf{I}_j} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

  That is, if the ith node is contained by the jth motif instance, the ith node and the jth motif instance are regarded as connected.
- $\forall i = 1, 2, \ldots, \bar{n}, j = 1, 2, \ldots, \bar{n}$,

$$\hat{\mathbf{A}}_{n+i,n+j} = \hat{\mathbf{A}}_{n+j,n+i} = \begin{cases} 1 & \text{if } \mathcal{V}_{\mathbf{I}_i} \cap \mathcal{V}_{\mathbf{I}_j} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

  That is, if the ith and jth motif instances share at least one common nodes, their corresponding virtual nodes are regarded as connected.

The motif-augmented node attribute vector set $\hat{\mathcal{F}} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n, \bar{\mathbf{f}}_1, \bar{\mathbf{f}}_2, \ldots, \bar{\mathbf{f}}_{\bar{n}}\}$ represents the attribute vectors of the $n+\bar{n}$ motif-augmented nodes. In particular, the first n attribute vectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ are the attribute vectors of the n original nodes. And the rest $\bar{n}$ attribute vectors $\bar{\mathbf{f}}_1, \bar{\mathbf{f}}_2, \ldots, \bar{\mathbf{f}}_{\bar{n}}$ are the attribute vectors of the $\bar{n}$ virtual nodes corresponding to the $\bar{n}$ motif instances, each of which is defined as the mean of the three attribute vectors of the three nodes contained by the corresponding motif instance. That is, let j_1, j_2, and j_3 denote the three indexes (numerically ordered) of the three nodes contained by the jth motif, i.e., $\{v_{j_1}, v_{j_2}, v_{j_3}\} = \mathcal{V}_{\mathbf{I}_j}, \forall j$, we have

$$\bar{\mathbf{f}}_j = \frac{1}{3}(\mathbf{f}_{j_1} + \mathbf{f}_{j_2} + \mathbf{f}_{j_3}) \in \mathbb{R}^{d\times 1}. \tag{3}$$

4.2 Hybrid-Order Attributed Network Encoder

For the second issue, in order to model the complex mutual influence between nodes and motif instances in terms of topological structure or attribute proximity, a Hybrid-Order Graph Attention Network (HO-GAT) is designed to automatically capture the importance: 1) of a node to another node; 2) of a node to a motif instance; 3) of a motif instance to a node; and 4) of a motif instance to another motif instance.

In particular, a graph attention layer is used to transform the original attribute vector of each node and the original attribute vector of each motif instance into low-dimensional representation vectors, in which the complex relations between nodes, between nodes and motif instances, and between motif instances can be encoded.

The input of the graph attention layer is a set of motif-augmented node feature vectors, i.e., $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+\bar{n}}\}$.

1) The first n motif-augmented node feature vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ denote the feature vectors of the n original nodes, which are initialized as the original attribute vectors in the first graph attention layer, i.e., $\mathbf{x}_i = \mathbf{f}_i, \forall i = 1, 2, \ldots, n$.

2) And the remaining $\bar{n}$ motif-augmented node feature vectors $\{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+\bar{n}}\}$ denote the feature vectors of the $\bar{n}$ motif instances, which are initialized as the attribute vectors of the $\bar{n}$ motif instances, i.e., $\mathbf{x}_{n+j} = \bar{\mathbf{f}}_j, \forall j = 1, 2, \ldots, \bar{n}$.

The output of the graph attention layer is a new set of motif-augmented node feature vectors, denoted as $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{y}_{n+1}, \ldots, \mathbf{y}_{n+\bar{n}}\}$, where $\mathbf{y}_i \in \mathbb{R}^{d' \times 1}, \forall i = 1, \ldots, n+\bar{n}$ denotes the higher-level feature vector of the ith motif-augmented node in the motif-augmented attributed network $\mathcal{G}$.

The main procedure of the graph attention layer is described as follows. First of all, a linear transformation which is shared by all the motif-augmented nodes and parametrized by a weight matrix $\mathbf{W}^{xy} \in \mathbf{R}^{d' \times d}$, is performed on each motif-augmented node, i.e., $\mathbf{W}^{xy}\mathbf{x}_i, \forall i = 1, 2, \ldots, n+\bar{n}$. Then, a *self-attention* is performed on the motif-augmented nodes, which is a shared attentional mechanism $a : \mathbb{R}^{d' \times 1} \times \mathbb{R}^{d' \times 1} \rightarrow \mathbb{R}$ for computing the attention coefficients as follows:

$$e_{ij} = a\big(\mathbf{W}^{xy}\mathbf{x}_i, \mathbf{W}^{xy}\mathbf{x}_j\big). \tag{4}$$

The attention coefficient e_{ij} indicates the importance of node j's features to node i. In particular, following the setting in [42], the attention mechanism a is a single-layer feedforward neural network parametrized by a weight vector $\mathbf{a} \in \mathbb{R}^{2d' \times 1}$ and applying the LeakyReLU nonlinearity with negative input slope 0.2. That is, Eq. (4) can be expanded as

$$e_{ij} = \mathrm{LeakyReLU}\Big(\mathbf{a}^\top \big[\mathbf{W}^{xy}\mathbf{x}_i \| \mathbf{W}^{xy}\mathbf{x}_j\big]\Big), \tag{5}$$

where $.^\top$ denotes the transposition and $\|$ is the concatenation operation.

In order to inject the topological structure of the motif-augmented attributed network into the above attention coefficients and make them easily comparable across different motif-augmented nodes, a neighbor-node based softmax function is applied, by which the normalized attention coefficients are obtained as follows:

$$\alpha_{ij} = \mathrm{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}, \tag{6}$$

where $\mathcal{N}_i$ denotes the indexes of the neighbor motif-augmented nodes of motif-augmented node i in the motif-augmented attributed network, including itself. That is, $\mathcal{N}_i = \{k | \forall k = 1, \ldots, n+\bar{n}, \hat{A}_{ik} = 1\}$.

The normalized attention coefficients α_{ij} are further used to compute a linear combination of the corresponding feature vectors, resulting in the output feature vector for each motif-augmented node, i.e.,

$$\mathbf{y}_i = \sum_{j \in \mathcal{N}_i} \alpha_{ij}\mathbf{W}^{xy}\mathbf{x}_j. \tag{7}$$

Notice that, in the above procedure, the weight matrix $\mathbf{W}^{xy} \in \mathbf{R}^{d' \times d}$ and the weight vector $\mathbf{a} \in \mathbb{R}^{2d' \times 1}$ are shared by all the motif-augmented nodes and all the motif-augmented node pairs respectively in one specific graph attention layer.

In our experiments, two graph attention layers are utilized, which are able to capture the complex mutual influence between nodes and motif instances in terms of topological structure or attribute proximity. And the dimension of the first layer is set to be two times of the second layer, i.e., the dimension of latent representations. After the HO-GAT-based encoding procedure, a set of latent representations of the $n+\bar{n}$ motif-augmented nodes (i.e., n original nodes and $\bar{n}$ motif instances) can be obtained, denoted as $\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n, \mathbf{h}_{n+1}, \ldots, \mathbf{h}_{n+\bar{n}}\}$, which are of size l, i.e., $\mathbf{h}_i \in \mathbb{R}^{l \times 1}, \forall i = 1, 2, \ldots, n+\bar{n}$.

4.3 Hybrid-Order Attributed Network Decoder

After obtaining the latent representations of the $n+\bar{n}$ motif-augmented nodes, the motif-augmented attributed network is reconstructed, denoted as $\hat{\mathcal{G}}^{Rec} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}^{Rec}, \hat{\mathcal{F}}^{Rec}\}$. $\hat{\mathbf{A}}^{Rec} \in \mathbb{R}^{(n+\bar{n}) \times (n+\bar{n})}$ denotes the reconstructed motif-augmented adjacency matrix, with the (i,j)th entry indicating the connection strength of the motif-augmented nodes i and j. And $\hat{\mathcal{F}}^{Rec} = \{\mathbf{f}_1^{Rec}, \ldots, \mathbf{f}_n^{Rec}, \bar{\mathbf{f}}_1^{Rec}, \ldots, \bar{\mathbf{f}}_{\bar{n}}^{Rec}\}$ denotes the reconstructed attribute vectors with $\mathbf{f}_i^{Rec}, \forall i = 1, \ldots, n$ being the reconstructed attribute vector for the ith original node, and $\bar{\mathbf{f}}_j^{Rec}, \forall j = 1, \ldots, \bar{n}$ being the reconstructed attribute vector for the jth motif instance. The reconstruction error will be utilized for calculating the abnormal scores of nodes and motif instances as will be elaborated in the next subsection.

4.3.1 Topological Structure Reconstruction

One widely adopted strategy of the topological structure reconstruction lies in calculating the inner product between the latent representations of the two nodes and applying a sigmoid activation function. We have

$$\hat{\mathbf{A}}_{i,j}^{Rec} = \mathrm{sigmoid}(\mathbf{h}_i^\top \mathbf{h}_j). \tag{8}$$

Therefore, for the ith motif-augmented node, the reconstructed motif-augmented adjacency vector can be obtained, denoted as $\hat{\mathbf{A}}_{i,:}^{Rec}$. In particular, the reconstructed motif-augmented adjacency vector of the ith original node is $\hat{\mathbf{A}}_{i,:}^{Rec}, \forall i = 1, 2, \ldots, n$. And the reconstructed motif-augmented adjacency vector of the jth motif instance is $\hat{\mathbf{A}}_{n+j,:}^{Rec}, \forall j = 1, 2, \ldots, \bar{n}$.

4.3.2 Attribute Information Reconstruction

The attribute vectors of the n original nodes and the $\bar{n}$ motif instances are reconstructed by a fully-connected layer. In particular, the attribute vector of the ith original node is reconstructed as follows:

$$\mathbf{f}_i^{Rec} = \sigma(\mathbf{W}^{Rec}\mathbf{h}_i + \mathbf{b}^{Rec}), \tag{9}$$

where $\mathbf{W}^{Rec} \in \mathbb{R}^{d \times l}$ and $\mathbf{b}^{Rec} \in \mathbb{R}^{d \times 1}$ are the learnable weight matrix and bias term respectively. Similarly, the attribute vector of the jth motif instance is reconstructed by using the same fully-connected layer as follows:

$$\bar{\mathbf{f}}_j^{Rec} = \sigma(\mathbf{W}^{Rec}\mathbf{h}_{n+j} + \mathbf{b}^{Rec}). \tag{10}$$

4.4 Loss Function and Anomaly Detection

4.4.1 Loss Function

The loss function of the proposed HO-GAT model consists of two main components, namely the structure reconstruction loss and the attribute reconstruction loss.

For the structure reconstruction loss, we aim to maximize the following likelihood,

$$\prod_{i,j}^{n+\bar{n}} (\hat{\mathbf{A}}_{i,j}^{Rec})^{\hat{\mathbf{A}}_{i,j}} \left(1 - \hat{\mathbf{A}}_{i,j}^{Rec}\right)^{(1-\hat{\mathbf{A}}_{i,j})}. \tag{11}$$

Maximizing the above likelihood is equal to minimizing the following negative log-likelihood error,

$$L_s = -\sum_{i,j}^{n+\bar{n}} \left[\hat{\mathbf{A}}_{i,j} \log\left(\hat{\mathbf{A}}_{i,j}^{Rec}\right) + (1 - \hat{\mathbf{A}}_{i,j}) \log\left(1 - \hat{\mathbf{A}}_{i,j}^{Rec}\right)\right]. \tag{12}$$

On the other hand, the attribute reconstruction error is designed to be the sum of the differences between the original attribute vectors and the reconstructed attributed vectors,

$$L_a = \sum_{i=1}^{n} \|\mathbf{f}_i - \mathbf{f}_i^{Rec}\|^2 + \sum_{j=1}^{\bar{n}} \|\bar{\mathbf{f}}_j - \bar{\mathbf{f}}_j^{Rec}\|^2. \tag{13}$$

Finally, the overall loss function of the proposed model is as follows,

$$L = L_s + L_a + \lambda L_{reg}(\Phi), \tag{14}$$

where $L_{reg}(\Phi)$ is the regularization term that is used to avoid the overfitting issue with Φ being all the parameters in the HO-GAT model and λ is a hyper-parameter, which is set to be 0.0015 in the experiments.

4.4.2 Anomaly Detection

In the literature [12], [43], the anomalies are ranked according to the reconstruction errors, i.e., the larger the reconstruction errors, the more likely the objects are anomalies. This is because the anomalies fail to conform to the patterns of the majority and hence cannot be accurately reconstructed. Therefore, in our method, the abnormal scores corresponding to the four target types of the anomaly are respectively defined as follows.

1) *Structure-Abnormal Nodes*: The structure-abnormal score of the ith original node v_i is

$$\mathrm{SAScore}^{\mathrm{node}}(v_i) = \|\hat{\mathbf{A}}_{i,:} - \hat{\mathbf{A}}_{i,:}^{Rec}\|^2. \tag{15}$$

2) *Attribute-Abnormal Nodes*: The attribute-abnormal score of the ith original node v_i is

$$\mathrm{AAScore}^{\mathrm{node}}(v_i) = \|\mathbf{f}_i - \mathbf{f}_i^{Rec}\|^2. \tag{16}$$

3) *Structure-Abnormal Motif Instances*: The structure-abnormal score of the jth motif instance $\mathbf{I}_j$ is

$$\mathrm{SAScore}^{\mathrm{motif}}(\mathbf{I}_j) = \|\hat{\mathbf{A}}_{n+j,:} - \hat{\mathbf{A}}_{n+j,:}^{Rec}\|^2. \tag{17}$$

4) *Attribute-Abnormal Motif Instances*: The attribute-abnormal score of the jth motif instance $\mathbf{I}_j$ is

$$\mathrm{AAScore}^{\mathrm{motif}}(\mathbf{I}_j) = \|\bar{\mathbf{f}}_j - \bar{\mathbf{f}}_j^{Rec}\|^2. \tag{18}$$

For clarity, Algorithm 1 summarizes the main procedure of the proposed approach for hybrid-order anomaly detection on attributed networks.

Algorithm 1. Hybrid-Order Anomaly Detection on Attributed Networks

Input: $\mathcal{G} = \{\mathcal{V}, \mathbf{A}, \mathcal{F}\}$: attributed network.

1: Obtain motif instance set $\mathcal{M} = \{\mathbf{I}_1, \ldots, \mathbf{I}_{\bar{n}}\}$ by Section 4.1.1.
2: Construct motif-augmented attributed network $\hat{\mathcal{G}} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}, \hat{\mathcal{F}}\}$ by Section 4.1.2.
3: Perform the hybrid-order attributed network encoder and decoder to obtain the reconstructed motif-augmented attributed network $\hat{\mathcal{G}}^{Rec} = \{\hat{\mathcal{V}}, \hat{\mathbf{A}}^{Rec}, \hat{\mathcal{F}}^{Rec}\}$ by Sections 4.2 and 4.3 respectively with the loss function in Eq. (14).
4: Calculate the structure/attribute-abnormal scores for nodes, namely $\mathrm{SAScore}^{\mathrm{node}}(v_i)$ and $\mathrm{AAScore}^{\mathrm{node}}(v_i)$ via Eqs. (15) and (16) respectively.
5: Calculate the structure/attribute-abnormal scores for motif instances, namely $\mathrm{SAScore}^{\mathrm{motif}}(\mathbf{I}_j)$ and $\mathrm{AAScore}^{\mathrm{motif}}(\mathbf{I}_j)$ via Eqs. (17) and (18) respectively.
6: Obtain the structure/attribute-abnormal nodes and motif instances via ranking according to the above abnormal scores.

Output: The structure/attribute-abnormal nodes and motif instances.

5 Experiments

In this section, extensive experiments will be conducted on six real-world datasets to confirm the superiority of the proposed HO-GAT method on simultaneously detecting structure/attribute-abnormal nodes and motif instances.

5.1 Experimental Settings

5.1.1 Datasets and Anomaly Generation

In our experiments, six real-world datasets are adopted, including one scholar social network namely Scholat, one scientific co-author network namely AMiner, and four WebKB hyperlinked web page datasets.

1) *Scholat*: Scholat is a scholar social network obtained from https://www.scholat.com/.[1] The nodes represent the scholars and the edges denote the message interactions between the two ending nodes. The breadth first search with random restart is used to craw the nodes. And the nodes of degree no larger than 50 are preserved. The node attribute vectors are extracted by means of applying Principal Components Analysis (PCA) on the brief biographies of the corresponding scholars. After preprocessing and subset selection, the Scholat dataset contains 2022 nodes, 2500 edges and 329 triangle motif instances.

1. The original Scholat dataset and its two anomaly injected versions are available at https://www.scholat.com/portaldownloadFile.html?fileId=11251.

The dimension of the node attribute vectors is 500. The number of the motif-augmented edges is 8361.

2) *AMiner*: AMiner is a scientific co-author network obtained from https://www.aminer.org/data/. The nodes represent the authors and the edges denote the co-author relations between the two ending nodes. The breadth first search with random restart is used to select the nodes from the entire dataset. And the nodes of degree no larger than 50 are preserved. The node attribute vectors are the Bag of Words (BOW) representation of the keywords of the authors' publications. After preprocessing and subset selection, the AMiner dataset contains 2079 nodes, 3812 edges and 2611 triangle motif instances. The dimension of the node attribute vectors is 133. The number of the motif-augmented edges is 55486.
3) *WebKB*:[2] The WebKB datasets are hyperlinked web page datasets consisting of 877 web pages that are gathered from four universities. Accordingly, the WebKB datasets consist of four datasets, namely Cornell, Texas, Washington and Wisconsin, which contain 195, 187, 230 and 265 web pages respectively. Each web page is regarded as a node and the hyperlink between the two web pages is regarded as the edge linking the two corresponding nodes. The node attribute vector is a 0/1-valued word vector representing the absence/presence of each word in the web page from the dictionary of size 1703. Therefore, the dimension of the node attribute vector is 1703. The Cornell dataset contains 195 nodes, 283 edges, 59 triangle motif instances and 819 motif-augmented edges. The Texas dataset contains 187 nodes, 280 edges, 67 triangle motif instances and 1866 motif-augmented edges. The Washington dataset contains 230 nodes, 366 edges, 99 triangle motif instances and 3065 motif-augmented edges. The Wisconsin dataset contains 265 nodes, 459 edges, 120 triangle motif instances and 3039 motif-augmented edges.

Because there is no ground-truth anomaly on the above six datasets, the structure-abnormal nodes (*resp.* structure-abnormal motif instances) are injected by collecting/generating isolated nodes (*resp.* isolated triangle motif instances), the topological structures of which significantly deviate from the other nodes and triangle motif instances. Let ρ_1 and ρ_2 denote the percentages of the structure-abnormal nodes and the structure-abnormal motif instances respectively. The attribute-abnormal nodes and motif instances are generated by following the same settings as [10], [44]. In particular, we will generate the attribute-abnormal motif instances first. For experimental purpose, the percentage of the attribute-abnormal motif instances is set to be the same as the percentage of the structure-abnormal motif instances, i.e., ρ_2. Similarly, the percentage of the attribute-abnormal nodes is set to be the same as the percentage of the structure-abnormal nodes, i.e., ρ_1. For each candidate attribute-abnormal motif instance $\mathbf{I}_j$, 50 random motif instances are selected from the dataset, and among the 50 random motif instances, the motif instance $\mathbf{I}_i$ whose average attribute vector deviates the most from that of the motif instance $\mathbf{I}_j$ in terms of the euclidean distance between the average attribute vectors is picked. Then the attribute vectors of the three nodes contained by this candidate attribute-abnormal motif instance $\mathbf{I}_j$ are replaced by the attribute vectors of the three nodes contained by the motif instance $\mathbf{I}_i$. In this way, an attribute-abnormal motif instance $\mathbf{I}_j$ is generated along with the three attribute-abnormal nodes contained by $\mathbf{I}_j$. That is, during generating the attribute-abnormal motif instances, some attribute-abnormal nodes are also generated. In order to make the percentage of the attribute-abnormal nodes equal to ρ_1, it may be necessary to generate more attribute-abnormal nodes. In particular, for each candidate attribute-abnormal node v_j, 50 random nodes are selected from the dataset, and among the 50 random nodes, the node v_i whose attribute vector deviates the most from that of node v_j in terms of the euclidean distance between the attribute vectors is picked. Then the attribute vector of this candidate attribute-abnormal node v_j is replaced by the attribute vector of node v_i.

5.1.2 Baselines and Settings

The baselines include the shallow machine learning based anomaly detection methods and the deep learning based anomaly detection methods.

1) *AMEN* [16]:[3] It is designed for detecting the abnormal neighborhoods on attributed networks by means of integrating modularity and weighted attributed similarity. However, it cannot simultaneously discover abnormal nodes and subgraphs.
2) *ANOMALOUS* [21]:[4] It is a joint modeling method for anomaly detection on attributed networks which jointly conducts attribute selection and anomaly detection based on CUR decomposition and residual analysis.
3) *Dominant* [10]:[5] It is a graph convolutional autoencoder framework which consists of an attributed network encoder and two reconstruction decoders for structure and attribute respectively.
4) *Radar* [5]:[6] It is an anomaly detection method via residual analysis on attributed networks, which takes into account both of the residuals of attribute information and its coherence with network information.

Among the above four baselines, AMEN is only able to detect abnormal subgraphs while fails to discover abnormal nodes. To this end, the nodes contained by the detected abnormal subgraphs are regarded as the detected abnormal nodes with the abnormal scores of nodes being equal to the abnormal score of the subgraph. On the contrary, the remaining three baselines, namely ANOMALOUS, Dominant and Radar, cannot discover abnormal subgraphs. In order to obtain the abnormal motif instances for the ANOMALOUS, Dominant and Radar methods, for any triangle motif instance, if the three contained nodes are detected as abnormal, then it is regarded as the detected abnormal motif

2. The original datasets are from http://www.cs.cmu.edu/~webkb/, and we use the version released in https://linqs.soe.ucsc.edu/data.

3. https://github.com/phanein/amen.
4. http://people.virginia.edu/~jl6qk/code/ANOMALOUS.zip.
5. https://github.com/kaize0409/GCN_AnomalyDetection.
6. https://github.com/szumbrunn/radar-java.

TABLE 2
Comparison Results on Detecting Structure-Abnormal Nodes (SAN), Attribute-Abnormal Nodes (AAN), Structure-Abnormal Motif Instances (SAM) and Attribute-Abnormal Motif Instances (AAM) on the Scholat Dataset Under Different Anomaly Ratios

	Ratios	$\{\rho_1 = 1\%, \rho_2 = 3\%\}$								$\{\rho_1 = 1\%, \rho_2 = 5\%\}$							
	Measures	Precision@k				Recall@k				Precision@k				Recall@k			
		k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70
SAN	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	**0.025**	**0.020**	**0.017**	**0.014**	**0.100**	**0.100**	**0.100**	**0.100**	**0.050**	0.040	**0.050**	**0.043**	**0.200**	0.200	**0.300**	**0.300**
	Dominant	0.000	**0.020**	**0.017**	**0.014**	0.000	**0.100**	**0.100**	**0.100**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Radar	**0.025**	**0.020**	**0.017**	**0.014**	**0.100**	**0.100**	**0.100**	**0.100**	**0.050**	0.040	0.033	0.029	**0.200**	0.200	0.200	0.200
	HO-GAT	**0.025**	**0.020**	**0.017**	**0.014**	**0.100**	**0.100**	**0.100**	**0.100**	**0.050**	**0.060**	**0.050**	**0.043**	**0.200**	**0.300**	**0.300**	**0.300**
	Imp. (%)	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	0%	0%	50%	0%	0%
AAN	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Dominant	0.020	**0.080**	0.067	0.057	0.200	**0.400**	0.400	0.400	0.050	**0.080**	**0.083**	0.071	0.200	**0.400**	**0.500**	0.500
	Radar	0.000	0.020	0.017	0.014	0.000	0.100	0.100	0.100	0.025	0.020	0.033	0.029	0.100	0.100	0.200	0.200
	HO-GAT	**0.075**	0.060	**0.083**	**0.071**	**0.300**	0.300	**0.500**	**0.500**	**0.075**	0.060	**0.083**	**0.100**	**0.300**	0.300	**0.500**	**0.700**
	Imp. (%)	275%	−25%	24%	25%	50%	−25%	25%	25%	50%	−25%	0%	41%	50%	−25%	0%	40%
SAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.025	0.020	0.033	0.029	0.100	0.100	0.200	0.200	0.025	0.020	0.033	0.043	0.063	0.063	0.125	0.188
	Dominant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Radar	0.000	0.020	0.017	0.014	0.000	0.100	0.100	0.100	0.000	0.000	0.017	0.014	0.000	0.000	0.063	0.063
	HO-GAT	**0.100**	**0.100**	**0.083**	**0.071**	**0.400**	**0.500**	**0.500**	**0.500**	**0.175**	**0.180**	**0.167**	**0.157**	**0.438**	**0.563**	**0.625**	**0.688**
	Imp. (%)	300%	400%	152%	145%	300%	400%	150%	150%	600%	800%	406%	265%	595%	794%	400%	266%
AAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Dominant	0.075	0.100	0.100	0.100	0.300	0.500	0.600	0.700	0.225	0.220	0.183	0.171	0.563	0.688	0.688	0.750
	Radar	0.000	0.000	0.017	0.014	0.000	0.000	0.100	0.100	0.000	0.020	0.017	0.029	0.000	0.063	0.063	0.125
	HO-GAT	**0.225**	**0.180**	**0.167**	**0.143**	**0.900**	**0.900**	**1.000**	**1.000**	**0.400**	**0.320**	**0.267**	**0.229**	**1.000**	**1.000**	**1.000**	**1.000**
	Imp. (%)	200%	80%	67%	43%	200%	80%	67%	43%	78%	45%	46%	34%	78%	45%	45%	33%

The best results in each case are highlighted in bold. The improvement percentage is defined as $\text{Imp.}(\%) = \frac{\text{HO-GAT result-Best baseline result}}{\text{Best baseline result}} \times 100\%$.

instance with the abnormal score being the average abnormal score of the three nodes. For all the baselines, the default parameters suggested in the codes are adopted. In addition, because Dominant uses a trade-off parameter for telling apart structure-abnormal anomalies and attribute-abnormal anomalies, the structure-abnormal score and the attribute-abnormal score are calculated separately for each node so as to discover structure-abnormal nodes and attribute-abnormal nodes respectively. For the proposed HO-GAT method, the dimension of latent representations (i.e., l) is set to be 64, the impact of which will be analyzed in the parameter analysis.

5.1.3 Evaluation Measures

In our experiments, two widely used evaluation measures, namely Precision@k and Recall@k, are adopted for measuring the performance of the proposed method and baselines. The Precision@k measures the proportion of the ground-truth anomalies that an anomaly detection method detects in the top k ranked nodes according to the abnormal scores. Larger Precision@k value indicates better anomaly detection result. The Recall@k measures the proportion of the ground-truth anomalies that an anomaly detection method detects in the total number of ground-truth anomalies. Similarly, larger Recall@k value indicates better anomaly detection result.

5.2 Comparison Results

In this subsection, we will report and analyze the comparison results on the three types of attributed networks.

5.2.1 Comparison Results on Scholat

First of all, the comparison results on the Scholat dataset under different anomaly ratios will be reported and analyzed. In this experiment, two sets of anomaly ratios will be adopted, namely $\{\rho_1 = 1\%, \rho_2 = 3\%\}$ and $\{\rho_1 = 1\%, \rho_2 = 5\%\}$.

From Table 2, we can see that the proposed HO-GAT approach has no significant improvement compared with the four baselines on detecting abnormal nodes. That is, the Precision@k and Recall@k values obtained by HO-GAT are not significantly larger than those obtained by the four baselines in the cases of detecting structure-abnormal nodes (SAN) and attribute-abnormal nodes (AAN). However, it is clear that HO-GAT achieves significant improvements compared with the four baselines with the average improvements being larger than 100% in the cases of detecting structure-abnormal motif instances (SAM) and attribute-abnormal motif instances (AAM). For instance, for the Recall@60 value in the setting of $\{\rho_1 = 1\%, \rho_2 = 3\%\}$ on structure-abnormal motif instance detection (SAM), the Recall@60 value obtained by HO-GAT is 0.500, which results in 150% improvement compared with the second best result namely 0.200 obtained by ANOMALOUS. In addition, for some cases on attribute-abnormal motif instance detection

TABLE 3
Comparison Results on Detecting Structure-Abnormal Nodes (SAN), Attribute-Abnormal Nodes (AAN), Structure-Abnormal Motif Instances (SAM) and Attribute-Abnormal Motif Instances (AAM) on the AMiner Dataset Under Different Anomaly Ratios

	Ratios	$\{\rho_1 = 4\%, \rho_2 = 1.5\%\}$								$\{\rho_1 = 4\%, \rho_2 = 2.5\%\}$							
	Measures	Precision@k				Recall@k				Precision@k				Recall@k			
		k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70
SAN	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.075	0.060	0.050	0.043	0.081	0.081	0.081	0.081	0.025	0.020	0.033	0.029	0.029	0.029	0.059	0.059
	Dominant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Radar	0.075	0.060	0.050	0.043	0.081	0.081	0.081	0.081	0.050	0.040	0.033	0.029	0.059	0.059	0.059	0.059
	HO-GAT	**0.175**	**0.140**	**0.133**	**0.143**	**0.189**	**0.189**	**0.216**	**0.270**	**0.150**	**0.120**	**0.100**	**0.086**	**0.177**	**0.177**	**0.177**	**0.177**
	Imp. (%)	133%	133%	166%	233%	133%	133%	167%	233%	200%	200%	203%	197%	200%	200%	200%	200%
AAN	AMEN	0.025	0.020	0.017	0.014	0.027	0.027	0.027	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.000	0.000	0.017	0.014	0.000	0.000	0.027	0.027	0.000	0.000	0.017	0.014	0.000	0.000	0.029	0.029
	Dominant	0.025	0.020	0.017	0.014	0.027	0.027	0.027	0.027	0.025	0.020	0.017	0.014	0.029	0.029	0.029	0.029
	Radar	0.000	**0.060**	**0.050**	**0.043**	0.000	**0.081**	**0.081**	**0.081**	0.000	0.020	0.017	0.029	0.000	0.029	0.029	0.059
	HO-GAT	**0.075**	**0.060**	**0.050**	**0.043**	**0.081**	**0.081**	**0.081**	**0.081**	**0.100**	**0.080**	**0.083**	**0.071**	**0.118**	**0.118**	**0.147**	**0.147**
	Imp. (%)	200%	0%	0%	0%	200%	0%	0%	0%	300%	300%	388%	145%	307%	307%	407%	149%
SAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.025	0.020	0.033	0.043	0.026	0.026	0.051	0.077	0.050	0.080	0.067	0.057	0.031	0.062	0.062	0.062
	Dominant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Radar	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HO-GAT	**0.100**	**0.080**	**0.067**	**0.071**	**0.103**	**0.103**	**0.103**	**0.128**	**0.225**	**0.200**	**0.167**	**0.171**	**0.139**	**0.154**	**0.154**	**0.185**
	Imp. (%)	300%	300%	103%	65%	296%	296%	102%	66%	350%	150%	149%	200%	348%	148%	148%	198%
AAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.025**	**0.020**	0.017	0.014	**0.015**	**0.015**	0.015	0.015
	ANOMALOUS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Dominant	**0.050**	**0.040**	**0.033**	**0.029**	**0.051**	**0.051**	**0.051**	**0.051**	**0.025**	**0.020**	**0.033**	**0.043**	**0.015**	**0.015**	**0.031**	**0.046**
	Radar	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HO-GAT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Imp. (%)	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%	−100%

The best results in each case are highlighted in bold. The improvement percentage is defined as $\text{Imp.}(\%) = \frac{\text{HO-GAT result} - \text{Best baseline result}}{\text{Best baseline result}} \times 100\%$.

(AAM), the Recall@k value obtained by HO-GAT reaches 1.000. Overall, the above comparison result on the Scholat dataset has confirmed the superiority of the proposed HO-GAT method over the existing baselines, in particular on detecting the abnormal motif instances.

5.2.2 Comparison Results on AMiner

Similar to the case of Scholat, on the AMiner dataset, two sets of anomaly ratios will be adopted, namely $\{\rho_1 = 4\%, \rho_2 = 1.5\%\}$ and $\{\rho_1 = 4\%, \rho_2 = 2.5\%\}$. And the comparison results on detecting structure-abnormal nodes (SAN), attribute-abnormal nodes (AAN), structure-abnormal motif instances (SAM) and attribute-abnormal motif instances (AAM) are listed in Table 3.

From the results shown in Table 3, similar observation to that on the Scholat dataset can be obtained. That is, in the tasks of structure-abnormal node detection (SAN) and attribute-abnormal node detection (AAN), only marginal improvements have been achieved by HO-GAT compared with the baselines. However, when further analyzing the results in the tasks of structure-abnormal motif instance detection (SAM) and attribute-abnormal motif instance detection (AAM), the performance improvement is quite different from that on the Scholat dataset. That is, on the Scholat dataset, significant improvement has been achieved by HO-GAT compared with baselines in the tasks of structure-abnormal motif instance detection (SAM) and attribute-abnormal motif instance detection (AAN). However, on the AMiner dataset, the performance improvement in the task of structure-abnormal motif instance detection (SAM) is not that significant. And even worse, in the task of attribute-abnormal motif instance detection (AAM), most of the methods, including the proposed HO-GAT method, fail to detect attribute-abnormal motif instances, i.e., most of the values of Precision@k and Recall@k are 0.000. After investigating the AMiner dataset, it is found that the aforementioned anomaly generation method for the attribute-abnormal motif instances fails to generate abnormal attribute vectors of motif instances. This is because the average vector of the attribute vectors of the three nodes contained by the motif instance only changes slightly when replacing the attribute vectors of the three nodes by those in the other motif instance. That is to say, the generated attribute-abnormal motif instances do not deviate significantly from the other motif instances in terms of attribute proximity. Therefore, they cannot be detected by most of the methods.

In order to further investigate the performance of all the methods on attribute-abnormal motif instance detection on the AMiner dataset, we further add some random noises to the attribute vectors of the generated attribute-abnormal motif instances to generate the new testing dataset and report the

TABLE 4
Comparison Results on Attribute-Abnormal Motif Instance Detection on the AMiner Dataset Under Different Anomaly Ratios With Adding Random Attribute Noises

Ratios	$\{\rho_1 = 4\%, \rho_2 = 1.5\%\}$								$\{\rho_1 = 4\%, \rho_2 = 2.5\%\}$							
Measures	Precision@k				Recall@k				Precision@k				Recall@k			
	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70	k=40	k=50	k=60	k=70
AMEN	0.025	0.020	0.017	0.014	0.026	0.026	0.026	0.026	0.025	0.040	0.033	0.029	0.015	0.031	0.031	0.031
ANOMALOUS	0.050	0.040	0.033	0.029	0.051	0.051	0.051	0.051	0.200	0.220	0.183	0.157	0.123	0.169	0.169	0.169
Dominant	0.450	**0.720**	**0.650**	**0.557**	0.462	**0.923**	**1.000**	**1.000**	0.300	0.380	0.517	0.714	0.185	0.292	0.477	0.769
Radar	0.300	0.300	0.250	0.214	0.308	0.385	0.385	0.385	0.300	0.320	0.383	0.371	0.185	0.246	0.354	0.400
HO-GAT	**0.675**	0.680	0.633	0.543	**0.692**	0.872	0.974	0.974	**0.875**	**0.840**	**0.783**	**0.757**	**0.538**	**0.646**	**0.723**	**0.815**
Imp. (%)	50%	−6%	−3%	−3%	50%	−6%	−3%	−3%	192%	121%	51%	6%	191%	121%	52%	6%

The best results in each case are highlighted in bold. The improvement percentage is defined as $\text{Imp.}(\%) = \frac{\text{HO-GAT result} - \text{Best baseline result}}{\text{Best baseline result}} \times 100\%$.

comparison results in Table 4. From Table 4, we can see that the performances of all the methods on attribute-abnormal motif instance detection on the AMiner dataset have been improved compared with those in Table 3, and in most cases, the proposed HO-GAT method has achieved relatively large improvement compared with the baselines. This analysis has further confirmed the superiority of the proposed HO-GAT method on simultaneously detecting structure/attribute-abnormal nodes and motif instances.

5.2.3 Comparison Results on Four WebKB Datasets

In this subsection, we will report and compare the performance on the four WebKB datasets. Due to the space limit, we only report the results under the same set of anomaly ratios $\{\rho_1 = 5\%, \rho_2 = 5\%\}$ on the four WebKB datasets. In addition, only the values of Precision@50 and Recall@50 are reported in the tasks of structure/attribute-abnormal node detection, while in the tasks of structure/attribute-abnormal motif instance detection, only the values of Precision@k and

TABLE 5
Comparison Results on Detecting Structure-Abnormal Nodes (SAN), Attribute-Abnormal Nodes (AAN), Structure-Abnormal Motif Instances (SAM) and Attribute-Abnormal Motif Instances (AAM) on the Four WebKB Datasets Under the Same Anomaly Ratios $\{\rho_1 = 5\%, \rho_2 = 5\%\}$

Datasets		Cornell		Texas		Washington		Wisconsin	
	Measures	Precision@50	Recall@50	Precision@50	Recall@50	Precision@50	Recall@50	Precision@50	Recall@50
SAN	AMEN	0.020	0.200	0.000	0.000	0.000	0.000	0.020	0.167
	ANOMALOUS	**0.060**	**0.600**	0.060	0.750	**0.060**	**0.600**	**0.600**	**0.500**
	Dominant	0.020	0.200	0.000	0.000	0.000	0.000	0.000	0.000
	Radar	0.020	0.200	0.000	0.000	0.040	0.400	0.040	0.333
	HO-GAT	0.000	0.000	**0.080**	**1.000**	0.000	0.000	0.060	**0.500**
	Imp. (%)	−100%	−100%	33%	33%	−100%	−100%	−90%	0%
	Measures	Precision@50	Recall@50	Precision@50	Recall@50	Precision@50	Recall@50	Precision@50	Recall@50
AAN	AMEN	0.020	0.200	0.040	0.500	0.020	0.200	0.020	0.167
	ANOMALOUS	0.040	0.400	0.000	0.000	0.020	0.200	0.000	0.000
	Dominant	0.020	0.200	**0.060**	**0.750**	0.040	0.400	0.080	0.667
	Radar	0.040	0.400	0.040	0.500	0.040	0.400	0.060	0.500
	HO-GAT	**0.080**	**1.000**	**0.060**	**0.750**	**0.080**	**0.800**	**0.120**	**1.000**
	Imp. (%)	100%	150%	0%	0%	100%	100%	50%	50%
	Measures	Precision@10	Recall@10	Precision@50	Recall@50	Precision@10	Recall@10	Precision@20	Recall@20
SAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.000	0.000	**0.060**	**1.000**	**0.200**	**0.400**	0.100	0.333
	Dominant	0.000	0.000	**0.060**	**1.000**	0.000	0.000	0.000	0.000
	Radar	**0.100**	**0.200**	**0.060**	**1.000**	0.000	0.000	0.050	0.167
	HO-GAT	**0.100**	**0.200**	**0.060**	**1.000**	**0.200**	**0.400**	**0.150**	**0.500**
	Imp. (%)	0%	0%	0%	0%	0%	0%	50%	50%
	Measures	Precision@10	Recall@10	Precision@50	Recall@50	Precision@10	Recall@10	Precision@20	Recall@20
AAM	AMEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ANOMALOUS	0.000	0.000	**0.060**	**1.000**	0.000	0.000	0.000	0.000
	Dominant	0.200	0.400	**0.060**	**1.000**	0.000	0.000	0.150	0.500
	Radar	0.100	0.200	0.020	0.333	0.100	0.200	0.150	0.500
	HO-GAT	**0.400**	**0.800**	**0.060**	**1.000**	**0.400**	**0.800**	**0.300**	**1.000**
	Imp. (%)	100%	100%	0%	0%	300%	300%	100%	100%

The best results in each case are highlighted in bold. The improvement percentage is defined as $\text{Imp.}(\%) = \frac{\text{HO-GAT result} - \text{Best baseline result}}{\text{Best baseline result}} \times 100\%$.

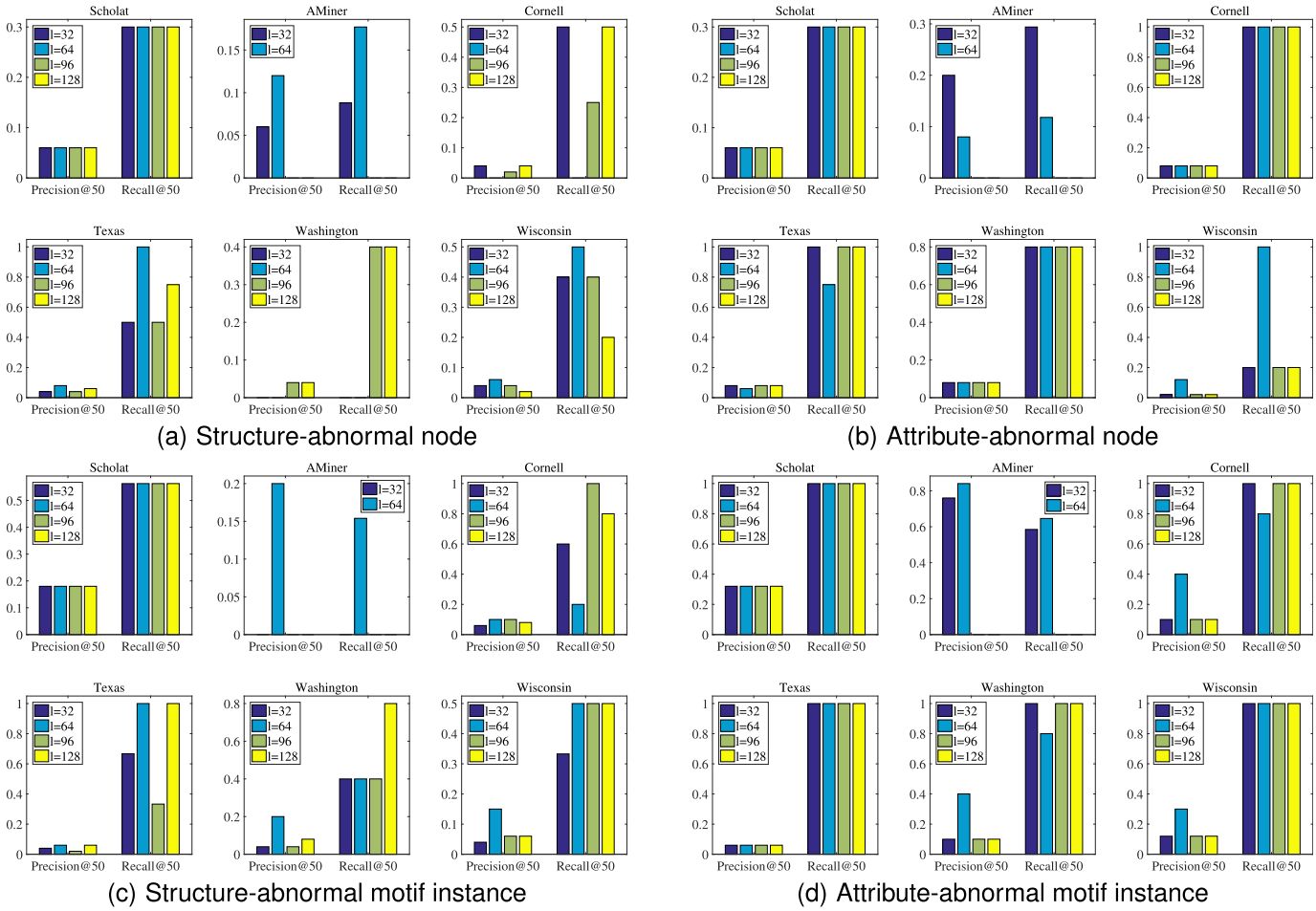


Fig. 2. Parameter analysis of l on the six datasets: The anomaly detection performances when using different dimensions of the latent representations (i.e., l). On the AMiner dataset, it is not suitable to use such large dimension (i.e., $l = 96$ and $l = 128$) due to the relatively small original dimension and hence no result is generated.

Recall@k with one randomly selected k are reported, i.e., $k = 10, 50, 10, 20$ for Cornell, Texas, Washington and Wisconsin respectively. The comparison results on detecting the four different types of anomalies are listed in Table 5. From the table, it is clear that, similar to the case of the Scholat and AMiner datasets, the improvements on detecting structure/attribute-abnormal nodes are not so significant. However, HO-GAT achieves significant improvements on detecting structure/attribute-abnormal motif instances. Therefore, similar conclusion can be drawn that HO-GAT is superior in detecting structure/attribute-abnormal motif instances while being comparable in detecting structure/attribute-abnormal nodes.

5.3 Parameter Analysis

5.3.1 *Latent Dimension l*

In this subsection, we will analyze how the dimension of the latent representations (i.e., l) affects the performance of the proposed HO-GAT method by reporting the performance when setting $l = 32, 64, 96, 128$. In this experiment, the anomaly ratios for Scholat and AMiner are set as $\{\rho_1 = 1\%, \rho_2 = 5\%\}$ and $\{\rho_1 = 4\%, \rho_2 = 2.5\%\}$ respectively. And the anomaly ratios for the four WebKB datasets are set as $\{\rho_1 = 5\%, \rho_2 = 5\%\}$. The results are plotted in Fig. 2. From the table, we can see that in most cases, the performance of HO-GAT is relatively insensitive to the value of dimension l. In particular, for the Scholat dataset, setting l to 32, 64, 96 and 128 generates completely the same performance. However, in some cases, e.g., structure-abnormal motif instance detection on Cornell, the performance varies when using different l. Overall, from the table, it can be seen that setting the dimension of latent representations as 64 generates good results in most cases, which is therefore adopted in our experiments.

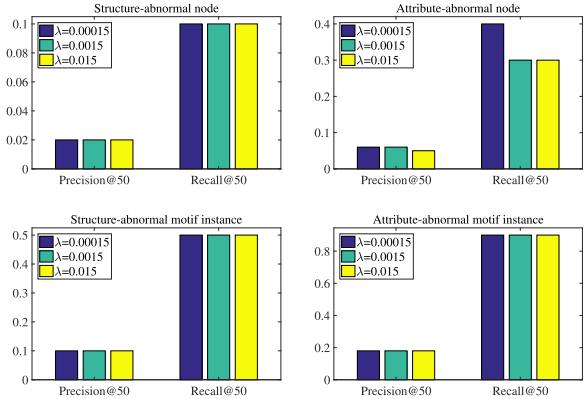
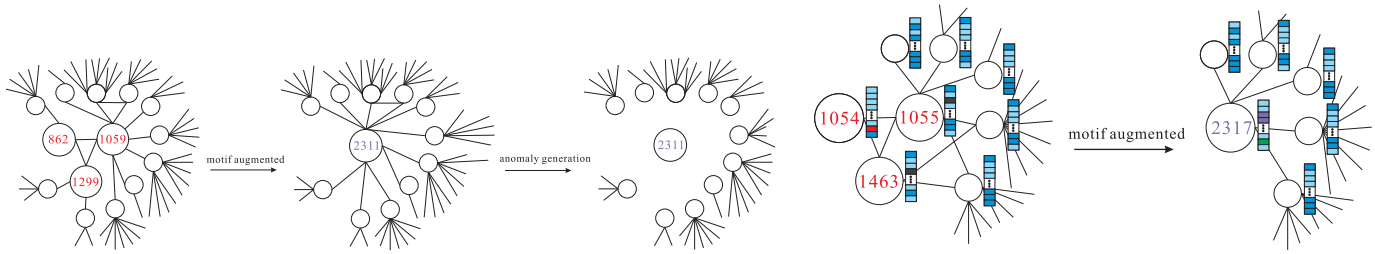


Fig. 3. Parameter analysis of λ on the Scholat dataset with the anomaly ratio $\{\rho_1 = 1\%, \rho_2 = 5\%\}$: The anomaly detection performances when using different λ values.



(a) Structure-abnormal motif instance (b) Attribute-abnormal motif instance

Fig. 4. Case study on the Scholat dataset with the anomaly ratio $\{\rho_1 = 1\%, \rho_2 = 5\%\}$: Illustration of detecting typical structure-abnormal motif instance and attribute-abnormal motif instance.

5.3.2 Regularization Parameter λ

In this subsection, we will analyze how the regularization parameter λ affects the performance of the proposed HO-GAT method when setting $\lambda = 0.00015$, $\lambda = 0.0015$ and $\lambda = 0.015$ respectively. Due to the space limit, only the results on the Scholat dataset with the anomaly ratio $\{\rho_1 = 1\%, \rho_2 = 5\%\}$ are reported. As shown in Fig. 3, the performance of HO-GAT is very stable in most cases. In particular, in the case of structure-abnormal node detection, structure-abnormal motif instance detection and attribute-abnormal motif instance detection, setting different values of λ generates the same results. In the case of attribute-abnormal node detection, the performance varies slightly when using different λ values. By considering the effect of regularization terms, λ is set to be the median value we have tried, i.e., 0.0015, which is a default value widely adopted in the literature.

5.4 Case Study

In this subsection, case study will be conducted on the Scholat dataset with the anomaly ratio $\{\rho_1 = 1\%, \rho_2 = 5\%\}$. Due to the space limit, we only investigate the cases of structure-abnormal motif instance detection and attribute-abnormal motif instance detection. Some typical cases are plotted in Fig. 4. Fig. 4a shows a topological network structure diagram, wherein the motif instance 2311 is composed of nodes 862, 1059 and 1299. During the anomaly generation process, we remove the edges between the motif instance 2311 and other nodes to get the isolated motif instance. In the proposed method, the motif instance 2311 has the largest structure-abnormal score among all the motif instances. Therefore, it can be correctly identified as a structure-abnormal motif instance. Fig. 4b shows an attribute network diagram after the attribute anomaly injection, wherein the motif instance 2317 is composed of nodes 1054, 1055 and 1463. In the figure, the attribute of the node is represented by the color block. We can observe that the attributes of nodes 1054, 1055 and 1463 are significantly different from those of their neighbor nodes. And after the motif augmented process, the attributes of the motif instance 2317 still remain different from the those of the other nodes and motif instances. In the proposed method, the motif instance 2317 has the largest attribute-abnormal score among all the motif instances. Therefore, it can be correctly identified as an attribute-abnormal motif instance.

6 Conclusion

In this paper, we have defined a new problem of hybrid-order anomaly detection on attributed networks, which aims to detect not only structure/attribute-abnormal nodes but also stucture/attribute-abnormal motif instances. To this end, we have developed a new deep learning model called Hybrid-Order Graph Attention Network (HO-GAT). First of all, a motif-augmented attributed network is constructed to model the hybrid-order structure and attribute information. Then the motif-augmented attributed network is fed into a hybrid-order attributed network encoder to generate the latent representations of the motif-augmented nodes. A novel hybrid-order self-attention mechanism is designed to model the complex mutual influence between nodes and motif instances in terms of topological structure or attribute proximity. After obtaining the latent representations of the motif-augmented nodes, a hybrid-order attributed network decoder is designed to reconstruct the attribute information of the nodes and motif instances, and the hybrid-order topological structure among nodes and motif instances. Finally, the reconstruction errors are utilized as the abnormal score of nodes and motif instances respectively. Extensive experiments conducted on real-world datasets have confirmed the effectiveness of the HO-GAT method.

References

[1] J. Liang, P. Jacobs, J. Sun, and S. Parthasarathy, "Semi-supervised embedding in attributed networks with outliers," in *Proc. SIAM Int. Conf. Data Mining*, 2018, pp. 153–161.

[2] L. Zheng, Z. Li, J. Li, Z. Li, and J. Gao, "AddGraph: Anomaly detection in dynamic graph using attention-based temporal GCN," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 4419–4425.

[3] S. Bandyopadhyay, N. Lokesh, and M. N. Murty, "Outlier aware network embedding for attributed networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 12–19.

[4] H. Peng *et al.*, "LIME: Low-cost incremental learning for dynamic heterogeneous information networks," *IEEE Trans. Comput.*, early access, Feb. 11, 2021, doi: 10.1109/TC.2021.3057082.

[5] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 2152–2158.

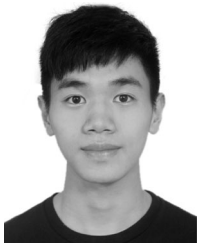
[6] L. Guti érrez-Gómez, A. Bovet, and J.-C. Delvenne, "Multi-scale anomaly detection on attributed networks," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 678–685.

[7] K. Ding, J. Li, and H. Liu, "Interactive anomaly detection on attributed networks," in *Proc. Int. Conf. Web Search Data Mining*, 2019, pp. 357–365.

[8] X. Teng, Y.-R. Lin, and X. Wen, "Anomaly detection in dynamic networks using multi-view time-series hypersphere learning," in *Proc. Int. Conf. Inf. Knowl. Manage.*, 2017, pp. 827–836.

[9] M. R. Gahrooei and K. Paynabar, "Change detection in a dynamic stream of attributed networks," *J. Qual. Technol.*, vol. 50, no. 4, pp. 418–430, 2018.
[10] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *Proc. SIAM Int. Conf. Data Mining*, 2019, pp. 594–602.
[11] Y. Li, X. Huang, J. Li, M. Du, and N. Zou, "SpecAE: Spectral autoencoder for anomaly detection in attributed networks," in *Proc. Int. Conf. Inf. Knowl. Manage*, 2019, pp. 2233–2236.
[12] Z. Peng, M. Luo, J. Li, L. Xue, and Q. Zheng, "A deep multi-view framework for anomaly detection on attributed networks," *IEEE Trans. Knowl. Data Eng.*, early access, Aug. 7, 2020, doi: 10.1109/TKDE.2020.3015098.
[13] M. Gupta, J. Gao, X. Yan, H. Cam, and J. Han, "On detecting association-based clique outliers in heterogeneous information networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2013, pp. 108–115.
[14] H. Zhuang *et al.*, "Mining query-based subnetwork outliers in heterogeneous information networks," in *Proc. Int. Conf. Data Mining*, 2014, pp. 1127–1132.
[15] Y. Xiong, Y. Zhu, P. S. Yu, and J. Pei, "Towards cohesive anomaly mining," in *Proc. AAAI Conf. Artif. Intell.*, 2013, pp. 984–990.
[16] B. Perozzi and L. Akoglu, "Scalable anomaly ranking of attributed neighborhoods," in *Proc. SIAM Int. Conf. Data Mining*, 2016, pp. 207–215.
[17] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015.
[18] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.
[19] R. Milo, S. S.-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
[20] N. Liu, X. Huang, and X. Hu, "Accelerated local anomaly detection via resolving attributed networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 2337–2343.
[21] Z. Peng, M. Luo, J. Li, H. Liu, and Q. Zheng, "ANOMALOUS: A joint modeling approach for anomaly detection on attributed networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 3513–3519.
[22] L. Xue, Y. Chen, M. Luo, Z. Peng, and J. Liu, "An anomaly detection framework for time-evolving attributed networks," *Neurocomputing*, vol. 407, pp. 39–49, 2020.
[23] Z. Liu, Y. Dou, P. S. Yu, Y. Deng, and H. Peng, "Alleviating the inconsistency problem of applying graph neural network to fraud detection," in *Proc. Special Int. Group Inf. Retrieval*, 2020, pp. 1569–1572.
[24] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph neural network-based fraud detectors against camouflaged fraudsters," in *Proc. Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 315–324.
[25] Z. Chen, B. Liu, M. Wang, P. Dai, J. Lv, and L. Bo, "Generative adversarial attributed network anomaly detection," in *Proc. Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1989–1992.
[26] R. Yu, X. He, and Y. Liu, "GLAD: Group anomaly detection in social media analysis," *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 2, pp. 18:1–18:22, 2015.
[27] H. Peng, J. Li, Q. Gong, Y. Ning, S. Wang, and L. He, "Motif-matching based subgraph-level attentional convolutional network for graph classification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 5387–5394.
[28] R. Milo *et al.*, "Superfamilies of evolved and designed networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, 2004.
[29] L. Huang, C.-D. Wang, and H.-Y. Chao, "HM-Modularity: A harmonic motif modularity approach for multi-layer network community detection," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2520–2533, Jun. 2021.
[30] W. Lin, X. Xiao, X. Xie, and X. Li, "Network motif discovery: A GPU approach," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 3, pp. 513–528, Mar. 2017.
[31] C. E. Tsourakakis, J. Pachocki, and M. Mitzenmacher, "Scalable motif-aware graph clustering," in *Proc. Int. Conf. World Wide Web*, 2017, pp. 1451–1460.
[32] P.-Z. Li, L. Huang, C.-D. Wang, and J.-H. Lai, "EdMot: An edge enhancement approach for motif-aware community detection," in *Proc. Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 479–487.
[33] L. Huang, H.-Y. Chao, and G. Xie, "MuMod: A micro-unit connection approach for hybrid-order community detection," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 107–114.
[34] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs," *Bioinformatics*, vol. 20, no. 11, pp. 1746–1758, 2004.
[35] F. Schreiber and H. Schwöbbermeyer, "MAVisto: A tool for the exploration of network motifs," *Bioinformatics*, vol. 21, no. 17, pp. 3572–3574, 2005.
[36] S. Wernicke and F. Rasche, "FANMOD: A tool for fast network motif detection," *Bioinformatics*, vol. 22, no. 9, pp. 1152–1153, 2006.
[37] J. A. Grochow and M. Kellis, "Network motif discovery using subgraph enumeration and symmetry-breaking," in *Proc. 11th Annu. Int. Conf. Res. Comput. Mol. Biol.*, 2007, pp. 92–106.
[38] Z. R. M. Kashani *et al.*, "Kavosh: A new algorithm for finding network motifs," *BMC Bioinf.*, vol. 10, no. 1, 2009, Art. no. 318.
[39] Y. Zhu *et al.*, "Exploiting a novel algorithm and GPUs to break the ten quadrillion pairwise comparisons barrier for time series motifs and joins," *Knowl. Inf. Syst.*, vol. 54, no. 1, pp. 203–236, 2018.
[40] S. Schbath, V. Lacroix, and M. Sagot, "Assessing the exceptionality of coloured motifs in networks," *EURASIP J. Bioinf. Syst. Biol.*, vol. 2009, pp. 1–9, 2009.
[41] G. Micale, R. Giugno, A. Ferro, M. Mongiovì, D. E. Shasha, and A. Pulvirenti, "Fast analytical methods for finding significant labeled graph motifs," *Data Min. Knowl. Discov.*, vol. 32, no. 2, pp. 504–531, 2018.
[42] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. 6th Int. Conf. Learn. Representations, Vancouver, BC, Canada*, 2018, pp. 1–12.
[43] H. Tong and C. Lin, "Non-negative residual matrix factorization with application to graph anomaly detection," in *Proc. SIAM Int. Conf. Data Mining*, 2011, pp. 143–153.
[44] X. Song, M. Wu, C. M. Jermaine, and S. Ranka, "Conditional anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 631–645, May 2007.

Ling Huang (Member, IEEE) received the PhD degree in computer science from Sun Yat-sen University, Guangzhou, in 2020. She joined South China Agricultural University in 2020 as an associate professor. She has authored or coauthored more than ten papers in international journals and conferences, including *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Industrial Informatics*, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Intelligent Systems and Technology*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *Pattern Recognition*, KDD, AAAI, IJCAI, and ICDM. Her research focuses on data mining.

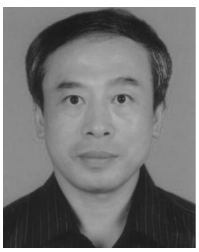


Ye Zhu is currently working toward the undergraduation degree with South China Agricultural University. His research interests include algorithm design, implementation, and data mining.

Yuefang Gao received the PhD degree in computer science from the South China University of Technology, Guangzhou, China, in 2009. From March 2016 to April 2017, she was a visiting scholar with the University of Sydney. She joined South China Agricultural University in 2003, where she is currently an associate professor with the School of Mathematics and Informatics. She has authored or coauthored more than ten scientific papers in international journals and conferences, including *IEEE Transactions on Circuits and Systems for Video Technology*, *Pattern Recognition*, and ACM MM. Her research interests include computer vision and machine learning.

Tuo Liu is currently working toward the undergraduation degree in computer science with Sun Yat-Sen University. His research interests include algorithm design, implementation, and data mining.

Chao Chang received the BS degree from Fujian Normal University, Fuzhou, China, in 2016 and the MS degree from South China Normal University, Guangzhou, China, in 2019. She is currently working toward the PhD degree in software engineering with South China Normal University. Her research interests include knowledge graph, social network services, and big data analysis for education.



Caixing Liu received the BS degree from Computer Science College, Nanjing University, in 1984. He is currently a professor with the College of Mathematics and Informatics, South China Agriculture University, Guangzhou, China. His research interests include artificial intelligent and smart agriculture.

Yong Tang received the BS and MSc degrees in computer science from Wuhan University in 1985 and 1990, respectively, and PhD degree in computer science from the University of Science and Technology of China in 2001. He is the founder of SCHOLAT, a kind of scholar social network. He is currently a professor and the dean of the School of Computer Science, South China Normal University. He was the vice dean of the School of Information of Science and Technology, Sun Yat-Sen University. He has authored or coauthored more than 200 papers and books. He has been supervising more than 40 PhD students since 2003 and more than 100 master students since 1996. His research interests include data and knowledge engineering, social networking, and collaborative computing. He is currently the director of the technical committee on collaborative computing, China Computer Federation and the executive vice president of Guangdong Computer Academy.

Chang-Dong Wang (Member, IEEE) received the PhD degree in computer science from Sun Yat-sen University, Guangzhou, China, in 2013. From January 2012 to November 2012, he was a visiting student with the University of Illinois at Chicago. He joined Sun Yat-sen University in 2013, where he is currently an associate professor with the School of Computer Science and Engineering. He has authored or coauthored more than 80 scientific papers in international journals and conferences, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Neural Networks and Learning Systems*, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Intelligent Systems and Technology*, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *IEEE Transactions on Industrial Informatics*, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* KDD, AAAI, IJCAI, CVPR, ICDM, CIKM, and SDM. His research interests include machine learning and data mining. He is currently an associate editor for *Journal of Artificial Intelligence Research*. He was the recipient of the Honorable Mention for Best Research Paper Awards for ICDM 2010 Paper, 2012 Microsoft Research Fellowship Nomination Award, and the 2015 Chinese Association for Artificial Intelligence Outstanding Dissertation.