Journal of King Saud University - Computer and Information Sciences 36 (2024) 102110

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

Review Article

Detection of misbehaving individuals in social networks using overlapping communities and machine learning

Wejdan Alshlahy [a], Delel Rhouma [a,b,*]

[a] Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
[b] Modeling of Automated Reasoning Systems Research Laboratory LR17ES05, Higher Institute of Computer Science and Telecom, University of Sousse, Sousse, Tunisia

ARTICLE INFO

Keywords:
Social network
Graph mining
Overlapping community
Contextual anomaly
Structural anomaly
Anomaly detection
Machine learning
Deep learning

ABSTRACT

Detecting misbehavior in social networks is essential for maintaining trust and reliability in online communities. Traditional methods of identification often rely on individual attributes or structural network properties, which may overlook subtle or complex misbehavior patterns. This paper introduces a novel approach called OCMLMD that leverages network overlapping community structure and machine learning techniques to detect misbehavior. Our method combines graph-based analyses of network topology with state-of-theart machine learning algorithms to identify suspicious behavior indicative of misbehavior. Specifically, we target nodes that belong to multiple communities or exhibit weak connections within their community, utilizing a novel metric for selecting overlapping nodes. Additionally, we develop a machine learning model trained on relevant attributes extracted from social network data to detect misbehavior accurately. Extensive experiments on synthetic and real-world social network datasets demonstrate the superior performance of OCMLMD compared to baseline methods. Overall, our proposed approach offers a promising solution to the challenge of detecting misbehavior in social networks.

1. Introduction

Social networks have become integral parts of modern society that facilitate communication, collaboration, and information sharing among individuals worldwide. However, the openness and anonymity of social networks also present challenges, particularly concerning the detection of misbehavior and malicious activities, such as spreading misinformation, harassment, spamming, and phishing (Shu et al., 2017; Ahmed et al., 2020; Lin et al., 2022). Detecting and mitigating such activities is crucial for maintaining trust, security, and integrity within social network platforms.

Social networks can be effectively represented as graphs: nodes represent

individuals, while edges represent interactions. Social communities are naturally formed within social networks (Keyvanpour et al., 2020) consisting of individuals who interact frequently and have much in common, resulting in fewer relationships being formed between communities (Ghoshal and Ramakrishnan, 2017; Ghoshal et al., 2019). In recent years, there has been growing scholarly interest in leveraging attributed graphs to identify misbehavior within complex networks. Unlike plain graphs that rely solely on topological information to detect misbehavior in traditional networks, attributed graphs provide a rich source of attribute information for each node or edge. In social networks, for example, users engage in diverse social activities and possess abundant personal information.

Integrating topological structures and attribute information for detecting misbehavior in social networks is quite challenging. Traditional methods often rely on either structural or attribute information, which may not fully capture the complexity of misbehavior. Structure-based approaches may struggle in contexts where behaviors are closely tied to attributes (Akoglu et al., 2010), while attribute-based approaches may falter when misbehavior extends beyond node attributes (Xu et al., 2018).

To address these limitations, recent research has explored hybrid approaches that combine both structural and attribute information to enhance misbehavior detection in social networks. These hybrid methods aim to leverage the strengths of both structural and attribute-based techniques while mitigating their respective weaknesses. However, designing effective integration strategies remains an ongoing research area.

* Corresponding author at: Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia.
E-mail addresses: 411200390@qu.edu.sa (W. Alshlahy), d.rhouma@qu.edu.sa (D. Rhouma).

https://doi.org/10.1016/j.jksuci.2024.102110
Received 11 February 2024; Received in revised form 19 April 2024; Accepted 23 June 2024
Available online 2 July 2024

Misbehavior in social networks is typically characterized by users who either belong to multiple communities or have weak relationships within their communities (Xu et al., 2007). While several studies have been conducted to identify misbehavior in social networks (Akoglu et al., 2009; Gupta et al., 2013; Hu et al., 2016; Win and Lynn, 2018; Helling et al., 2018), many of these methods run into scalability issues when applied to large network datasets.

This paper proposes a novel approach to detecting misbehavior in social networks. It combines graph-based analyses of social networks with state-of the-art machine learning algorithms. By focusing specifically on overlapping nodes and pendant nodes, which are likelier to represent misbehavior, the search space is significantly reduced.

The primary contributions of this paper are outlined as follows:

- Proposing a new metric for selecting overlapping nodes in social networks, which enables the identification of suspicious relationships between individuals and detect misbehavior that spans multiple communities. To the best of our knowledge, this is the first study to achieve this.
- The proposed approach is designed to efficiently handle large-scale network data and address the scalability challenge commonly encountered for misbehavior detection methods within social networks.
- Developing an unsupervised machine learning model trained on relevant attributes extracted from social network data to effectively detect misbehavior.
- Conducting extensive experiments on synthetic and real-world social network datasets to evaluate our approach's effectiveness in terms of detecting misbehavior, compare our method against baseline approaches and demonstrate its superior performance in terms of ROC-AUC, precision, recall, and scalability.

We have organized the rest of the paper into sections as follows: Section 2 presents a review of related literature, Section 3 introduces preliminary material, Section 4 demonstrates how structure and individual attributes can be used to detect misbehavior, Section 5 presents the experimental results, and Section 6 concludes the paper.

2. Related work

This section provides an overview of existing literature on detecting misbehavior in social networks. It categorizes methods for detecting misbehavior into three main approaches:

2.1. Structure-based approaches

These techniques analyze network structure, focusing on properties such as node connections, edge formations, and a graph's characteristics. Examples include spectral analysis of the adjacency matrix and its variations (T. Ide, H. Kashima, Eigenspace-based anomaly detection in computer systems, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004; Miller et al., 2010), defining structural similarity measures, and applying clustering methods for misbehavior detection (Xu et al., 2007; Peng et al., 2018). Additionally, it can utilize statistical features computed based on graph structure (Akoglu et al., 2010; Ding et al., 2012; Hooi et al., 2016). However, these approaches may miss information from node attributes, leading to a high rate of false detection, particularly for joint-type misbehavior.

2.2. Attribute-based approaches

These approaches involve analyzing nodes' attribute information to detect contextual patterns and misbehavior. Various methods have been proposed to analyze the distribution of node attributes for identifying misbehavior (Breunig et al., 2000; Liu et al., 2012; Bandyopadhyay et al., 2019). For instance, one common technique involves employing the k nearest neighbor algorithm on node attributes to identify isolated nodes (Angiulli and Pizzuti, 2002). In recent years, autoencoders have gained popularity for their effectiveness in misbehavior detection (Sakurada and Yairi, 2014). Autoencoders leverage neural networks to reduce the dimensionality of data, aiming to capture the principal components while excluding sparse misbehavior (Hinton and Salakhutdinov, 2006). By compressing data, autoencoders generate representations that aim to reconstruct normal data. Monitoring the reconstruction loss allows for the identification of misbehavior within the normal data. However, these techniques are often limited in their capacity to identify misbehavior beyond contextual patterns (Xu et al., 2018).

2.3. Structural and attribute-based approaches

These approaches integrate both network structure and node attributes, considering connections, topology, and node-specific attributes. Techniques falling under this approach include clustering (Bojchevski, 2018; Perozzi and Akoglu, 2014), human-expert interaction (Ding et al., 2019), and group merging (Zhu and Zhu, 2020). Autoencoders integrated with graph neural networks (GNNs) offer a comprehensive approach to combining graph structures and node attributes to detect misbehavior, as demonstrated in various studies (Ding et al., 2019; Fan et al., 2020; Kipf and Welling, 1611; Bandyopadhyay et al., 2020). These methods typically assess misbehavior by evaluating the reconstruction loss of node attributes or links. However, instead of reconstructing the entire neighborhood for misbehavior, they utilize reconstruction errors, which are sometimes supplemented by estimating Gaussian mixture density (Li et al., 2019). Some approaches consider nodes from multiple perspectives—a node's misbehavior status may vary across different views, reflecting attributes from diverse perspectives. To handle such multiview scenarios, multiple GNNs are often employed for misbehavior detection (Liu et al., 2212; Peng et al., 2020; Sheng et al., 2019; Wu et al., 2014; Wu et al., 2013). Additionally, advanced techniques such as self-supervised learning (Hu et al., 2304; Huang et al., 2022; Jin et al., 2021; Liu et al., 2021; Xu et al., 2022; Zhou et al., 2023) and reinforcement learning (Ding et al., 2019; Langford and Zhang, 2007; Morales et al., 2021) have recently emerged as effective strategies in this domain. These methods leverage both graph structure and node attributes, potentially leading to optimal misbehavior detection in social networks.

Despite these advancements, current misbehavior detection methods face limitations, including challenges with attribute-rich networks, identifying relevant features, and scalability issues as social networks grow in size and complexity. These limitations contribute to the subpar performance of many existing misbehavior detection methods for social networks.

3. Preliminaries

In this section, we introduce fundamental concepts and definitions used in our model.

Given a social network represented as an attributed graph G, we will represent it as a graph structure (V,E) and attribute information A, characterized as follows:

- Each individual is represented by a node $v \in V$ and is connected by an edge $(v_i,v_j) \in E$, where $v_i \neq v_j$, for all $v_i,v_j \in V$.
- Each individual is described by attribute information $A = (a_1,a_2,\ldots,a_r)$ in an r-dimensional data space, where $a_i \in \mathrm{R}$ for $i = 1,2,\ldots,r$.
- Community structure C in social networks involves groups of individuals $C = \{c_1,c_2,\ldots,c_k\}$, allowing robust communication within

the same community, unlike weak connections between communities (Ghoshal et al., 2021).

3.1. Related definitions

Definition 1. (*Overlapping communities: The overlapping communities refer to a structural pattern where nodes can simultaneously belong to multiple communities. Unlike traditional disjoint communities, where nodes belong to only one community, overlapping communities enable nodes to have memberships in more than one community, reflecting the complex nature of social networks.*) Example 1. *For a social network represented as an attributed graph G a possible division of a graph structure (V,E) into disjoint or overlapping communities is illustrated in* Fig. 1.

Definition 2. (*Overlapping nodes: The overlapping nodes refer to those nodes that belong to multiple communities within a graph structure (Ghoshal et al., 2021).*) Example 2. Fig. 1(*B*) *illustrates the division of network G into overlapping communities, each consisting of a set of overlapping nodes.*

Definition 3. (*Index of community connectivity: The index of community connectivity, denoted as IC, is a metric used to quantify the cohesion and separation of communities within a social network. It is designed to find a partition of a graph structure into communities that are both internally cohesive (high compactness) and externally separated (low separability) (Rhouma and Romdhane, 2014). Mathematically, the IC of a community c is calculated as:*)

$$IC(c) = \frac{Com(c) - Sep(c)}{\sqrt{Com(c) + Sep(c)}} \tag{1}$$

where $Com(c)$ represents compactness, which is the number of edges connecting individuals in a community c, and $Sep(c)$ represents separability, which is the number of edges outside a community c.

Definition 4. (*Overlapping node connectivity: The overlapping node connectivity, denoted as OvpCon, measures the strength of the connection between a node and its associated communities within a network. It is given by:*)

$$OvpCon(v) = \frac{\deg(v)}{\frac{\sum_{k=1}^{c} nb(v_k)}{|C|}} \tag{2}$$

where $deg(v)$ denotes a degree of an overlapping node v, $nb(v)$ represents the number of neighboring nodes of v that belong to each community c, and $|C|$ signifies the number of overlapping node communities.

Definition 5. (*Overlapping node score: The overlapping node score, denoted as OvpScore, assesses the significance of a node's presence in multiple communities within a network, expressed by the equation:*)

$$OvpScore(v) = 1 - CC(v)^{*}OvpCon(v) \tag{3}$$

where $CC(v)$ denotes the clustering coefficient of an overlapping node v, which measures the degree of interconnectedness among its neighboring nodes, and $OvpCon(v)$ stands for overlapping node connectivity.

Definition 6. (*Mutual information: The mutual information, denoted as MI, is a statistical measure quantifying the information shared between two random variables. In relevant attributes selection, it assesses the dependency between an attribute and a target variable, commonly employed to evaluate attribute relevance in machine learning tasks (Bishop, 2006), and is calculated using Equation (4):*)

$$MI(X;Y) = H(X) - H(X|Y) \tag{4}$$

where $H(X)$ represents the entropy of attribute X and $H(X|Y)$ is the conditional entropy of attribute X given Y.

Definition 7. (*Autoencoder: The autoencoder, denoted as AE, is a type of artificial neural network used for unsupervised learning. It comprises two key components: an encoder and a decoder. The encoder compresses the input attribute data A into a lower-dimensional representation while the decoder reconstructs the original input from this compressed representation. The primary objective of utilizing an autoencoder in our study is to minimize the reconstruction error, measured using a reconstruction loss function. The reconstruction loss function of the attribute autoencoder is formulated as the comparison between the input attribute A_i and its corresponding reconstructed attribute A_{r-i}:*)

$$L_{recon} = \frac{1}{N}\sum_{i=1}^{N} \|A_i - A_{r-i}\|^2 \tag{5}$$

3.2. Types of misbehavior

Two major types of misbehavior identified by researchers based on realworld patterns are illustrated in Fig. 2 (Liu et al., 2022). These types have been extensively studied from various perspectives and have garnered significant academic attention (Akoglu et al., 2015; Bandyopadhyay et al., 2019; Ding et al., 2019; Ioannidis et al., 2021; Li et al., 2017; Ma et al., 2021).

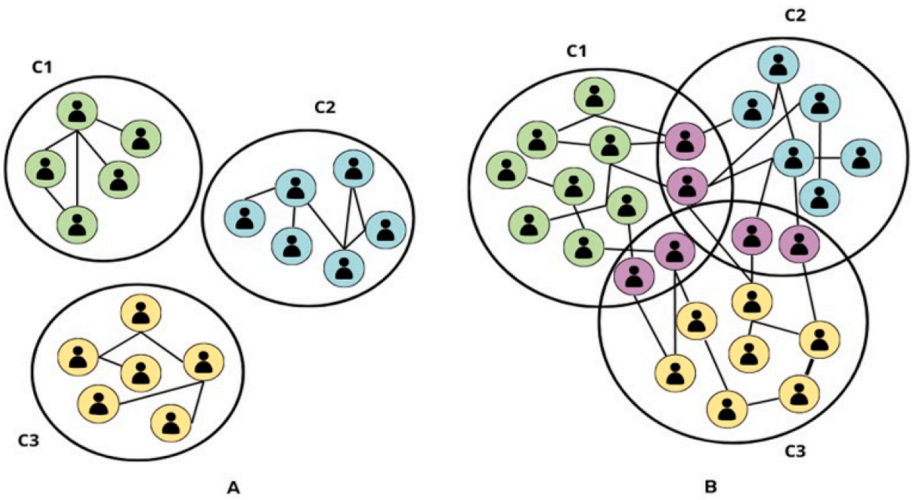

Fig. 1. Structures of disjoint communities (A) and overlapping communities (B).

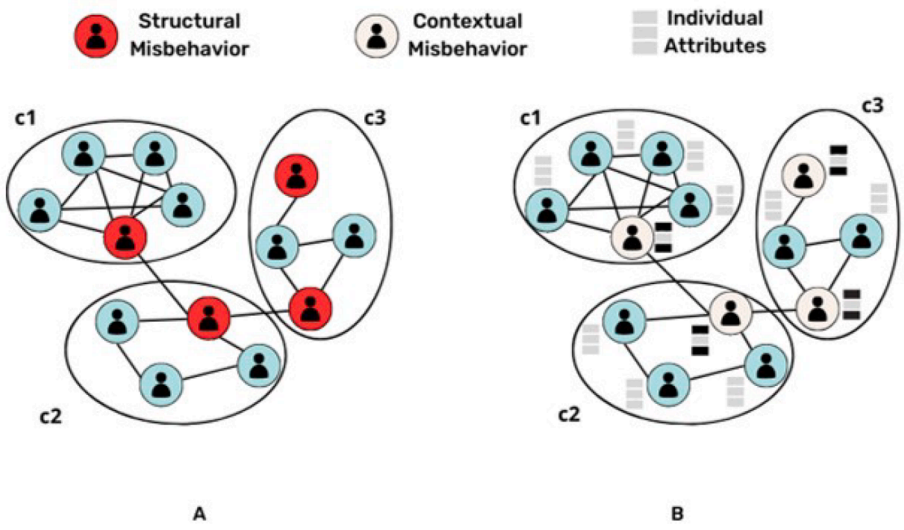

Fig. 2. Two primary types of misbehavior in social networks: A) structural misbehavior and B) contextual misbehavior.

- **Structural misbehavior** refers to anomalous behavior exhibited by nodes in a social network related to structural characteristics or relationships within a network. In the context of social networks, structural misbehavior may manifest as unexpected patterns of connectivity, unusual deviations within communities, or deviations from a network topology.
- **Contextual misbehavior** refers to anomalous behavior exhibited by nodes in a network with respect to their attributes or features within a specific context or situation. In social networks in which nodes have associated attributes (e.g., location data, purchase history, online activity), contextual misbehavior involves nodes that deviate from the norm or expected behavior based on their attributes. This may include nodes that change their attributes or manipulate attribute information to deceive and exploit other nodes in the network.

3.3. Problem formulation

Our approach looks to identify the set of misbehaving nodes $Misbhv \in V$ in a social network $G(V,E,A)$, where $Misbhv$ represents nodes that either belong to multiple communities or exhibit weak connections within their local communities.

4. Our proposed model: OCMLMD

4.1. Basic idea

Our approach, called OCMLMD (**O**verlapping **C**ommunities and **M**achine **L**earning for **M**isbehavior **D**etection), integrates both topological and semantic information to identify misbehaving individuals within social networks. These individuals evolve within the network's

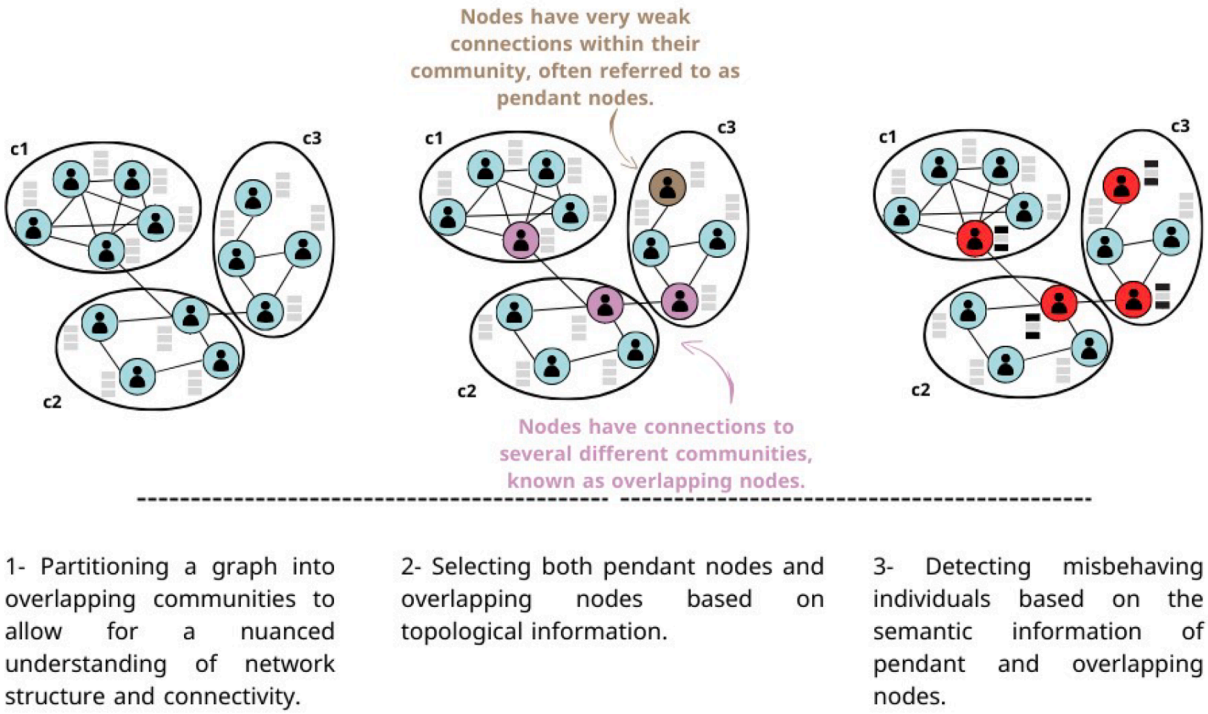

Fig. 3. The proposed approach involves two phases: selecting nodes using topological information and filtering them based on semantic information.

structure and exhibit deviant attributes, characteristics that make them challenging to detect using global approaches that focus on a single aspect of the network.

Traditional misbehavior detection methods often consider all nodes and attributes; our approach, however specifically targets two types of nodes for detecting misbehavior: pendant nodes and overlapping nodes. Pendant nodes reside within a single community with weak connections, whereas overlapping nodes are linked to multiple communities. Additionally, we define a subset of relevant attributes through context selection.

Our approach consists of two phases: first, the detection of pendant and overlapping nodes in a social network using topological information, and second, the filtration of these nodes using machine learning based on semantic information, as illustrated in Fig. 3.

The first step involves partitioning a network into overlapping communities to identify pendant nodes and overlapping nodes that deviate from the community structure. We employ an efficient algorithm called Docnet to identify overlapping communities (Rhouma and Romdhane, 2014). To mitigate the large number of overlapping nodes, we assess each node's importance using a clustering coefficient measure (Watts and Strogatz, 1998). A high clustering coefficient indicates strong connections to multiple communities while misbehaving overlapping nodes typically exhibit a low clustering coefficient and weak connections. Pendant nodes, on the other hand, are characterized by weak connections within their communities.

In the second step, we identify misbehavior in the network by computing the reconstruction error of relevant attributes that are inputted into an autoencoder.

index of community connectivity, is met in Eq. (1). However, this process often results in a large number of overlapping nodes. To refine the Docnet algorithm's outcome and address the issue of numerous overlapping nodes, we compute the overlapping node score in Eq. (3). This score is determined based on the clustering coefficient and overlapping node connection in Eq. (2) for each node identified as overlapping by the Docnet algorithm. A low score indicates weak connections between the overlapping node and its communities, potentially indicating misbehavior. For pendant nodes, their connection to the community is considered low, with a degree of 1. Further details regarding this process are outlined in the pseudo-code detailed in PNI-ONS (Algorithm 1).

4.2.2. Second Phase: Semantic information utilization

In this phase, the paper introduces an unsupervised machine-learning technique employing an autoencoder for behavior classification. Initially, a subset of attributes with the highest mutual information scores is selected in Eq. (4), followed by training the autoencoder on relevant attributes from only normal nodes. Subsequently, the trained autoencoder is utilized to compute the reconstruction error for attributes of each node within the set of pendant and overlapping nodes according to Eq. (5). Finally, the autoencoder is optimized by minimizing L_{recon} to its minimum, and nodes with reconstruction error exceeding a predefined threshold λ, indicative of significant deviations from normal behavior, are identified as misbehavior.

$$misbehavior = \begin{cases} 1, L_{recon} > \lambda \\ 0, L_{recon} < \lambda \end{cases} \tag{6}$$

For further details, refer to OCMLMD (Algorithm 2).

Algorithm 1: PNI-ONS Algorithm: Pendant Node Identification and Overlapping Node Selection

Input: Social network graph structure $G(V,E)$
Output: Set of pendant nodes and overlapping nodes *PAndO*

```
PAndO ← Empty set;
OverComm ← Partition G using the Docnet algorithm ;
OvpN ← Set of overlapping nodes;
CC ← clustering coefficient for all v ∈ OvpN;
for Each node v ∈ V do
    if degree(v)=1 then
        Add v to PAndO ;
for Each node v ∈ OvpN do
    Compute OvpCon (v) Using Eq.(2);
    Compute OvpScore (v) Using Eq.(3);
    if OvpScore < 1 then
        Add v to PAndO ;
```

4.2. Proposed misbehaving individuals detection approach

4.2.1. First Phase: Topological information utilization

The first phase of OCMLMD involves partitioning a graph into overlapping communities using the Docnet algorithm (Rhouma and Romdhane, 2014). This algorithm begins by identifying a starting core and gradually adds nodes until a stopping requirement, defined by the

5. Experiments

5.1. Experimental environment settings

The computer used in our experiment runs the Ubuntu 18.04.4 operating system, has an Intel(R) Core(TM) i7-1065G7 CPU running at 1.30 GHz with a clock speed of 1498 MHz, a 2 TB SSD for storage, and

16 GB of DDR4 RAM. The programming environment consists of Python 3.6, TensorFlow 1.15, and NumPy 1.16.6.

Algorithm 2: OCMLMD Algorithm: Overlapping Communities and Machine Learning for Detecting Misbehavior

Input: Social network attribute data A, Threshold for reconstruction error λ

Output: Set of misbehaving individuals *Misbhv*

Misbhv ← Empty set
PAndO ← Run PNI-ONS algorithm to obtain a set of pendant and overlapping nodes
For all $a_r \in A$ calculate mutual information (MI) using Eq. (4)
AttrSelect ← subset of N attributes with the highest mutual information score
AE ← trained autoencoder with *AttrSelect* for normal node

for *Each node* $v \in$ *PAndO* **do**
    Compute reconstruction error by *AE* Using Eq.(5)
    if *Reconstruction error exceeds* λ **then**
        Add v to *Misbhv*

5.2. Datasets

5.2.1. Real-world datasets

- **Disney Network** (Huang, 1998): The Disney network, a subset Amazon's copurchase network, contains information on Disney DVDs. Each item in the network is described by 30 attributes, including price, customer ratings, etc. Despite its small size, the network's intricate graph and attribute structure render it a compelling dataset that is widely employed for evaluating misbehavior detection models. To provide a ground truth for the dataset, a user experiment was conducted, in which high school students individually labeled each product as either normal or abnormal.

- **Books** (P. Iglesias Śanchez, , 2015 (2015).): The Book Network is an attributed network dataset that originated from Amazon's co-purchase network and specifically focuses on books. Similar to the Disney Network, each product in this network is characterized by a set of attributes, that may include features such as price, ratings, genre, and more. This dataset offers a valuable resource for evaluating misbehavior detection models within the context of online book purchases.

Table 1
Real-life datasets.

Dataset	Num.Nodes	Num.Edges	Num.Attrs	Mis. Ratio
Disney	124	334	30	4.8 %
Books	1468	3695	28	2.0 %
Enron	13,533	176,967	20	0.4 %

- **Enron** (Metsis et al., 2006): The Enron Network is an attributed network dataset derived from the Enron email corpus,. It is comprised of email communications from employees of the Enron Corporation. In this attributed network, nodes represent individual employees, while edges represent email exchanges between them. Each node is associated with various attributes such as job title, department, email frequency, and other metadata extracted from the email corpus. This dataset is frequently used in research endeavors spanning email categorization, network examination, and misbehavior identification (Table 1).

5.2.2. Synthetic datasets

Evaluating misbehavior detection models is challenging due to the limited availability of suitable datasets and ground truth. Synthetic datasets offer a solution for assessing model performance by simulating real-world scenarios. These datasets vary in terms of size and attribute complexity, and they aim to replicate characteristics observed in authentic networks (S'anchez et al., 2013). Thus, a graph is generated using a power-law distribution, with 50 % of attributes categorized as relevant and the other 50 % deemed irrelevant. Relevant attributes follow a Gaussian distribution, while irrelevant ones are assigned values uniformly. To prevent outliers, a hyperellipsoid truncates the tails of Gaussian distributions (see Fig. 4). Misbehaving nodes have their attributes manipulated with random values outside their communities' defined ranges. The proportion of misbehaving nodes is set at 10 %. The synthetic datasets consist of a graphml file containing node attributes and a true file providing ground truth.

5.3. Comparison methods and parameter configurations

5.3.1. Comparison methods

To evaluate the efficacy of our proposed approach, we conduct a comparative analysis against several baseline methods commonly used to detect misbehavior in social networks. The baseline methods included in our comparative analysis are:

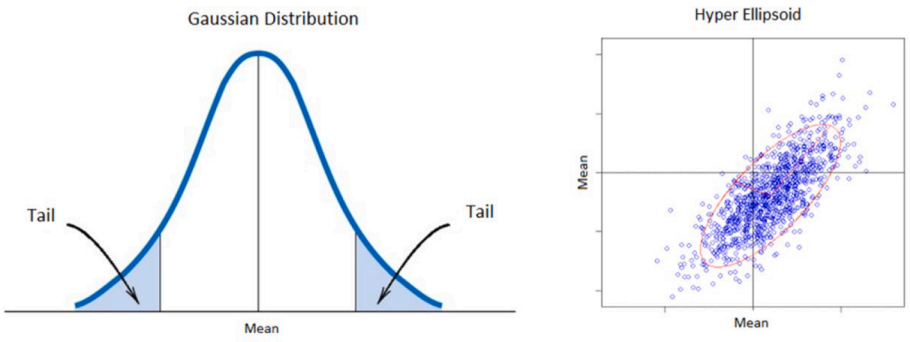

Fig. 4. Gaussian distribution with tail truncation.

- LOF (Breunig et al., 2000) utilizes node attributes to detect misbehavior by comparing the local density of data points. Nodes with significantly lower density than their neighbors are flagged for misbehavior.
- SCAN (Xu et al., 2007) identifies clusters or communities within a network by analyzing the local neighborhood structure of each node and then clustering nodes based on structural similarities. It relies solely on topological information to detect misbehavior.
- RADAR (Li et al., 2017) integrates both topological and semantic information to identify misbehavior by analyzing residuals, which represent differences between observed and predicted values from a model. These differences can indicate deviations from expected patterns.
- ANOMALOUS (Peng et al., 2018) uses CUR decomposition for attributes selection and residual analysis to identify misbehavior within social networks.
- DeepAD (Zhu et al., 2020) is an unsupervised misbehavior detection method that employs graph convolutional networks (GCN) to capture both topological and semantic information. By leveraging reconstruction errors, DeepAD effectively identifies misbehavior within a network.
- AnomMAN (Chen et al., 2023) is a graph convolution framework for detecting misbehavior in multi-view attributed networks. It utilizes an attention mechanism to consider the importance of different views and a graph autoencoder module for misbehavior detection.

5.3.2. Parameter configurations

For OCMLMD, we set the parameters as follows: number of attributes: selecting the top 20 %; number of layers: 4; batch size = 256; epochs: 100; optimizer = Adam; threshold: a prediction loss of 2 %. To ensure fairness in comparison, the parameters of other methods are configured to the optimal default values as documented in their respective papers.

5.4. Model evaluation

This paper compares the effectiveness of various approaches for detecting misbehavior using three widely used evaluation indicators. The indicators for evaluating the experiment include:

1) Receiver Operating Characteristic – Area Under the Curve (ROC-AUC): The results found by the detection approach and true individual misbehavior in networks suggest that there are four possible outcomes:

2) A normal individual is recognized as a misbehaving individual (FN).

3) A misbehaving individual is recognized as a normal individual (FP).

4) A normal individual is recognized as a normal individual (TN).

5) A misbehaving individual is recognized as a misbehaving individual (TP).

Therefore, the true positive rate (TPR) and false positive rate (FPR) for the detection of misbehaving individuals are defined as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP} \tag{7}$$

The AUC value is the area under the ROC curve, with the ordinate being TPR and the abscissa being FPR. A higher AUC value denotes a

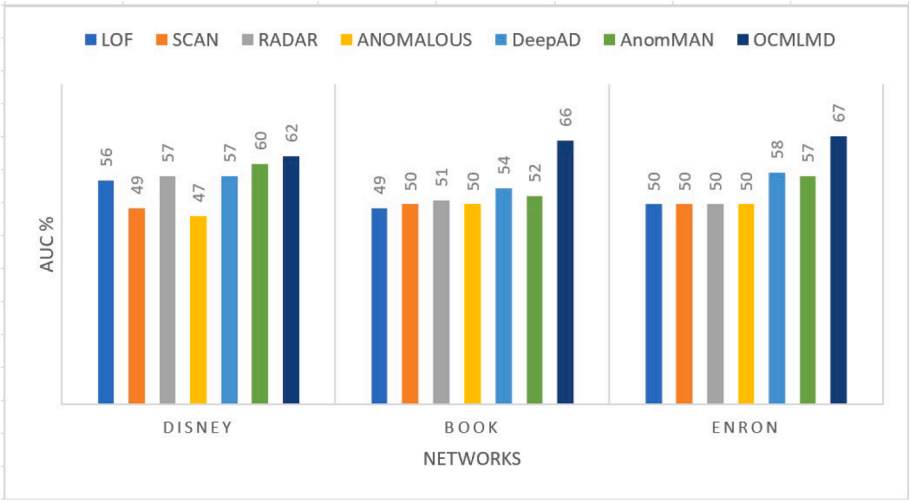

Fig. 5. ROC-AUC scores for various misbehavior detection methods evaluated on realworld social networks.

Table 2
Evaluation of real-world social networks using precision values for various misbehavior detection methods.

Dataset	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
Disney	0.562	0.475	0.646	0.474	0.624	0.562	**0.702**
Book	0.489	0.501	0.508	0.501	0.592	0.541	**0.647**
Enron	0.481	0.501	0.491	0.501	0.526	0.513	0.526

Table 3
Evaluation of real-world social networks using recall values for various misbehavior detection methods.

Dataset	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
Disney	0.562	0.491	0.574	0.474	0.561	0.571	**0.621**
Book	0.491	0.502	0.508	0.501	**0.572**	0.531	0.520
Enron	0.482	0.501	0.473	0.501	0.512	0.551	**0.573**

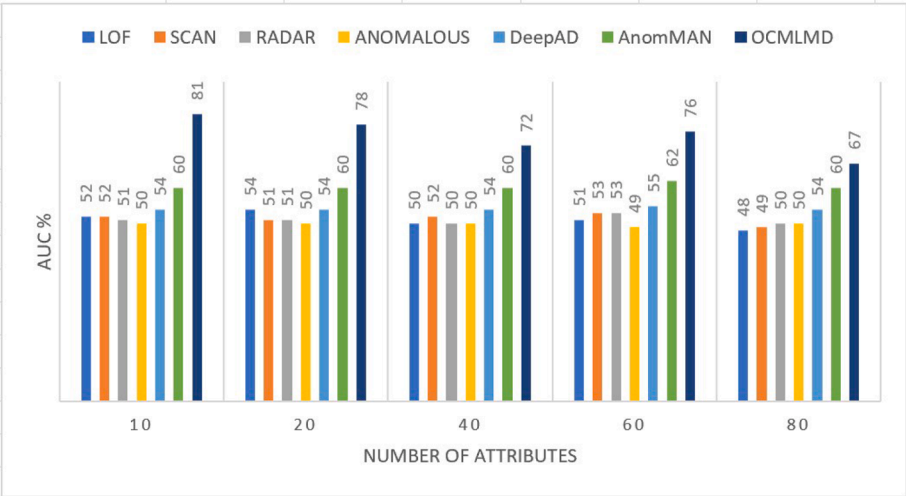

Fig. 6. ROC-AUC scores for various misbehavior detection methods with varying numbers of attributes.

more effective detection approach.

6) Precision: Precision serves as a significant evaluation metric for gauging the efficacy of a classification model, particularly concerning the detection of misbehavior within social networks. It quantifies the ratio of accurately identified positive cases (true positives) relative to all instances classified as positive (comprising both true positives and false positives). Mathematically, precision is measured as:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{8}$$

A higher precision value indicates that the model is better at correctly identifying misbehavior and avoiding incorrectly classifying normal individuals as misbehaving.

3) Recall: Also referred to as sensitivity or true positive rate, this represents another essential evaluation metric employed to evaluate the effectiveness of a classification model, especially in the realm of detecting misbehavior within social networks. Recall quantifies the proportion of accurately identified positive cases (true positives) out of all actual positive cases present in the dataset. Mathematically, recall can be computed as:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{9}$$

A higher recall value indicates that the model is better at capturing all instances of misbehavior in the dataset, minimizing the number of misbehaving individuals that are overlooked (false negatives).

5.5. Experimental results and analysis

5.5.1. Real-world results

- **Comparative analysis of AUC values for various detection methods**: The experimental results shed light on the performance of different misbehavior detection methods across real-world datasets shown in Fig. 5. OCMLMD emerges as the most effective method for identifying misbehavior across all datasets, given it has the highest AUC values. DeepAD performs well on the Book and Enron datasets, while AnomMAN is competitive with relatively high AUC values.

Table 4
Evaluation of synthetic network datasets using precision values for different misbehavior detection methods.

NumAttrs	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
10	0.517	0.509	0.516	0.498	0.524	0.553	**0.721**
20	0.539	0.518	0.535	0.495	0.512	0.542	**0.701**
40	0.502	0.527	0.497	0.497	0.551	0.664	**0.682**
60	0.511	0.544	0.581	0.489	0.551	**0.674**	0.653
80	0.474	0.485	0.501	0.497	0.504	0.605	**0.651**

Table 5
Evaluation of synthetic network datasets using recall values for different misbehavior detection methods.

NumAttrs	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
10	0.516	0.524	0.505	0.499	0.535	0.519	**0.634**
20	0.535	0.513	0.512	0.495	0.532	0.513	**0.635**
40	0.502	0.517	0.499	0.497	0.509	0.517	**0.622**
60	0.509	0.534	0.526	0.489	0.509	**0.676**	0.651
80	0.477	0.489	0.501	0.497	0.517	0.629	**0.657**

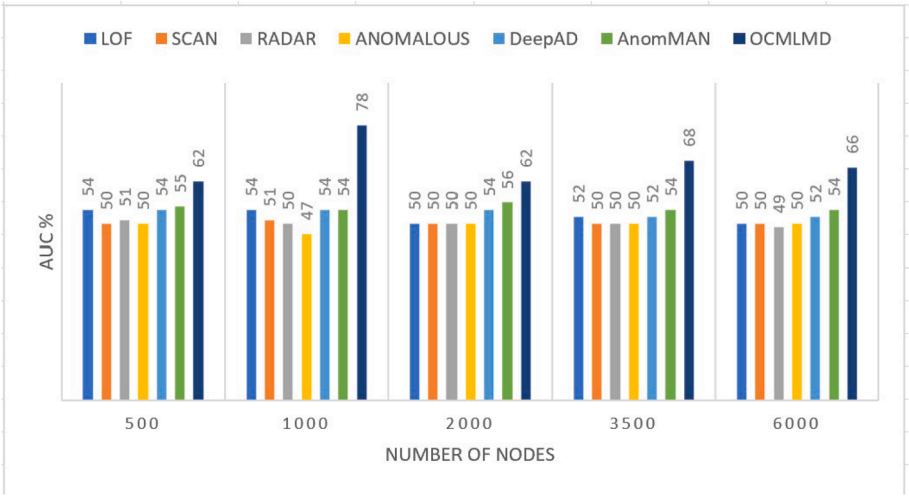

Fig. 7. Variations in ROC-AUC scores across misbehavior detection methods with varying numbers of nodes.

Table 6
Evaluation of synthetic network datasets using precision values for different misbehavior detection methods.

Num.Nodes	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
500	0.554	0.498	0.546	0.499	0.546	0.546	**0.698**
1000	0.539	0.515	0.506	0.474	0.549	0.545	**0.684**
2000	0.501	0.449	0.447	0.459	0.557	0.554	**0.691**
3500	0.522	0.504	0.501	0.501	0.552	0.554	**0.721**
6000	0.496	0.501	0.481	0.496	0.551	0.576	**0.712**

RADAR exhibits moderate performance that is slightly higher than baseline methods. LOF, SCAN, and ANOMALOUS demonstrate lower AUC values, suggesting poorer performance in identifying misbehavior. In summary, OCMLMD is the most effective method, followed by DeepAD and AnomMAN. This suggests that deep learning-based approaches effectively leverage network topology and semantic information, leading to improved misbehavior detection performance.

- **Comparative analysis of the precision and recall values for various detection methods:** Tables 2 and 3 show that OCMLMD consistently achieves the highest values across all datasets, indicating its effectiveness in identifying misbehavior. DeepAD also outperforms other methods, particularly on the Book dataset. RADAR and AnomMAN show good precision values, slightly lower than OCMLMD and DeepAD. LOF, SCAN, and ANOMALOUS have lower precision values, suggesting poorer performance in identifying misbehavior. In summary, OCMLMD is the most effective method for misbehavior detection, followed by DeepAD.

5.5.2. Synthetic network results

- **First Experiment**:
  - **Comparative analysis of AUC values for various detec-tion methods:** Fig. 6 reveals that OCMLMD is the most effective method, consistently achieving the highest AUC values and maintaining good scalability across various numbers of attributes. DeepAD and AnomMAN also exhibit competitive performance in terms of AUC values and scalability. RADAR performs moderately well but shows slightly lower AUC values compared to OCMLMD, DeepAD, and AnomMAN. LOF, SCAN, and ANOMALOUS demonstrate poorer performance in terms of AUC values and scalability. Overall, OCMLMD, DeepAD, and AnomMAN are considered the top-performing methods.
  - **Comparative analysis of the precision and recall values for-various detection methods**: In Tables 4 and 5, OCMLMD appears to be the most effective method, consistently achieving high values and maintaining optimal scalability across various numbers of attributes. AnomMAN also demonstrates satisfactory competitive performance in terms of precision, recall values, and scalability. DeepAD exhibits stable performance but may have slightly lower precision and recall compared to OCMLMD and AnomMAN. Radar, LOF, and SCAN perform moderately well but may encounter scalability challenges. ANOMALOUS showcases poorer performance in terms of precision, recall values, and scalability.

  Overall, OCMLMD and AnomMAN stand out as the top-performing methods.

- **Second Experiment**:

Table 7
Evaluation of synthetic network datasets using recall values for different misbehavior detection methods.

Num.Nodes	LOF	SCAN	RADAR	ANOMALOUS	DeepAD	AnomMAN	OCMLMD
500	0.543	0.499	0.513	0.499	0.503	0.519	**0.633**
1000	0.536	0.512	0.502	0.473	0.513	0.519	**0.626**
2000	0.501	0.501	0.502	0.513	0.513	0.512	**0.623**
3500	0.521	0.503	0.501	0.501	0.513	0.518	**0.632**
6000	0.497	0.501	0.492	0.496	0.502	0.521	**0.635**

- **Comparative analysis of AUC values for various detection-methods**: Fig. 7 shows that OCMLMD is the most effective method due to its consistent achievement of the highest AUC values and maintaining good scalability across various numbers of nodes. AnomMAN also demonstrates competitive performance in terms of AUC values and scalability. DeepAD is a stable performer that may exhibit slightly lower AUC compared to OCMLMD and AnomMAN. RADAR and LOF perform moderately well but may face scalability challenges. SCAN and ANOMALOUS have poorer performance in terms of AUC values and scalability. Overall, OCMLMD and AnomMAN are the best misbehavior detection methods.
- **Comparative analysis of the precision and recall valuesfor various detection methods**: Tables 6 and 7 indicate that OCMLMD is the most effective method, as made evident by its high precision, recall values, and good scalability across various numbers of nodes. AnomMAN also demonstrates competitive performance in terms of precision, recall values, and scalability. DeepAD exhibits stable performance but has slightly lower precision and recall. LOF, SCAN, and RADAR perform moderately well but may face scalability challenges. ANOMALOUS exhibits poorer performance than other methods in terms of its precision, recall values, and scalability. Overall, OCMLMD and AnomMAN are the top-performing methods, followed by DeepAD.

6. Conclusion

This paper presents OCMLMD, a novel approach for detecting misbehavior in social networks through a combination of graph-based analyses and machine-learning techniques. In the structure-based analysis, OCMLMD leverages graph-based analyses to identify pendant nodes within single communities and overlapping nodes that belong to multiple communities. By focusing on these nodes, OCMLMD reduces the search space for detecting misbehavior, enabling more efficient and effective analysis of social network data. In the machine learning techniques part, OCMLMD utilizes machine learning techniques, particularly an autoencoder model, to classify network individuals based on reconstruction error. By training the autoencoder on relevant attributes from normal nodes, OCMLMD learns to distinguish between normal and abnormal behavior. The reconstruction error serves as a measure of deviation from normal behavior, allowing OCMLMD to accurately identify misbehavior in social networks. Through experimental evaluations on both real-world and synthetic datasets, OCMLMD demonstrates superior performance compared to existing methods. Its ability to effectively capture misbehavior across diverse social network structures highlights its effectiveness and robustness in detecting misbehavior.

Future research directions may include investigating the robustness of OCMLMD against adversarial attacks and data noise, which are common challenges in real-world social network environments. By enhancing the method's resilience to such adversities, we can bolster its effectiveness in accurately detecting misbehavior amidst varying degrees of interference. Additionally, there is potential for extending OCMLMD to incorporate temporal dynamics and evolving community structures within social networks. This expansion could yield deeper insights into the nuanced dynamics of misbehavior detection over time, allowing for more proactive and adaptive detection strategies. Moreover, investigating the integration of multimodal data, such as text, images, and videos, into OCMLMD could enrich the analysis of social network behavior and detect misbehavior across different types of content shared on social networks, enhancing its versatility and effectiveness.

CRediT authorship contribution statement

Wejdan Alshlahy: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft. **Delel Rhouma:** Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2024-9/1).

References

Ahmed, N., Shahbaz, T., Shamim, A., Khan, K.S., Hussain, S., Usman, A., 2020. The covid-19 infodemic: a quantitative analysis through facebook. Cureus 12 (11).
L. Akoglu, M. McGlohon, C. Faloutsos, Anomaly detection in large graphs, CMU-C S-09-173 Technical Report (2009).
L. Akoglu, M. McGlohon, C. Faloutsos, Oddball: Spotting anomalies in weighted graphs, in: Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 2124, 2010. Proceedings. Part II 14, Springer, 2010, pp. 410–421.
Akoglu, L., Tong, H., Koutra, D., 2015. Graph based anomaly detection and description: a survey. Data Min. Knowl. Disc. 29, 626–688.
Angiulli, F., Pizzuti, C., 2002. Fast outlier detection in high dimensional spaces. In: European Conference on Principles of Data Mining and Knowledge Discovery. Springer, pp. 15–27.
S. Bandyopadhyay, N. Lokesh, M. N. Murty, Outlier aware network embedding for attributed networks, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 12–19.
Bandyopadhyay, S., L. N, Vivek, S.V., Murty, M.N., 2020. Outlier resistant unsupervised deep architectures for attributed network embedding, in. In: Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 25–33.
Bishop, C.M., 2006. Pattern recognition and machine learning. Springer Google Schola 2, 645–678.
A. Bojchevski, S. Günnemann, Bayesian robust attributed graph clustering: Joint learning of partial anomalies and group structure, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018, p.2738–2745.
M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander, Lof: identifying density-based local outliers, in: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 2000, pp. 93–104.
Chen, L.-H., Li, H., Zhang, W., Huang, J., Ma, X., Cui, J., Li, N., Yoo, J., 2023. Anomman: Detect anomalies on multi-view attributed networks. Inf. Sci. 628, 1–21.
Q. Ding, N. Katenka, P. Barford, E. Kolaczyk, M. Crovella, Intrusion as (anti) social communication: characterization and detection, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 886–894.
K. Ding, J. Li, R. Bhanushali, H. Liu, Deep anomaly detection on attributed networks, in: Proceedings of the 2019 SIAM International Conference on Data Mining, SIAM, 2019, pp. 594–602.
K. Ding, J. Li, H. Liu, Interactive anomaly detection on attributed networks, in: Proceedings of the twelfth ACM international conference on web search and data mining, 2019, pp. 357–365.

Fan, H., Zhang, F., Li, Z., 2020. Anomalydae: Dual autoencoder for anomaly detection on attributed networks. In: In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5685–5689.
D. Ghoshal, L. Ramakrishnan, Madats: Managing data on tiered storage for scientific workflows, in: Proceedings of the 26th International Symposium on High-Performance Parallel and Distributed Computing, 2017, pp. 41–52.
Ghoshal, A.K., Das, N., Bhattacharjee, S., Chakraborty, G., 2019. A fast parallel genetic algorithm based approach for community detection in large networks. In: In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS). IEEE, pp. 95–101.
Ghoshal, A.K., Das, N., Das, S., 2021. Influence of community structure on misinformation containment in online social networks. Knowl.-Based Syst. 213, 106693.
M. Gupta, J. Gao, J. Han, Community distribution outlier detection in heterogeneous information networks, in: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part I 13, Springer, 2013, pp. 557–573.
T. J. Helling, J. C. Scholtes, F. W. Takes, A community-aware approach for identifying node anomalies in complex networks, in: Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7, Springer, 2019, pp. 244–255.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the Dimensionality of Data with Neural Networks, Science 313 (5786), 504–507.
B. Hooi, H. A. Song, A. Beutel, N. Shah, K. Shin, C. Faloutsos, Fraudar: Bounding graph fraud in the face of camouflage, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 895–904.
Hu, R., Aggarwal, C.C., Ma, S., Huai, J., 2016. An embedding approach to anomaly detection. In: In: 2016 IEEE 32nd International Conference on Data Engineering (ICDE). IEEE, pp. 385–396.
Y. Hu, C. Chen, B. Deng, Y. Lai, H. Lin, Z. Zheng, J. Bian, Decoupling anomaly discrimination and representation learning: self-supervised learning for anomaly detection on attributed graph, arXiv preprint arXiv:2304.05176 (2023).
Huang, Z., 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min. Knowl. Disc. 2 (3), 283–304.
T. Huang, Y. Pei, V. Menkovski, M. Pechenizkiy, Hop-count based selfsupervised anomaly detection on attributed networks, in: Joint European conference on machine learning and knowledge discovery in databases, Springer, 2022, pp. 225–241.
Iglesias Śanchez, P., 2015 (2015).. Context selection on attributed graphs for outlier and community detection, Ph.D. thesis, Karlsruhe, Karlsruher Institut für Technologie (KIT). Diss.
Ioannidis, V.N., Berberidis, D., Giannakis, G.B., 2021. Unveiling anomalous nodes via random sampling and consensus on graphs. In: In: ICASSP 20212021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5499–5503.
Jin, M., Liu, Y., Zheng, Y., Chi, L., Li, Y.-F., Pan, S., 2021. Anemone: Graph anomaly detection with multi-scale contrastive learning, in. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3122–3126.
Keyvanpour, M.R., Shirzad, M.B., Ghaderi, M., 2020. Ad-c: a new node anomaly detection based on community detection in social networks. Int. J. Electron. Bus. 15 (3), 199–222.
T. N. Kipf, M. Welling, Variational graph auto-encoders, arXiv preprint arXiv: 1611.07308 (2016).
Langford, J., Zhang, T., 2007. The epoch-greedy algorithm for multi-armed bandits with side information. Adv. Neural Inf. Proces. Syst. 20.
J. Li, H. Dani, X. Hu, H. Liu, Radar: Residual analysis for anomaly detection in attributed networks., in: IJCAI, Vol. 17, 2017, pp. 2152– 2158.
Li, Y., Huang, X., Li, J., Du, M., Zou, N., 2019. Specae: Spectral autoencoder for anomaly detection in attributed networks, in. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2233–2236.
Lin, Z., Wang, H., Li, S., 2022. Pavement anomaly detection based on transformer and self-supervised learning. Autom. Constr. 143, 104544.
Liu, K., Dou, Y., Zhao, Y., Ding, X., Hu, X., Zhang, R., Ding, K., Chen, C., Peng, H., Shu, K., et al., 2022. Bond: Benchmarking unsupervised outlier node detection on static attributed graphs. Adv. Neural Inf. Proces. Syst. 35, 27021–27035.
Z. Liu, C. Cao, J. Sun, Mul-gad: a semi-supervised graph anomaly detection framework via aggregating multi-view information, arXiv preprint arXiv:2212.05478 (2022).
Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., Karypis, G., 2021. Anomaly detection on attributed networks via contrastive self-supervised learning. IEEE Trans. Neural Networks Learn. Syst. 33 (6), 2378–2392.
Liu, F.T., Ting, K.M., Zhou, Z.-H., 2012. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD) 6 (1), 1–39.
Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q.Z., Xiong, H., Akoglu, L., 2021. A comprehensive survey on graph anomaly detection with deep learning. IEEE Trans. Knowl. Data Eng.
V. Metsis, I. Androutsopoulos, G. Paliouras, Spam filtering with naive bayes-which naive bayes?, in: CEAS, Vol. 17, Mountain View, CA, 2006, pp. 28–69.
Miller, B., Bliss, N., Wolfe, P., 2010. Subgraph detection using eigenvector l1 norms. Adv. Neural Inf. Proces. Syst. 23.
Morales, P., Caceres, R.S., Eliassi-Rad, T., 2021. Selective network discovery via deep reinforcement learning on embedded spaces. Applied Network Science 6, 1–20.
Peng, Z., Luo, M., Li, J., Liu, H., Zheng, Q., et al., 2018. Anomalous: A joint modeling approach for anomaly detection on attributed networks. IJCAI, in, pp. 3513–3519.
Peng, Z., Luo, M., Li, J., Xue, L., Zheng, Q., 2020. A deep multi-view framework for anomaly detection on attributed networks. IEEE Trans. Knowl. Data Eng. 34 (6), 2539–2552.
B. Perozzi, L. Akoglu, P. Iglesias Sánchez, 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, 2014, pp. 1346–1355.
Rhouma, D., Romdhane, L.B., 2014. An efficient algorithm for community mining with overlap in social networks. Expert Syst. Appl. 41 (9), 4309–4321.
S'anchez, P.I., Müller, E., Laforet, F., Keller, F., Böhm, K., 2013. Statistical selection of congruent subspaces for mining attributed graphs. IEEE 13th international conference on data mining, IEEE 2013, 647–656.
M. Sakurada, T. Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, 2014, pp. 4–11.
X.-R. Sheng, D.-C. Zhan, S. Lu, Y. Jiang, Multi-view anomaly detection: Neighborhood in locality matters, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 4894–4901.
Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19 (1), 22–36.
T. Ide, H. Kashima, Eigenspace-based anomaly detection in computer systems, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 440–449.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world'networks. Nature 393 (6684), 440–442.
Win, H.N., Lynn, K.T., 2018. Community and outliers detection in social network. In: International Conference on Big Data Analysis and Deep Learning Applications. Springer, pp. 58–67.
Wu, J., Zhu, X., Zhang, C., Cai, Z., 2013. Multi-instance multi-graph dual embedding learning. In: In: 2013 IEEE 13th International Conference on Data Mining, pp. 827–836.
Wu, J., Pan, S., Zhu, X., Cai, Z., 2014. Boosting for multi-graph classification. IEEE Trans. Cybern. 45 (3), 416–429.
X. Xu, N. Yuruk, Z. Feng, T. A. Schweiger, Scan: a structural clustering algorithm for networks, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 824–833.
H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng, et al., Unsupervised anomaly detection via variational autoencoder for seasonal kpis in web applications, in: Proceedings of the 2018 world wide web conference, 2018, pp. 187–196.
Xu, Z., Huang, X., Zhao, Y., Dong, Y., Li, J., 2022. Contrastive attributed network anomaly detection with data augmentation, in. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer 444–457.
Zhou, S., Huang, X., Liu, N., Zhou, H., Chung, F.-L., Huang, L.-K., 2023. Improving generalizability of graph anomaly detection models via data augmentation. IEEE Trans. Knowl. Data Eng.
M. Zhu, H. Zhu, Mixedad: A scalable algorithm for detecting mixed anomalies in attributed graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 1274–1281.
D. Zhu, Y. Ma, Y. Liu, Deepad: A joint embedding approach for anomaly detection on attributed networks, in: Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part II 20, Springer, 2020, pp. 294–307.