IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 February 2024, accepted 3 April 2024, date of publication 8 May 2024, date of current version 15 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3398555

RESEARCH ARTICLE

Detecting Anomalies in Attributed Networks Through Sparse Canonical Correlation Analysis Combined With Random Masking and Padding

WASIM KHAN[1], MOHAMMAD ISHRAT[1], AHMAD NEYAZ KHAN[1], (Member, IEEE), MOHAMMAD ARIF[2], ANWAR AHAMED SHAIKH[1], MOUSA MOHAMMED KHUBRANI[3], SHADAB ALAM[3], (Senior Member, IEEE), MOHAMMED SHUAIB[3], (Member, IEEE), AND RAJAN JOHN[3], (Member, IEEE)

[1]Koneru Lakshmaiah Education Foundation, Andhra Pradesh 522502, India
[2]Vellore Institute of Technology, Vellore 632014, India
[3]Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan 82817, Saudi Arabia

Corresponding author: Mousa Mohammed Khubrani (mmkhubrani@jazanu.edu.sa)

The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia through Project Number: GSSRD-24.

ABSTRACT Attributed networks are prevalent in the current information infrastructure, where node attributes enhance knowledge discovery. Anomaly detection in attributed networks is gaining attention for its potential uses in cybersecurity, finance, and healthcare. Recognizing the complicated relationship between node attributes and network topology is crucial for attributed network embedding and anomaly detection. Nevertheless, there are few approaches available to directly represent the relationship between these two perspectives of the node property and the network topology. Approaches utilizing the reconstruction error rely on straightforward, simple mappings, which introduce a substantial risk of overfitting in high-dimensional data, wherein the model acquires patterns that are exclusive to the training data and fails to generalize to new data. To do this, we suggest a new way to find graph anomalies on attributed networks using random masking and padding along with sparse canonical correlation analysis. Motivated by the limitations of existing methodologies in effectively addressing these challenges, our research introduces a novel methodology for anomaly detection in attributed networks by leveraging Sparse Canonical Correlation Analysis (SCCA) in conjunction with Random Masking and Padding (RMP). This dual approach uniquely addresses the challenges of high-dimensional data and the sparsity of attributes, which are prevalent issues in anomaly detection. Unlike previous works that primarily focus on either dimensionality reduction or attribute sparsity independently, our method synergizes these aspects to enhance detection performance. Initially, we randomly mask and pad nodes in the attributed network and use the Graph Convolutional Network (GCN) to map them to latent space. Next, we optimize the distribution alignment of node attributes and graph structure latent space representations by using Kullback-Leibler (KL) divergence regularization, which increases their comparability. Finally, we use sparse canonical correlation analysis (SCCA) to quantify the correlation between node attributes and network structure views in latent space. SCCA incorporates sparsity by making the model choose fewer variables, which adds another level of complexity. It improves interpretability and reduces overfitting in high-dimensional data analysis by highlighting only the key variables. To optimize our model, we maximize the correlation between attribute and structural aspects of normal nodes, and anomalies are detected by measuring the correlation between these two views. Our approach is the first of its kind to provide a novel remedy to the fundamental problems preventing efficient and accurate anomaly identification, thereby establishing a new standard in this field.

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

The proposed model has been extensively tested on four real-world datasets, and its effectiveness has been demonstrated in comparison to state-of-the-art approaches. The empirical evaluation across multiple benchmark datasets validates the potential of the proposed approach as a pivotal tool in advancing anomaly detection research and applications.

INDEX TERMS Attributed networks anomaly detection, random masking and padding, correlation analysis, graph convolution network, KL divergence, deep learning.

I. INTRODUCTION

Attributed networks, a common form of graph data, represent entities as nodes and their relationships as links [1], [2], [3], [4]. In recent years, there has been a significant focus among researchers on graph analysis, encompassing areas like node classification, link prediction, and, notably, anomaly detection. Within this spectrum, the detection of anomalies in attributed networks stands out as a crucial area of study. Identification of nodes with unusual behavior is required. The interaction between network topologies and node characteristics makes anomaly detection in attributed networks difficult [5], [6], [7]. Considering both aspects of graph learning increases this complexity. Traditional anomaly identification approaches, while beneficial in other situations, fail to completely understand the graph's informative breadth in attributed network anomaly detection [8], [9], [10].

Factorization-based graph embedding learning [11], [12], [13] assumes a node's embedding is a linear combination of its neighbors' embeddings. DeepWalk [14] revolutionizes graph embedding learning with deep learning. Bayesian random walks are then used by Node2vec [15] to learn node embeddings. Node2vec uses Bayesian random walks to learn node embeddings, which can capture both the local and global structure of the network. The performance of Node2vec can be sensitive to the choice of hyperparameters. Dominant [16] uses Graph Convolutional Networks (GCN) to handle network sparsity and represent nonlinear interactions and complexity among nodes in an attributed network. Dominant uses Graph Convolutional Networks (GCN) to handle network sparsity and represent nonlinear interactions and complexity among nodes in an attributed network. It struggles with high-dimensional data and overfitting. Integrating cross-modal interactions between node characteristics and network structure gives AnomalyDAE [17] better embeddings. This method integrates cross-modal interactions between node characteristics and network structure for better embeddings but not perform well on data with complex structures or high-dimensional data. An improved hybrid embedding approach, DeepAD [18], uses attributes and network architectures' non-linear properties. Using complicated data patterns, this approach analyses reconstruction flaws to find abnormalities. DeepAD uses attributes and network architectures' non-linear properties for improved hybrid embedding. It analyses reconstruction flaws to find abnormalities, which may not always be accurate or efficient. Furthermore, a highly developed method in this area is DUAL-SVDAE [19]. The structure and attribute autoencoders learn node embedding representations. Next, a dual-hypersphere training method learns two normal node hyperspheres, improving anomaly detection precision and efficacy. This approach uses a dual-hypersphere training method to learn two normal node hyperspheres, improving anomaly detection precision and efficacy but it is computationally expensive and complex to implement. Unlike reconstruction error approaches, ResGCN [20] ranks anomalies using residual information from the input network. This technique uses graph GCN to manage network data sparsity and nonlinearity. ResGCN uses the network's intrinsic features to detect anomalies by focusing on residual information. ResGCN uses the network's intrinsic features to detect anomalies by focusing on residual information. It does not perform well on data with complex structures or high-dimensional data.

Although graph embedding learning may discover aberrant nodes in associated networks, it can be limited when using complicated deep neural network models in real-world applications. Many methods focus on extracting information from attribute networks by creating intricate interaction learning structures, as seen in studies like [9], [17], [21], and [22]. These approaches tend to emphasize complex data interactions rather than directly developing new anomaly detection protocols. This can lead to less effective identification of anomalies as the primary focus isn't on innovating detection strategies. Some methods of self-supervised learning only look at adding more data to find local and global connections. They don't look at the important connections between network structures and node attributes in the latent space [23], [24]. Several approaches, including two-step methods and graph auto-encoder techniques, are developed primarily to understand graph embedding, not to identify anomalous nodes, deviating from the main goal of anomaly detection.

Specifically, in attributed networks, normal nodes and their neighbors usually exhibit similar structural and attribute distribution states. This implies a strong association among normal nodes, a trait not shared with abnormal nodes [25], [26], [27]. Consequently, using neighbor information to supplement the nodes in the reconstruction process can effectively highlight the inconsistencies of reconstructed abnormal nodes compared to their original forms. When calculating reconstruction error, this approach can more accurately identify anomalies as the discrepancies become more apparent [28], [29]. This methodology addresses the need to differentiate subtle but crucial variations between normal and seemingly normal-like abnormal nodes, enhancing the effectiveness of anomaly detection in attributed networks. Additionally, in high-dimensional data, there's a significant risk of overfitting where the model learns patterns specific to the training data, failing to generalize to new data, whereas the approaches using the reconstruction error rely on simple, direct mappings [28], [30], [31], [32].

Random masking and padding help mitigate the problem of overfitting by introducing randomness and variability in the training process [33], [34]. Masking introduces a form of data augmentation. By randomly hiding some features (masking), the model is encouraged to learn more robust patterns rather than overfitting to specific feature values. This helps in improving the generalization capabilities of the model. Padding adds a level of stochasticity to the data, further enhancing the model's ability to generalize well on unseen data, and by randomly masking features, the model is forced to learn deeper dependencies between features as it cannot rely on any single feature too heavily. But if we are using only random masking and padding, abnormal node reconstruction is difficult due to the introduced noise and non-alignment of neighbor node embedding. Since random masking and padding produce arbitrary randomness, it is also not possible to identify anomalies efficiently. Introduced noise results in overfitting and the diversification of anomalies is not generalized either because of biassed latent space exploration.

A single framework using random masking and padding with canonical correlation analysis (CCA) is an efficient technique for improving adaptability [35]. By using CCA, it is possible to align the introduced variability with the typical node pattern and find significant attribute and structure feature correlations. All the possible correlations are considered to achieve better accuracy in detecting anomalies. Attributed network anomaly detection can be seen as the problem of correlation measurement between the node features and topological structure. Many approaches used the CCA for the alignment of latent space distribution between the features and network structure [36], [37], [38]. When used with high-dimensional data, CCA faces major problems. More precisely, CCA tends to overfit and frequently produces solutions that are challenging to comprehend because they include an excessive number of variables. CCA models are computationally costly and difficult to comprehend since they maximize correlations without limiting the number of variables.

We proposed a model that utilizes sparse CCA, which uses sparsity-based regularization and integrates the L1 norm loss for constraining the CCA optimization. As a result, it overcomes the risk of overfitting while also improving the result's interpretability. Hence, SCCA is a more viable alternative to CCA because the final canonical variables only comprise the most important and impactful variables, making it more suitable for high-dimensional and complex datasets. This integrated model, thus, presents a comprehensive and robust solution for anomaly detection, addressing the multidimensional challenges posed by attributed networks. Initially, we perform random masking and padding on nodes and input them into GCN to extract the embeddings. Then, we smooth the distribution using Kullback-Leibler (KL) divergence regularization to make the abnormal nodes closer to the distribution of normal nodes. The following are the key aspects of this paper:

- The introduction of random masking and padding as a preprocessing step for node attributes before feeding them into the GCN contributes to more robust feature representation and helps the model learn to identify essential features, potentially improving its ability to detect anomalies.
- KL divergence regularization is used to align the distributions of the encoded attribute and structure features. This makes sure that the latent representations are well-aligned, which makes it easier to compare and find the anomalies.
- Incorporating sparse canonical correlation analysis (SCCA) to maximize the correlation between the latent representations of attributes and structure during training contributes to the field by providing a means to detect anomalies more effectively.
- It is possible to achieve more discriminative power in anomaly detection through the combined use of random masking and padding, followed by SCCA. Significant attribute and structure feature correlations are discovered by aligning the introduced variability with the normal node pattern, which makes the model more effective in detecting anomalies.
- As variability is introduced with the random operations, SCCA guarantees that the attribute features and structure characteristics are well aligned with the exploration. The comprehension of the dynamics of the network is enhanced by this alignment.
- KL divergence regularization is used to line up the distributions of the encoded attribute and structure features. This makes sure that the latent representations are well-aligned, which makes it easier to compare things and find problems.
- Random masking and padding, GCN, distribution alignment, and correlation analysis are collectively optimized inside a unified framework. This optimization allows one component to enhance the others, leading to improved anomaly detection.

The proposed model demonstrates superiority in both theoretical and practical perspectives:

Theoretical Perspective:

- *Addressing High-Dimensional Data and Attribute Sparsity:* The suggested methodology specifically tackles the difficulties provided by high-dimensional data and the scarcity of attributes, which are common problems in anomaly identification. This method combines dimensionality reduction with attribute sparsity to improve detection performance, unlike earlier works that just focus on one of these characteristics.
- *Optimizing Distribution Alignment:* The model employs Kullback-Leibler (KL) divergence regularization to

enhance the comparability of node characteristics and graph structure latent space representations by optimizing their distribution alignment.

- *Incorporating Sparsity:* The model uses Sparse Canonical Correlation Analysis (SCCA) to measure the correlation between node attributes and network structure views in a hidden space. SCCA includes sparsity by enforcing the selection of a reduced number of variables, hence introducing an additional layer of intricacy. By selectively emphasizing the essential variables, it enhances interpretability and mitigates overfitting in the analysis of complex data with a high number of dimensions.

Practical Perspective:

- *Enhanced Anomaly Detection:* The model aims to optimize the association between the attributes and structural characteristics of regular nodes. Anomalies are identified by evaluating the correlation between these two perspectives. This approach offers an innovative solution to the underlying issues that hinder the efficient and precise detection of anomalies.
- *Empirical Validation:* The proposed model has undergone thorough testing on five real-world datasets, and its efficacy has been proved when compared to cutting-edge methods. The suggested approach has been empirically evaluated on different benchmark datasets, confirming its promise as a crucial tool for furthering research and applications in anomaly identification.

In summary, the proposed model offers a unique and effective approach to anomaly detection in attributed networks by addressing key challenges in the field and demonstrating strong performance in empirical tests. The rest of this work is structured in the following manner: An analysis of the relevant literature on attributed network anomaly detection is provided in Section II. The problem of anomaly detection on attributed networks is clearly stated in Section III. Section IV describes the preliminaries. Section V presents the proposed anomaly detection framework in detail. Section VI presents empirical proof of the proposed framework's effectiveness for detecting anomalies in real-world networks using several assessment measures. Finally, in Section VII, we come to a logical conclusion.

II. RELATED WORK

Traditional anomaly detection and attributed anomaly detection are discussed in relation to each other in this section.

A. TRADITIONAL ANOMALY DETECTION

In traditional anomaly detection, the focus is on identifying outliers within Euclidean structural data, such as tables or images [39]. These methods generally fall into two primary categories. The first category involves one-class classification-based methods [19], [40], [41], which aim to encompass normal data within a defined hyperplane or hypersphere. Data points outside of this boundary are considered outliers. The second category is centered on reconstruction-based methods [17], [42], [43], [44]. These methods operate on the premise that anomalies cannot be accurately reconstructed from a compressed, low-dimensional space. Techniques like auto-encoder based methods use reconstruction errors to pinpoint outliers. These methods work well in their own areas, but they don't work well with non-Euclidean graph data. This shows that finding anomalies in attributed networks is still a problem that needs to be fully solved.

B. ATTRIBUTED NETWORK EMBEDDINGS

Attributed networks, commonly found in real-world scenarios, are analyzed using methods that combine topological structures and node attributes. These approaches can be categorized into three strategies: random-walk-based, matrix-factorization-based, and deep-learning-based methods [45].

Random-Walk-Based Methods: SANE (Sparse Attributed Network Embedding) creates node sequences through random walks and utilizes an attention mechanism for information aggregation from neighboring nodes, facilitating the learning of low-dimensional features [46]. Text-Associated DeepWalk (TADW) enhances DeepWalk by incorporating text features [47]. HSCA (Homophily, Structure, and Content Augmented) extends TADW by introducing a regularization mechanism for neighboring nodes to better capture network homogeneity [48].

Matrix-Factorization-Based Methods: AANE (Accelerated Attributed Network Embedding) [49] and BANE (Binarized Attributed Network Embedding) [50]merge node attributes with edge information to understand the connections between structural and attribute data. To solve the problem of attributed graph clustering, WSNMF [51] approach uses the similarity of node attributes to calculate a weight matrix. To maintain the geometric structure of data points and to detect irrelevant characteristics and data outliers, this technique integrates sparsity restrictions and graph regularization. They also confirm algorithmic convergence and offer an updated strategy to handle optimization complexity. A new model for attributed graph clustering based on Nonnegative Matrix Factorization (NMF) is presented in [52]. They eliminated background noise in the clustering process by applying Symmetric NMF and NMF. To deal with the problem of heterogeneity in the partitions, a novel regularization term is introduced that uses pairwise similarity spaces to include complementary information from the attribute partitions into the structure. Requirements for orthogonality on found communities promote the portrayal of separate, non-overlapping groups.

These methods assume a node's embedding is a linear combination of its neighbors' embeddings, which can be a simple and effective way to represent nodes in a network. The disadvantage is that these methods may not capture complex, non-linear relationships between nodes.

Deep-Learning-Based Methods: These have gained traction for learning node embeddings. HNE (Heterogeneous Network Embedding) is notable for mapping nodes and attributes into a shared latent space [53]. CSAN

(Co-embedding for Static Attributed Networks) uses a Variational Autoencoder (VAE) in a joint learning framework to extract embeddings [54]. DeepWalk based methods revolutionize graph embedding learning with deep learning, which can capture complex patterns in the data but these methods do not perform well on sparse data or data with complex structures.

C. ANOMALY DETECTION ON ATTRIBUTED GRAPHS

Node anomaly detection has been completely transformed by the introduction of deep learning and its utilization in graph data settings [55]. Deep learning-based graph anomaly detection (GAD) approaches outperform more conventional approaches because they more accurately and efficiently capture the intricate connections and structures observed in attributed networks [56]. For the purpose of detecting abnormal nodes, AE has recently been popular [7], [17], [28], [42], [57], [58]. Reconstruction errors are used as anomaly scores in AEs, meaning that nodes with larger reconstruction errors are seen as more abnormal. This serves as the justification for employing AEs for anomaly identification.

With the graph's structure and node properties, GNNs can learn node embeddings. These learned embeddings have the potential to detect anomalies by capturing intricate patterns. Graph Neural Networks (GNNs) have been more popular for identifying abnormal nodes in networks with attributes [59], [60], [61]. It is important to mention that Graph Neural Networks (GNNs) may be integrated with Autoencoders (AEs), where GNNs fulfill the roles of both the encoder and decoder inside the AE framework.

SES-AD [62] is a hybrid model that does not directly search for anomalies in the original time series. Instead, it projects the raw sequence onto a lower dimensional space. This allows it to quickly identify major abrupt change points in the new space using the dissimilarity vector. Ultimately, the possible abnormalities were identified using a statistical approach. The LRRDS (Local Recurrence Rate based Discord Search) [63] introduced a new computational framework for detecting discords in multivariate time series (MTS) data. LRRDS precisely detects the discrepancies by examining a recurrence plot, which is derived from the initial time series data. A novel approach was utilized to enhance the effectiveness of comparing the distances between two subsequences in pairs.

While surveying the current literature, we came across a significant void in how anomaly detection systems handle both sparse and high-dimensional data simultaneously. Our technique fills this need by providing an all-encompassing answer that is unexplored by existing approaches. We provide new capabilities and insights beyond the state-of-the-art by combining SCCA with RMP, which establishes a new standard for anomaly identification in attributed networks.

III. NOTATIONS AND PROBLEM STATEMENT

In this section, we describe the common notations used in this paper. Table 1. summarizes the most significant notations.

Problem Statement: For a given attributed network G with X and A as the node attributed matrix and adjacency matrix, respectively, anomaly detection for an attributed network is to find and rank all the rare nodes according to how they differ markedly from most of the other reference nodes from the perspective of both the attribute information and the topological structure.

TABLE 1. Notations.

Notation	Description
G	The graph of the attributed network.
V	The set of nodes in the attributed network.
A	The adjacency matrix.
X	The node feature matrix.
$\hat{\mathrm{x}}_{\mathrm{i}}^{[\mathrm{M}]}$	The masked node feature vector of x_i.
m_{i}	The masking vector.
$\tilde{\mathrm{x}}_{\mathrm{i}}^{[\mathrm{M}']}$	The padded node feature vector.
ϵ	A vector of random noise used for padding.
$\mathrm{H}_{\mathrm{attribute}}$	The latent space embedding of the attributes.
$\mathrm{H}_{\mathrm{structure}}$	The latent space embedding of the structure.
U	Canonical variable for the transformed attributes space.
V	Canonical variable for the transformed structure space.
$\mathrm{W}_{\mathrm{attribute}}$	Transformation matrix for node attributes.
$\mathrm{W}_{\mathrm{structure}}$	Transformation matrix for network structure.
L_{KL}	Kullback-Leibler divergence loss.

IV. PRELIMINARIES

A. RANDOM MASKING AND PADDING

Random masking is generally used to make the learning algorithm robust and improve generalization by forcing the approach to be less dependent on any special features. Padding is responsible for making changes to the input by adding extra data, either in the form of constant values or noise. The added noise or extra data works as a regularizer, resulting in the avoidance of overfitting. Random masking and padding are used in deep learning models for data preparation because feature masking and padding enable the model to capture the hidden patterns of the data, resulting in more robust and improved anomaly detection [64].

B. GRAPH CONVOLUTIONAL NETWORKS (GCN)

Graph Convolutional Networks (GCNs) are a type of neural network designed to operate directly on graphs. The Graph Convolutional Network (GCN) demonstrates the structure and relationships between features and nodes using the node adjacency matrix A and the feature matrix X. The method employs spectral convolution to perform the convolutional operation on graph data, resulting in the production of the transformation through the formula:

$$\mathrm{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathrm{H}^{(l)}\mathbf{W}^{(l)}\right). \tag{1}$$

where $\mathrm{H}^{(l)}$ and $\mathrm{H}^{(l+1)}$ are the convolutional input and output respectively in the layer l. $\mathrm{W}^{(l)}$ is the layer-specific trainable weight matrix, and σ denotes the activation function, and we

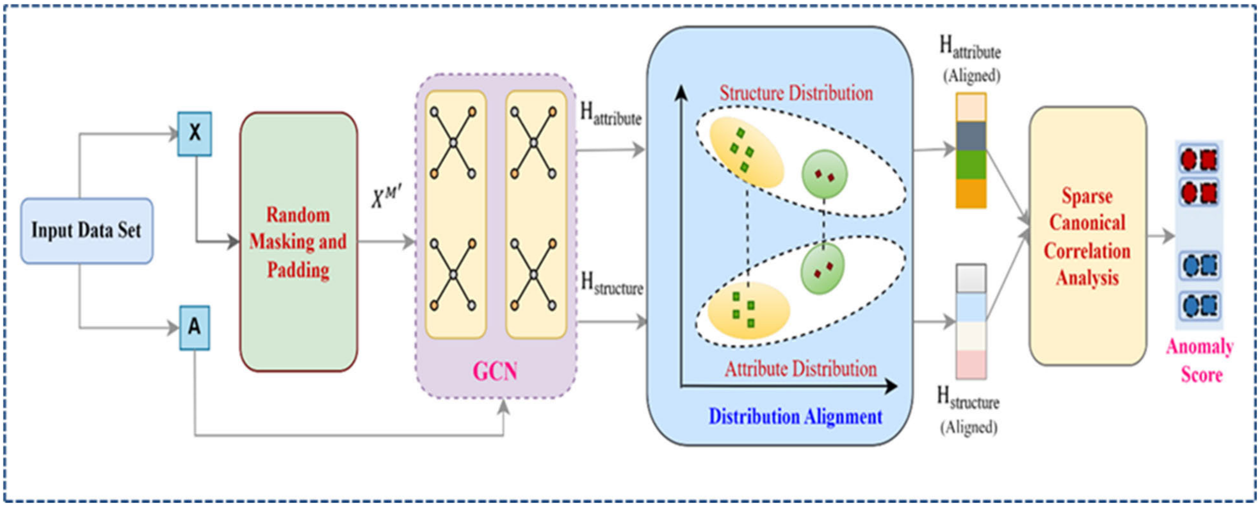


FIGURE 1. Proposed anomaly detection model.

chose ReLU function, $\mathrm{ReLU}(x) = \max(0,x)$. The network begins with $H^{(0)} = \mathbf{X}$, and $H^{(1)} = \sigma\left(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}^{(0)}\right)$.

C. KULLBACK-LEIBLER (KL) DIVERGENCE

The Kullback-Leibler divergence score, also known as the KL divergence score, measures the extent to which one probability distribution deviates from another probability distribution. Mathematically, the KL divergence between two distributions A and B is defined as:

$$\mathrm{D_{KL}}\,(\mathrm{A} \parallel \mathrm{B}) = \sum_{\mathrm{x}} \mathrm{P}\,(\mathrm{x}) \log \frac{A\,(\mathrm{x})}{B\,(\mathrm{x})} \tag{2}$$

D. SPARSE CANONICAL CORRELATION ANALYSIS (SCCA)

Sparse Canonical Correlation Analysis (SCCA) differs from the CCA as it includes sparsity constraint [65]. SCCA introduces an extra layer of complexity by forcing the model to select fewer variables, while CCA tries to detect the linear relationships between two sets of variables by finding pairs of maximally correlated linear combinations (canonical variables) from each set. SCCA helps with high-dimensional data analysis by emphasizing the most important variables, improving interpretability, and decreasing overfitting. Let $X \in \mathbb{R}^{a\times b}$ and $Y \in \mathbb{R}^{a\times c}$ be two matrices that represent two sets of variables, where a is the number of observations, and

b and c are the numbers of variables in each set. The objective in SCCA is to find two vectors $\mathbf{w}_x \in \mathbb{R}^a$ and $\mathbf{w}_y \in \mathbb{R}^b$ that maximize the correlation between the projections of X and Y along these vectors, while also being sparse. Mathematically, this is formulated as:

$$\begin{aligned} &\max_{\mathbf{w}_x,\mathbf{w}_y} \mathrm{corr}\left(X\mathbf{w}_x, Y\mathbf{w}_y\right) \\ &\text{subject to:} \quad \|\mathbf{w}_x\|_2 \leq 1,\ \left\|\mathbf{w}_y\right\|_2 \leq 1,\ \|\mathbf{w}_x\|_0 \leq k_x, \\ &\qquad\qquad\quad \left\|\mathbf{w}_y\right\|_0 \leq k_y \end{aligned} \tag{3}$$

Here, corr denotes the correlation, $\|\cdot\|_2$ is the L2 norm (Euclidean norm), enforcing the vectors to be unit-normed, and $\|\cdot\|_0$ is the LO norm, indicating the number of nonzero elements in the vectors, capped by k_x and k_y. The LO norm introduces the sparsity constraint, limiting the number of variables that contribute to each canonical variable.

V. PROPOSED FRAMEWORK

The proposed anomaly detection framework for attributed networks is described in this section, which combines random masking and padding with the Canonical Correlation Analysis approach. We offered a new way to combine SCCA and RMP, a technique that has never been used before for anomaly detection in attributed networks. Because of this integration, high-dimensional and sparse data can be handled more effectively, which greatly improves the identification of small abnormalities. Applying SCCA allows for the derivation of significant correlations between variables, while RMP ensures robust anomaly identification in tough settings by addressing the inherent sparsity and unpredictability of network data.

Figure 1 is an illustration of the proposed framework pipeline. It provides a visual representation of the core components of our proposed anomaly detection framework, illustrating the methodological innovations and the flow of data processing steps. Random masking and padding components illustrate the initial data preprocessing phase, where we apply Random Masking and Padding to the attributed network data. The GCN component takes the node features and the graph structure as input and produces a node-level embedding that captures both the features and the structure of the graph. In the distribution alignment component, node attributes and graph structure latent space representations are measured and minimized via KL Divergence. The SCCA component is used for extracting and correlating the latent features from both the network structure and node attributes, underscoring its role in uncovering subtle, complex anomalies that other methods might overlook. Finally, in the anomaly scoring mechanism, anomalies are identified based on the deviations in the correlated feature space generated by SCCA.

Firstly, we apply random masking and padding to each node in the attributed network. Then, masked features are fed as input to the weight-sharing GCN that provides $\mathrm{H_{attribute}}$ and $\mathrm{H_{structure.}}$, which are the encoded representation of node attributes, and encoded representation of the graph structure, respectively. After that, ensuring that the latent space representations of the node attributes and the graph structure, are aligned in a meaningful way, we did distribution Alignment with KL divergence. Finally, SCCA is applied to the latent space representations of the node attributes ($\mathrm{H_{attribute}}$) and the graph structure ($\mathrm{H_{structure}}$) and it aims to find pairs of linear combinations of the two sets of variables (in this case, the latent representations) that are maximally correlated with each other, and then anomalies are detected by measuring how much individual data points deviate from the established canonical correlation.

A. RANDOM MASKING AND PADDING

Each node in the network is associated with a feature vector. Let's consider $\mathrm{x_I}$ as the feature vector for node i, where $\mathrm{x_I}$ is a part of the node feature matrix X. To apply random masking, for each element in the feature vector $\mathrm{x_I}$, a corresponding masking vector $\mathrm{m_I}$ is generated. The elements of $\mathrm{m_i}$ are drawn from a Bernoulli distribution, which means each element is randomly set to 1 (keep the feature) or 0 (mask the feature) with a certain probability p. The masked feature vector $\mathrm{\hat{x}_i^M}$ is then obtained by element-wise multiplication (denoted by $\odot$) of the original feature vector $\mathrm{x_i}$ with the masking vector $\mathrm{m_i}$, as follows:

$$\mathrm{\hat{x}_i^M = x_i \odot m_i} \tag{4}$$

After the masking of feature vectors, random noise ϵ is added to reduce the overfitting to the unmasked nodes. This random noise is generally sampled from a gaussian distribution. We get $\mathrm{x_i^{M'}}$ as the feature vector after padding the noise to the masked feature vector:

$$\mathrm{x_i^{M'} = \hat{x}_i^M + \epsilon} \tag{5}$$

It works as a regularizer that ensures that model not to be dependent on any single feature heavily, thus making the generalization of the model better.

B. ENCODING WITH SHARED WEIGHT GCN

The GCN takes the node features and the graph structure as input and produces a node-level embedding that captures both the features and the structure of the graph. In our case, $X^{M'}$ is the node feature matrix after random masking and padding, and A is the adjacency matrix of the graph depicting the graph structure. The GCN takes two inputs: the feature matrix $X^{M'}$ which is the result of masking and padding, and the adjacency matrix A. The shared weights approach uses the same $W^{(l)}$ across different encodings (masked/padded features and structural information). The shared weights enable the model to learn a unified set of parameters that can effectively encode both the features and the structure of the graph. We have used two-layer GCN as it allows for each node's representation to encapsulate not just its immediate neighbors (first-order), but also the neighbors of its neighbors (second-order), capturing a richer and more complex structure of the network. The process can be split into the following steps:

1. The GCN processes the padded feature matrix $X^{M'}$ through two layers as described above. After the last layer, the resulting matrix $H_{\text{attribute}}$ is the encoded representation of the node features. Initialize $H^{(0)}_{\text{attribute}} = X^{M'}$.

$$\mathrm{H^{(1)}_{attribute}} = \sigma\left(\mathrm{\tilde{A}H^{(0)}_{attribute}W^{(0)}}\right) \tag{6}$$

$$\mathrm{H^{(2)}_{attribute}} = \sigma\left(\mathrm{\tilde{A}H^{(1)}_{attribute}W^{(1)}}\right) \tag{7}$$

Here, $\mathrm{\tilde{A}}$ is the modified adjacency matrix (typically $\mathrm{\tilde{A} = A+I_n}$), where I_n is the identity matrix of size $n \times n$, and $\mathrm{W^{(0)}}, \mathrm{W^{(1)}}$ are the weight matrices for the first and second layers, respectively. The final output after the third layer, $\mathrm{H^{(2)}_{attribute}}$, is the encoded representation of node attributes, $\mathrm{H_{attribute}}$.

2. To encode purely structural information, we used the same GCN layers with shared weights $W^{(l)}$, but this time the input $H^{(0)}$ would be different to emphasize structure. Initialize $H^{(0)}_{\text{structure}} = A$.

$$\mathrm{H^{(1)}_{structure}} = \sigma\left(\mathrm{\tilde{A}H^{(0)}_{structure}W^{(0)}}\right) \tag{8}$$

$$\mathrm{H^{(2)}_{structure}} = \sigma\left(\mathrm{\tilde{A}H^{(1)}_{structure}W^{(1)}}\right) \tag{9}$$

The final output after the second layer, $H^{(2)}_{\text{structure}}$, is the encoded representation of the graph structure, $H_{\text{structure}}$. In this two-layer model, the weights $W^{(0)} \text{and} W^{(1)}$, are shared between the processes of calculating $H_{\text{attribute}}$ and $H_{\text{structure}}$, implying that both types of representations are learned simultaneously using the same parameters.

C. DISTRIBUTION ALIGNMENT WITH KL DIVERGENCE

Here, KL Divergence aligns the latent representations distributions for node attributes ($\mathrm{H_{attribute}}$) and for graph structure ($\mathrm{H_{structure}}$). Optimizing distribution alignment for anomaly detection is the major goal of this stage. Node attributes and graph structure latent space representations are measured and minimized via KL Divergence. The mathematical expression for KL divergence regularization is:

$$\mathrm{L_{KL} = D_{KL}\left(P\left(H_{attribute}\right) \parallel \mathit{P}\left(H_{structure}\right)\right)} \tag{10}$$

The probability distributions of latent space representations for node attributes and graph structure are $\mathrm{P\left(H_{attribute}\right)}$ and $\mathrm{P\left(H_{structure}\right)}$. Standardizing latent space representation scale makes them more comparable. SCCA, the framework's next level, requires this phase to prepare representations. The methodology consistently compares Latent Space representations of attributes and network structure by aligning their distributions. This is essential for graph anomaly detection and pattern understanding.

Algorithm 1 Detecting Anomalies in Attributed Networks through Sparse Canonical Correlation Analysis combined with Random Masking and Padding

```
Input: Attributed Network $G$ with node attribute matrix $X$ and adjacency matrix $A$.
Number of training epochs $E$.
Learning rate $\eta$.
Weighting parameters $\alpha, \beta, \gamma, \lambda$.
Output: Anomaly Scores for each node in G.
Initialization: Initialize weights $W$ for GCN layers.
BEGIN
for epoch = 1 to $E$ do
  Generate binary mask matrix M with probability p by Eq. (5).
  Generate noise matrix $\varepsilon$ with standard deviation $\sigma$.
  Mask and pad features: $X \leftarrow (X \odot M) + \varepsilon$ by Eq. (6).
  Encode attributes and structure using single GCN:
      Compute $H_{\text{attribute}} = \text{GCN}(X', A)$ by Eq. (8).
      Compute $H_{\text{structure}} = \text{GCN}(A)$ by Eq. (9).
  Smooth the distribution using KL divergence by Eq. (11).
  Compute SCCA to maximize correlation between $H_{\text{attribute}}$ and $H_{\text{structure}}$ and
   compute canonical variables U and V by Eq. (12) and Eq. (13).
  Update parameters by optimizing the Loss Function $L$ given in Eq. (15).
End for
Compute anomaly scores based on deviations in canonical correlations for each
 node $i$ by Eq. (14).
END
```

D. SPARSE CANONICAL CORRELATION ANALYSIS (SCCA)

The anomaly detection process is based on the correlation between the latent representations of node attributes ($\text{H}_{\text{attribute}}$) and graph structure ($\text{H}_{\text{structure}}$) obtained from SCCA. Given two sets of variables, $\text{H}_{\text{attribute}}$ and $\text{H}_{\text{structure}}$, with dimensions $n \times p$ and $n \times q$ respectively, where n is the number of nodes and p, q are the dimensions of the attribute and structure latent spaces, following are the steps to compute the canonical variables. Now, the goal of SCCA is to determine two sets of canonical vectors $\mathbf{w}_{\text{attribute}} \in \mathbb{R}^p$ and $\mathbf{w}_{\text{structure}} \in \mathbb{R}^q$ that maximize the correlation between $H_{\text{attribute}}\mathbf{w}_{\text{attribute}}$ and $H_{\text{structure}}\mathbf{w}_{\text{structure}}$, subject to sparsity constraints. The optimization problem can be stated as:

$$\max_{\mathbf{w}_{\text{attribute}}, \mathbf{w}_{\text{structure}}} corr \begin{pmatrix} H_{\text{attribute}}\mathbf{w}_{\text{attribute}}, \\ H_{\text{structure}}\mathbf{w}_{\text{structure}} \end{pmatrix}$$

$$\text{subject to: } \|\mathbf{w}_{\text{attribute}}\|_2 \leq 1, \|\mathbf{w}_{\text{structure}}\|_2 \leq 1, \|\mathbf{w}_{\text{attribute}}\|_1 \leq c_{\text{attribute}}, \|\mathbf{w}_{\text{structure}}\|_1 \leq c_{\text{structure}} \quad (11)$$

The L1 norm constraints $\|\mathbf{w}_{\text{attribute}}\|_1 \leq c_{\text{attribute}}$ and $\|\mathbf{w}_{\text{attribute}}\|_1 \leq c_{\text{structure}}$ enforce sparsity in the canonical vectors, ensuring that each vector utilizes only a limited number of significant features from its respective feature set. The parameters $c_{\text{attribute}}$ and $c_{\text{structure}}$ control the level of sparsity. The optimization problem in SCCA can be challenging due to the sparsity constraints. We have used a common approach to solve it is through alternating least squares (ALS), which iteratively optimizes one variable while keeping the other fixed. The output of SCCA consists of sparse canonical variables:

$$U = H_{\text{attribute}}\mathbf{w}_{\text{attribute}} \quad (12)$$

$$V = H_{\text{structure}}\mathbf{w}_{\text{structure}} \quad (13)$$

E. ANOMALY DETECTION

Anomalies can be detected by measuring how much individual data points deviate from the established canonical correlation. We compute the anomaly scores based on the correlation of the canonical variables:

$$\text{Anomaly Score }(i) = 1 - \frac{U_i \cdot V_i}{\|U_i\|\|V_i\|} \quad (14)$$

Here, U_i and V_i are the canonical variable scores for node i. A higher anomaly score suggests a greater deviation from the expected canonical correlation, indicating a potential anomaly.

F. OBJECTIVE FUNCTION

The objective function for the proposed framework is defined as:

$$L = \alpha \cdot \text{Recon}(X, \hat{X}) + \beta \cdot D_{KL}(P\|Q) - \gamma \\ . \text{corr}\left(H_{\text{attribute}}\ \mathbf{w}_{\text{attribute}}\ , H_{\text{structure}}\ \mathbf{w}_{\text{structure}}\ \right) \\ + \lambda\left(\|\mathbf{w}_{\text{attribute}}\ \|_1 - \|\mathbf{w}_{\text{structure}}\ \|_1\right) \quad (15)$$

where:

- α, β, and γ are weighting parameters that balance the contribution of each term in the loss function.
- The first term, $\text{Recon}(X, \hat{X})$, is the reconstruction loss from the GCN.
- The second term, $D_{KL}(P \parallel Q)$, is the KL divergence for distribution alignment.
- $-\text{corr}\left(H_{\text{attribute}}\mathbf{w}_{\text{attribute}}, H_{\text{structure}}\mathbf{w}_{\text{structure}}\right)$, represents the negative correlation in SCCA, aiming to maximize the correlation while the negative sign converts it into a minimization problem.
- The final term, $\lambda\left(\|\mathbf{w}_{\text{attribute}}\|_1 + \|\mathbf{w}_{\text{structure}}\|_1\right)$, is the L1 regularization from SCCA promoting sparsity.

VI. EXPERIMENTS

The performance of our proposed framework on various datasets is discussed in this part. Two of the most important evaluation tasks are anomaly detection performance analysis and model parameter sensitivity analysis. The datasets are initially described in detail in this section. After that, the proposed framework is compared to the other baseline techniques, and the anomaly detection accuracy is given, as well as a comparison of the experimental data and analysis. Finally, we examine the experimental parameters' sensitivity.

A. DATASETS

To assess the performance of the proposed framework, we conducted evaluations using five widely recognized public datasets. This included three citation benchmarks

(Cora, Citeseer, and PubMed) and two social benchmarks (BlogCatalog, and Flickr). The particulars of these datasets are as outlined below:

Citation Networks: The datasets Cora, Citeseer, and PubMed constitute three commonly utilized benchmarks in citation networking. In these datasets, nodes represent scientific papers, and edges denote the citations among them.

Social Networks: The BlogCatalog, and Flickr datasets serve as benchmarks for social networking. These represent users as nodes and their follow relationships as edges.

TABLE 2. The statistics of the datasets.

Dataset	Nodes	Edges	Attributes
Cora	2,708	5,429	1,433
Citeseer	3,327	4,732	3,703
PubMed	19,717	44,338	500
BlogCatalog	5,196	121,743	8,189
Flickr	7,575	239,738	12,407

B. EVALUATION INDICATORS

This paper evaluates the contribution of different anomaly detection methods using two commonly used evaluation indicators that have been extensively used in earlier anomaly detection methods [66], [67], [68].

1) ROC-AUC

An ROC curve, also known as a receiver operating characteristic curve, is a graphical representation that illustrates the performance of a classification model over various thresholds. An ROC curve illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) across various categorization criteria. Decreasing the classification threshold results in the categorization of a greater number of items as positive, hence increasing the occurrence of both False Positives and True Positives. AUC (Area under the ROC Curve) assesses the total two-dimensional region under the total ROC curve. A higher AUC suggests a better anomaly detection system. AUC 1 denotes perfect classification, whereas 0.5 shows random chance-like differentiation. To conclude, the ROC curve and AUC assess an anomaly detection system's ability to distinguish normal from abnormal occurrences.

2) AVERAGE PRECISION (AP)

Average Precision (AP) is generally used where prediction ranking is considered more important than the individual scores. It adds the precision value calculated at each threshold value while ranking the prediction. A high AP value indicates that the anomaly detection system efficiently prioritizes anomalies over regular events, guaranteeing excellent accuracy even when recall rates change. AP can be more informative in imbalanced datasets, where the number of anomalies is much lower than the number of normal instances.

C. BASELINES

Our proposed framework is compared to the following techniques to demonstrate its ability to detect anomalies:

- One-Class Support Vector Machine (OC-SVM) [69]: A classical anomaly detection algorithm employing a hyperplane to identify anomalies.
- DOMINANT [16]: An advanced unsupervised deep learning approach that integrates Graph Convolutional Networks (GCN) with a deep Autoencoder. This method reconstructs the attributed network through both topological structure and node attributes, enabling anomaly detection.
- Adversarially Regularized Graph Autoencoder (ARGA) [70]: An adversarial graph embedding framework that builds upon Graph Autoencoders (GAE). It enforces the embeddings of topological structure and node attributes to conform to a prior distribution through adversarial training.
- ResGCN [20]: Rather than relying on reconstruction errors, ResGCN generates residual information from the input network to rank anomalies. It combines GCN for capturing network sparsity and nonlinearity, a deep neural network for residual information aggregation, and a residual-based attention mechanism to mitigate the influence of anomalous nodes.
- CCA-SSG (Canonical Correlation Analysis to Self-Supervised GNN) [23]: A self-supervised graph embedding learning model that utilizes canonical correlation analysis to link two view embeddings through data augmentation. It also discards augmentation-variant information, preventing degenerate solutions.
- ARISE [71]: A method for finding unusual patterns in networks with attributes. Unlike other methods, it concentrates on specific patterns in the network to identify abnormalities. A region proposal module is used to identify dense patterns in the network as suspicious areas. The average similarity between pairs of nodes indicates the degree of abnormality in the pattern. Graph contrastive learning scheme was also introduced to identify attribute anomalies.

D. EXPERIMENTAL DESIGN

In the experiment, we implemented proposed framework on Python, and trained it with 200 training epochs for all the datasets. Training for less than 200 epochs do not allow the model sufficient time to learn from the complexity of the data, because our datasets involve high-dimensional features. This can lead to underfitting, where the model fails to capture essential relationships. On the other hand, training for more than 200 epochs lead to poor model performance due to the overfitting which negatively impacts performance on new, unseen data. We used Python 3.8, leveraging widely

TABLE 3. AUC and AP results in percentages (%) of all methods based on five datasets.

Methods	Cora		Citeseer		BlogCatalog		Flickr		PubMed	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
OC-SVM	84.32	81.94	74.66	74.03	88.08	88.97	83.06	84.98	79.86	78.26
DOMINANT	80.34	75.35	73.23	67.46	88.71	90.31	85.78	87.56	84.32	82.73
ARGA	82.95	81.03	76.42	74.89	86.46	87.47	84.12	85.04	81.76	84.5
ResGCN	93.85	90.68	81.73	79.57	92.33	93.63	**94.05**	**96.43**	88.42	84.95
CCA-SSG	90.04	86.54	80.35	74.61	93.65	89.23	89.82	87.31	88.95	89.71
ARISE	94.02	91.44	82.58	79.92	95.57	94.05	92.45	90.52	89.64	**92.62**
PROPOSED	**96.43**	**95.01**	**84.41**	**81.73**	**97.41**	**96.16**	93.58	94.98	**91.92**	92.04

recognized libraries for machine learning and graph analysis. TensorFlow was utilized for constructing and training deep learning models. Pandas and NumPy were used for data manipulation and numerical computations, respectively, helping in the handling of dataset attributes and matrix operations.

For optimization, the Adam algorithm with a learning rate of 0.001 is being used. Using the Adam optimization algorithm with a learning rate of 0.001 is a common and effective choice for training deep learning models, including those involved in anomaly detection frameworks. The embedding dimension has been fixed at 64 for all the datasets. We tried other dimensions also, starting from 16 to 256, but we got the best result at 64.

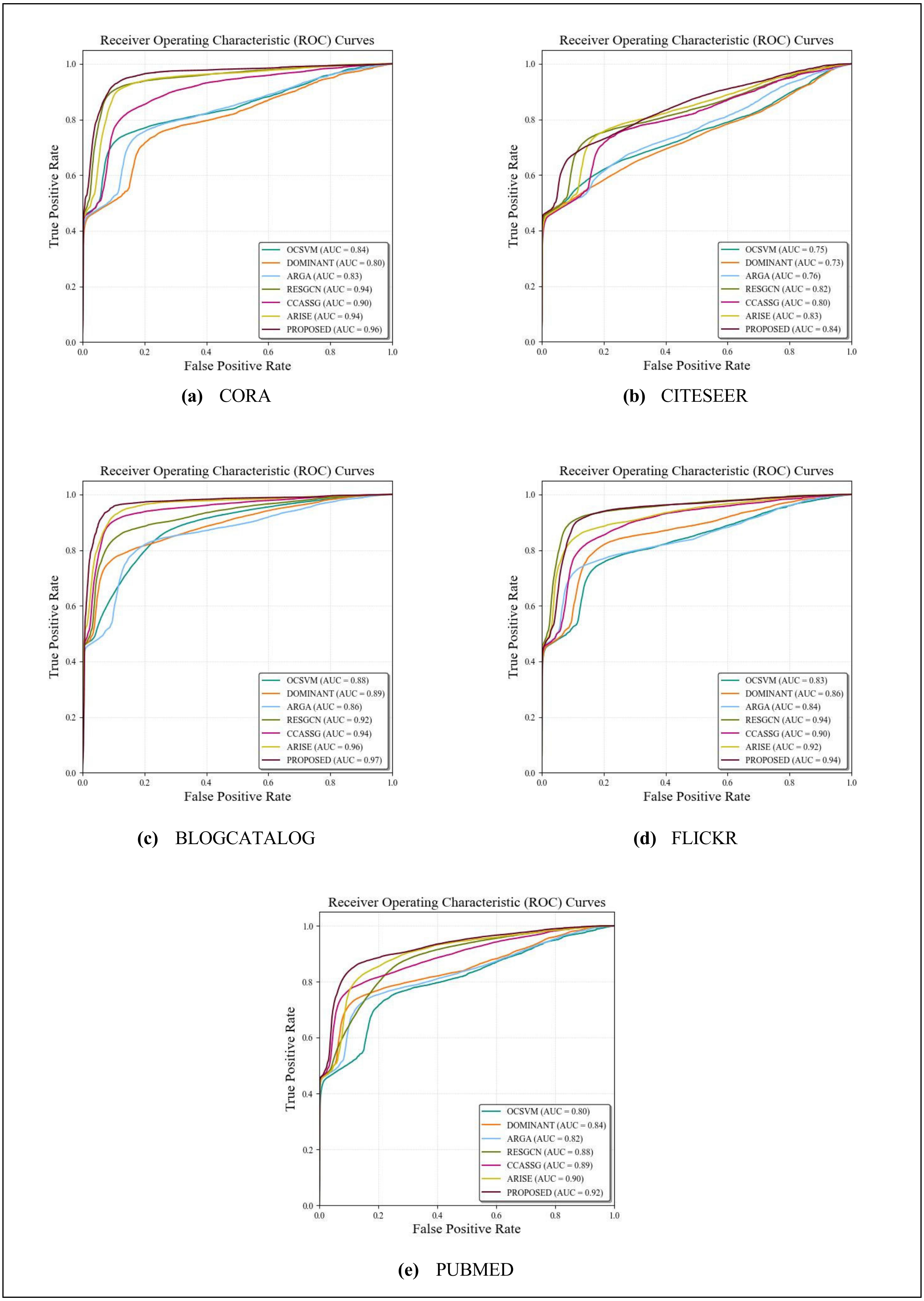
We used the grid search for finding the optimum values for α (Reconstruction Loss Weight), β (KL Divergence Weight), γ (Correlation Maximization Weight), and λ (L1 Regularization Weight). After performing grid search, we found that $\alpha = 0.1$, $\beta = 0.01$, $\gamma = 0.5$, and $\lambda = 0.01$ yielded the best performance, balancing the contributions of reconstruction loss, KL divergence, correlation maximization, and sparsity effectively for our problem. We iterated over all possible combinations of α, β, γ, and λ within the defined ranges (0.01 to 1).

We use the publicly accessible implementations from the source publications for the baseline techniques. The experiments were conducted on a workstation equipped with an Intel Core i7 processor, 16GB RAM, and an NVIDIA GeForce RTX 3080 GPU. This setup ensured the efficient processing of computationally intensive tasks, including model training and evaluation.

E. EXPERIMENTAL RESULTS

The quantitative evaluation of the proposed approach involved a comparison with the state-of-the-art methods. The outcomes of all methods on four datasets with respect to ROC-AUC and average precision values are displayed in Table 3. In addition, the ROC curve comparison of all the methods for all datasets are demonstrated in Figure 2.

An analysis of all datasets demonstrates that our proposed approach surpasses all existing baseline methodologies in terms of performance except the Flickr dataset. Our model obtains a significant improvement of 3.76% on AUC and 4.33% on AP compared to the second-best results in the baseline. The AUC and AP performance comparison is shown in Figure 3 and Figure 4, respectively. The OC-SVM technique, which is based on one-class graph neural networks, does not demonstrate competitive performance despite being specifically developed for extracting graph structure features and learning hyperspheres for anomaly identification. This is because aberrant nodes with a latent space comparable to normal nodes cannot be measured using the one class technique. Dominant integrates attribute and structural data for node embedding, however autoencoder-based approaches that rely on reconstruction errors fail to provide a sufficient metric for abnormality detection. The ARGA technique utilizes a variational autoencoder to acquire node embeddings, resulting in a more distinct separation in the latent space. However, the use of reconstruction error as an anomaly score renders it incapable of detecting abnormal nodes that resemble normal nodes. The self-supervised approach such as CCA-SSG employ graph embedding to enhance data augmentation by using past human expertise, although they may not encompass all exceptional patterns. This renders it incapable of detecting anomalous nodes that exhibit patterns similar to those of regular nodes. ResGCN outperforms all the other baseline methods except our model due to the attention based deep residual modeling approach. Unlike other baselines, our approach uses random masking and padding as a preprocessing step for node attributes before feeding them into the GCN that helps the model in learning to identify essential features, and regularization provided by the SCCA reduces the possibility of overfitting. The core idea of our approach is that it focuses on capturing the correlations between the structure of a network and the attributes of its nodes. When a node is abnormal, the network embedding will differ from the node attribute embedding. In contrast, other baselines primarily focus on learning node representation in a latent space,



(a) CORA

(b) CITESEER

(c) BLOGCATALOG

(d) FLICKR

(e) PUBMED

FIGURE 2. ROC curves comparison on (A) CORA (B) Citeseer (C) Blogcatalog (D) Flickr, and (E) Pubmed. The area under curve is larger, the anomaly detection performance is better.

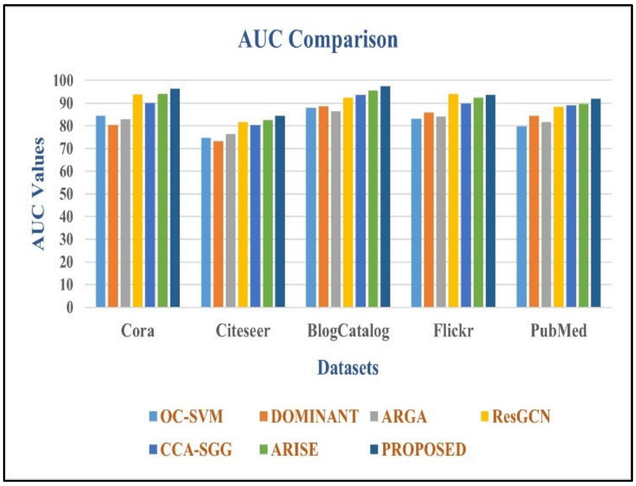


FIGURE 3. AUC comparison.

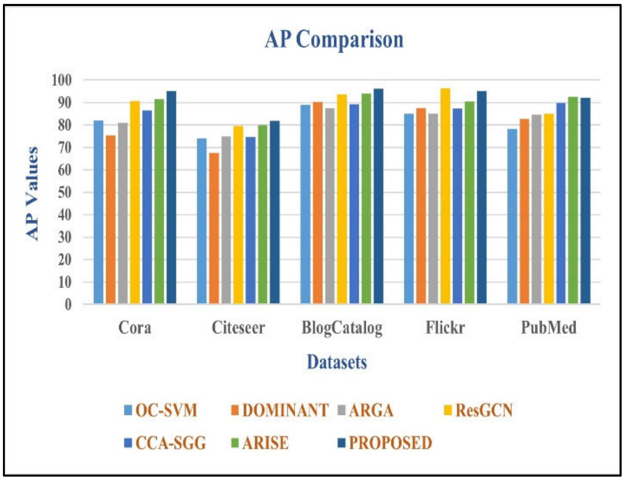


FIGURE 4. AP comparison.

disregarding the relationship between network topology and node properties.

The results of our empirical study show that our methodology outperforms the conventional approaches. Our technique demonstrated a significant improvement in detection accuracy, as indicated by better AUC and AP measures, during rigorous testing on benchmark datasets. The comparisons of the ROC curves, shown in Figure 2, demonstrate how our technique is more sensitive and specific. These findings validate our methodology's originality and efficacy in detecting outliers in attributed networks. Our results have two important consequences. Our research adds to the literature on anomaly identification in complex networks by providing a new approach to the problem that takes into account the high dimensionality and attribute sparsity. Our technique has the potential to greatly enhance anomaly identification in a variety of real-world applications, from cybersecurity to social network research, highlighting the work's broad applicability and effect. Our experiments on datasets of varying sizes demonstrate the proposed method's superior scalability. The ability to maintain high detection accuracy, even as dataset size increases, underscores our method's suitability for large, complex networks.

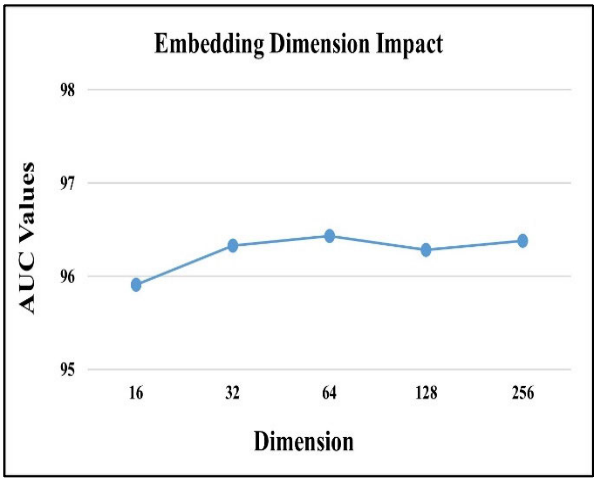


FIGURE 5. Embedding dimension impact.

The thorough comparison investigation confirms our work's innovative capabilities. Our approach establishes a new standard for detecting anomalies in attributed networks by exhibiting higher performance in terms of efficiency, scalability, resilience, and accuracy. These results support our theories and add to the evidence supporting the theoretical and applied significance of our study.

F. PARAMETER SENSITIVITY

This section examines the effects of various node embedding dimensions using the Cora dataset (Figure. 5). Regardless of the dimensionality of the node embedding, our approach maintains stable performance. The reason behind this is that our approach records crucial details on the network topology, and the node attributes that is used for detecting anomalies.

VII. CONCLUSION

An attribute network is a common type of graph data. Existing methods for graph anomaly detection primarily rely on feature extraction; however, they lack a targeted design for anomaly detection tasks and fail to account for the strong or weak association between anomalies and nearby information. As a result, these methods struggle to capture anomalies in attributed networks that are similar to normal ones. Therefore, we propose a novel approach that is based on Sparse Canonical Correlation Analysis combined with Random Masking and Padding. In our approach, random masking and padding is used as a preprocessing step to help mitigate the problem of overfitting by introducing randomness and variability in the training process. After that, we learned the distribution network topology and node characteristics using weight-sharing GCN as an encoder. Then, we employed Kullback-Leibler (KL) divergence to align them to a shared latent space. Then, we use the sparse canonical correlation analysis, which uses sparsity-based regularization and integrates the L1 norm loss for constraining the CCA optimization and reducing the possibility of overfitting. It helps the model to generalize to new data by emphasizing the pertinent alignment of latent space exploration. The goal of canonical correlation analysis is to maximize the correlation of normal nodes with respect to network structure and node characteristics. Finally, an anomaly score is defined as the correlation of two views to detect anomalous nodes. By presenting a new, comprehensive technique that surpasses current methodologies, our research significantly advances the area of anomaly identification in attributed networks. Combining SCCA with RMP is a novel approach that points the way for future study and lays the groundwork for even greater achievements in this field. This innovative approach addresses critical challenges in the field, including the high-dimensionality of data and the sparsity of attributes, which have traditionally hindered the effectiveness of anomaly detection methodologies. Our framework significantly improves the detection of subtle and complex anomalies in attributed networks, as evidenced by our extensive testing across multiple benchmark datasets. The use of SCCA ensures the effective capture of deep correlations within the data, while Random Masking and Padding enhances model robustness against overfitting. The practical applications of our research are broad and impactful, ranging from cybersecurity to social network analysis, where early and accurate anomaly detection can prevent fraudulent activities, identify misinformation spread, and much more. In future research, our focus will be on analyzing the relationship between global and local nodes in order to identify anomalous spots. We will also investigate the connections between nodes and their neighbors, with the aim of improving the model and testing it in particular attribute network anomaly detection instances. We will try to explore the scalability of our methodology to larger networks and its applicability to other types of data beyond attributed networks.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia through Project Number: GSSRD-24.

REFERENCES

[1] E. Nasiri, K. Berahmand, and Y. Li, "Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks," *Multimedia Tools Appl.*, vol. 82, no. 3, pp. 3745–3768, Jan. 2023.

[2] K. Srinivasan, "Graph data management, modeling, and mining," in *Encyclopedia of Data Science and Machine Learning*. Hershey, PA, USA: IGI Global, 2023.

[3] Q. Zhang, J. Dong, Q. Tan, and X. Huang, "Integrating entity attributes for error-aware knowledge graph embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 4, pp. 1667–1682, Apr. 2024, doi: 10.1109/tkde.2023.3310149.

[4] W. Khan and M. Haroon, "A pilot study and survey on methods for anomaly detection in online social networks," in *Human-Centric Smart Computing*, vol. 2022. Cham, Switzerland: Springer, 2022, pp. 119–128.

[5] L.-H. Chen, H. Li, W. Zhang, J. Huang, X. Ma, J. Cui, N. Li, and J. Yoo, "AnomMAN: Detect anomalies on multi-view attributed networks," *Inf. Sci.*, vol. 628, pp. 1–21, May 2023.

[6] M. Shao, Y. Lin, Q. Peng, J. Zhao, Z. Pei, and Y. Sun, "Learning graph deep autoencoder for anomaly detection in multi-attributed networks," *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110084.

[7] W. Khan, M. Haroon, A. N. Khan, M. K. Hasan, A. Khan, U. A. Mokhtar, and S. Islam, "DVAEGMM: Dual variational autoencoder with Gaussian mixture model for anomaly detection on attributed networks," *IEEE Access*, vol. 10, pp. 91160–91176, 2022.

[8] T. Barbariol, "Improving anomaly detection for industrial applications," Tech. Rep., 2023.

[9] R. Ball and L. Drevin, "Anomaly detection using autoencoders with network analysis features," *ORION*, vol. 39, no. 1, pp. 1–44, 2023.

[10] M. Education, W. Khan, and M. Haroon, "An exhaustive review on state-of-the-art techniques for anomaly detection on attributed networks," *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 10, pp. 6707–6722, Apr. 2021. [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/5537/4640

[11] S. Hu and M. Shao, "Dual perspective contrastive learning based subgraph anomaly detection on attributed networks," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2022, pp. 481–493.

[12] Y. Liu, Z. Liu, X. Feng, and Z. Li, "Robust attributed network embedding preserving community information," in *Proc. IEEE 38th Int. Conf. Data Eng. (ICDE)*, May 2022, pp. 1874–1886.

[13] X. Wu, H. Zhang, Y. Quan, Q. Miao, and P. G. Sun, "Graph embedding based on motif-aware feature propagation for community detection," *Phys. A, Stat. Mech. Appl.*, vol. 630, Nov. 2023, Art. no. 129205.

[14] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 701–710.

[15] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 855–864.

[16] 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.

[17] H. Fan, F. Zhang, and Z. Li, "Anomalydae: Dual autoencoder for anomaly detection on attributed networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 5685–5689.

[18] D. Zhu, Y. Ma, and Y. Liu, "DeepAD: A joint embedding approach for anomaly detection on attributed networks," in *Proc. Int. Conf. Comput. Sci.* Cham, Switzerland: Springer, 2020, pp. 294–307.

[19] F. Zhang, H. Fan, R. Wang, Z. Li, and T. Liang, "Deep dual support vector data description for anomaly detection on attributed networks," *Int. J. Intell. Syst.*, vol. 37, no. 2, pp. 1509–1528, Feb. 2022.

[20] Y. Pei, T. Huang, W. van Ipenburg, and M. Pechenizkiy, "ResGCN: Attention-based deep residual modeling for anomaly detection on attributed networks," *Mach. Learn.*, vol. 111, no. 2, pp. 519–541, Feb. 2022.

[21] H. Hojjati and N. Armanfard, "DASVDD: Deep autoencoding support vector data descriptor for anomaly detection," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 10, 2023, doi: 10.1109/TKDE.2023.3328882.

[22] W. Khan and M. Haroon, "An unsupervised deep learning ensemble model for anomaly detection in static attributed social networks," *Int. J. Cognit. Comput. Eng.*, vol. 3, pp. 153–160, Jun. 2022.

[23] H. Zhang, Q. Wu, J. Yan, D. Wipf, and P. S. Yu, "From canonical correlation analysis to self-supervised graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 76–89.

[24] W. Lu, H. Fan, K. Zeng, Z. Li, and J. Chen, "Self-supervised domain adaptation for cross-domain fault diagnosis," *Int. J. Intell. Syst.*, vol. 37, no. 12, pp. 10903–10923, Dec. 2022.

[25] Q. Tan, X. Zhang, X. Huang, H. Chen, J. Li, and X. Hu, "Collaborative graph neural networks for attributed network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 3, pp. 972–986, Mar. 2024.

[26] H. Du, W. Wang, and L. Bai, "Dual-channel embedding learning model for partially labeled attributed networks," *Pattern Recognit.*, vol. 142, Oct. 2023, Art. no. 109644.

[27] H. Kundra, W. Khan, M. Malik, K. P. Rane, R. Neware, and V. Jain, "Quantum-inspired firefly algorithm integrated with cuckoo search for optimal path planning," *Int. J. Modern Phys. C*, vol. 33, no. 2, Feb. 2022, Art. no. 2250018.

[28] K. Zhang, G. Lu, Y. Li, and C. Xu, "A graph autoencoder-based anomaly detection method for attributed networks," in *Proc. 5th Int. Conf. Natural Lang. Process. (ICNLP)*, Mar. 2023, pp. 330–337.

[29] X. Li, C. Xiao, Z. Feng, S. Pang, W. Tai, and F. Zhou, "Controlled graph neural networks with denoising diffusion for anomaly detection," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121533.

[30] S. Patil and S. Bhosale, "A novel feature selection-based algorithm for medical correlation of high dimensional data," in *Proc. IEEE IAS Global Conf. Emerg. Technol. (GlobConET)*, May 2023, pp. 1–7.

[31] X. Zhang, P. Wei, and Q. Wang, "A hybrid anomaly detection method for high dimensional data," *PeerJ Comput. Sci.*, vol. 9, p. e1199, Jan. 2023.

[32] M. Soni, M. A. Shnan, and Y. Bengio, "Scalable neural network algorithms for high dimensional data," *Mesopotamian J. Big Data*, vol. 2023, pp. 1–11, Jan. 2023.

[33] V. Mahalakshmi, A. Balobaid, B. Kanisha, R. Sasirekha, and M. R. Raja, "Artificial intelligence: A next-level approach in confronting the COVID-19 pandemic," *Healthcare*, vol. 11, no. 6, p. 854, Mar. 2023, doi: 10.3390/healthcare11060854.

[34] C. F. G. dos Santos, "Avoiding overfiting: new algorithms to improve generalisation in convolutional neural networks," Tech. Rep., 2022.

[35] M. Li, W. Li, Y. Liu, Y. Huang, and G. Yang, "Adaptive mask sampling and manifold to Euclidean subspace learning with distance covariance representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023.

[36] R. Wang, F. Zhang, X. Huang, C. Tian, L. Xi, and H. Fan, "CaCo: Attributed network anomaly detection via canonical correlation analysis," *IEEE Trans. Ind. Informat.*, vol. 20, no. 1, pp. 461–470, Jan. 2024.

[37] C. Liu, Z. Yang, G. Xiang, and Y. Yu, "A statistical feature-based anomaly detection method for PFC using canonical correlation analysis," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[38] H. Kwon and N. M. Nasrabadi, "Kernel canonical correlation analysis for hyperspectral anomaly detection," *Proc. SPIE*, vol. 6233, pp. 23–30, May 2006.

[39] S. Thudumu, P. Branch, J. Jin, and J. Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," *J. Big Data*, vol. 7, no. 1, pp. 1–30, Dec. 2020.

[40] T. Vaiyapuri, S. Algamdi, R. John, Z. Sbai, M. Al-Helal, A. Alkhayyat, and D. Gupta, "Metaheuristics with federated learning enabled intrusion detection system in Internet of Things environment," *Expert Syst.*, vol. 40, no. 5, Jun. 2023, Art. no. e13138.

[41] M. Kim, J. Kim, J. Yu, and J. K. Choi, "Active anomaly detection based on deep one-class classification," *Pattern Recognit. Lett.*, vol. 167, pp. 18–24, Mar. 2023.

[42] V. H. Son, U. Daisuke, H. Kiyoshi, M. Kazuki, S. Pranata, and S. M. Shen, "Anomaly detection with adversarial dual autoencoders," Feb. 2019, *arXiv:1902.06924*.

[43] A. Roy, "GAD-NR: Graph anomaly detection via neighborhood reconstruction KEYWORDS anomaly detection, graph neural network, auto-encoder ACM reference format," Tech. Rep., 2023, doi: 10.1145/3616855.3635767.

[44] Y. Li, X. Huang, J. Li, M. Du, and N. Zou, "SpecAE: Spectral AutoEncoder for anomaly detection in attributed networks," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2233–2236, doi: 10.1145/3357384.3358074.

[45] M. Xu, "Understanding graph embedding methods and their applications," *SIAM Rev.*, vol. 63, no. 4, pp. 825–853, Jan. 2021.

[46] H. Wang, E. Chen, Q. Liu, T. Xu, D. Du, W. Su, and X. Zhang, "A united approach to learning sparse attributed network embedding," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 557–566.

[47] Z. Ye, H. Zhao, K. Zhang, Y. Zhu, and Y. Xiao, "Text-associated max-margin DeepWalk," in *Proc. 6th CCF Conf., Big Data*, Xi'an, China. Cham, Switzerland: Springer, Oct. 2018, pp. 301–321.

[48] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Attributed network embedding via subspace discovery," *Data Mining Knowl. Discovery*, vol. 33, no. 6, pp. 1953–1980, Nov. 2019.

[49] X. Huang, J. Li, and X. Hu, "Accelerated attributed network embedding," in *Proc. SIAM Int. Conf. Data Mining*, 2017, pp. 633–641.

[50] H. Yang, S. Pan, P. Zhang, L. Chen, D. Lian, and C. Zhang, "Binarized attributed network embedding," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1476–1481.

[51] K. Berahmand, M. Mohammadi, R. Sheikhpour, Y. Li, and Y. Xu, "WSNMF: Weighted symmetric nonnegative matrix factorization for attributed graph clustering," *Neurocomputing*, vol. 566, Jan. 2024, Art. no. 127041, doi: 10.1016/j.neucom.2023.127041.

[52] V. Jannesari, M. Keshvari, and K. Berahmand, "A novel nonnegative matrix factorization-based model for attributed graph clustering by incorporating complementary information," *Expert Syst. Appl.*, vol. 242, May 2024, Art. no. 122799, doi: 10.1016/j.eswa.2023.122799.

[53] E. A. Alhazami and A. M. Sheneamer, "Graph-of-code: Semantic clone detection using graph fingerprints," *IEEE Trans. Softw. Eng.*, vol. 49, no. 8, pp. 3972–3988, Aug. 2023, doi: 10.1109/TSE.2023.3276780.

[54] Z. Meng, S. Liang, X. Zhang, R. McCreadie, and I. Ounis, "Jointly learning representations of nodes and attributes for attributed networks," *ACM Trans. Inf. Syst.*, vol. 38, no. 2, pp. 1–32, Apr. 2020.

[55] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12012–12038, Dec. 2023.

[56] P. Joseph, "Novel techniques using graph neural networks (GNNS) for anomaly detection," Tech. Rep., 2023.

[57] X. Yuan, N. Zhou, S. Yu, H. Huang, Z. Chen, and F. Xia, "Higher-order structure based anomaly detection on attributed networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 2691–2700.

[58] W. Khan and M. Haroon, "An efficient framework for anomaly detection in attributed social networks," *Int. J. Inf. Technol.*, vol. 14, no. 6, pp. 3069–3076, Oct. 2022, doi: 10.1007/s41870-022-01044-2.

[59] X. Wang, B. Jin, Y. Du, P. Cui, Y. Tan, and Y. Yang, "One-class graph neural networks for anomaly detection in attributed networks," *Neural Comput. Appl.*, vol. 33, no. 18, pp. 12073–12085, Sep. 2021.

[60] H. Fan, F. Zhang, Y. Wei, Z. Li, C. Zou, Y. Gao, and Q. Dai, "Heterogeneous hypergraph variational autoencoder for link prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4125–4138, Aug. 2022.

[61] K. Ding, Q. Zhou, H. Tong, and H. Liu, "Few-shot network anomaly detection via cross-network meta-learning," in *Proc. Web Conf.*, Apr. 2021, pp. 2448–2456.

[62] M. Hu, X. Feng, Z. Ji, K. Yan, and S. Zhou, "A novel computational approach for discord search with local recurrence rates in multivariate time series," *Inf. Sci.*, vol. 477, pp. 220–233, Mar. 2019, doi: 10.1016/j.ins.2018.10.047.

[63] Z. Ji, Y. Wang, K. Yan, X. Xie, Y. Xiang, and J. Huang, "A space-embedding strategy for anomaly detection in multivariate time series," *Expert Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117892, doi: 10.1016/j.eswa.2022.117892.

[64] Y. Chen, H. Peng, L. Huang, J. Zhang, and W. Jiang, "A novel MAE-based self-supervised anomaly detection and localization method," *IEEE Access*, vol. 11, pp. 127526–127538, 2023.

[65] M. Kim, E. J. Min, K. Liu, J. Yan, A. J. Saykin, J. H. Moore, Q. Long, and L. Shen, "Multi-task learning based structured sparse canonical correlation analysis for brain imaging genetics," *Med. Image Anal.*, vol. 76, Feb. 2022, Art. no. 102297.

[66] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2152–2158.

[67] Z. Peng, M. Luo, J. Li, H. Liu, and Q. Zheng, "ANOMALOUS: A joint modeling approach for anomaly detection on attributed networks," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3513–3519.

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

[69] P. Nader, P. Honeine, and P. Beauseroy, "L_p-norms in one-class classification for intrusion detection in SCADA systems," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2308–2317, Nov. 2014.

[70] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially regularized graph autoencoder for graph embedding," 2018, *arXiv:1802.04407*.

[71] J. Duan, B. Xiao, S. Wang, H. Zhou, and X. Liu, "ARISE: Graph anomaly detection on attributed networks via substructure awareness," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 22, 2023, doi: 10.1109/TNNLS.2023.3312655.

WASIM KHAN received the B.Tech., M.Tech., and Ph.D. degrees. He is currently an Assistant Professor with the Koneru Lakshmaiah Education Foundation (NAAC A++ University), India. He is a seasoned expert in computer science engineering with more than 17 years of experience in academia. He has made significant academic contributions, evidenced by numerous patents and a vast array of publications, including conference papers and book chapters. His research interests include machine learning, deep learning, social network analysis, anomaly detection, and network intrusion detection.

MOHAMMAD ISHRAT received the master's and Ph.D. degrees. He is currently an Associate Professor with the Koneru Lakshmaiah Education Foundation (NAAC A++ University), India. He has extensive teaching and research experience at various universities, including King Abdul Aziz University, Jeddah, Saudi Arabia; the University of Technology and Applied Science, Oman; and Integral University, Lucknow, India. He has more than 20 years of international and national academic careers in computer science and engineering. Previously, he has authored many research articles, patents, and book chapters in the fields of machine learning, data mining, deep learning, and future networks.

AHMAD NEYAZ KHAN (Member, IEEE) received the B.Sc. (Hons.) and master's degrees in computer applications from Aligarh Muslim University, India, in 2009 and 2012, respectively, and the Ph.D. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. He is currently an Assistant Professor with Integral University, India. His research interests include information security, machine learning, and reversible data hiding in the encrypted domain.

MOHAMMAD ARIF received the B.Tech. degree in computer science and engineering from CCS University, Meerut, India, in 2001, the M.Tech. degree in computer science and engineering from MNNIT, Allahabad, India, in 2008, and the Ph.D. degree in computer science and engineering from Integral University, Lucknow, India, in 2018. He has nearly 22 years of teaching experience. He is currently with the School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu. He has published nearly 40 peer-reviewed papers in Scopus and other international journals, and IEEE and other international and national conferences. He authored one book and two book chapters and holds two Indian patents. His research interests include ad-hoc networks, vehicular networks, cloud computing, mobile computing, security, machine learning, and deep learning.

ANWAR AHAMED SHAIKH received the B.Tech. and M.Tech. degrees in IT from Uttar Pradesh Technical University, Lucknow, Uttar Pradesh, India, and the Ph.D. degree from Integral University, Lucknow. He is currently an Assistant Professor with the Koneru Lakshmaiah Education Foundation, Vijayawada. He has more than 15 years of teaching experience. His current research interests include cloud computing, fault tolerance in cloud computing, edge computing, and network intrusion detection.

MOUSA MOHAMMED KHUBRANI is currently the Dean of the College of Engineering and Computer Sciences, Jazan University. His research interests include mobile applications, the IoT, machine learning, and blockchain.

SHADAB ALAM (Senior Member, IEEE) received the bachelor's, master's, and Ph.D. degrees in computer science from Aligarh Muslim University, Aligarh. He is currently an Assistant Professor with the Department of Computer Science, Jazan University, Jazan, Saudi Arabia. His research interests include cryptography and information security, the Internet of Things (IoT), blockchain technology, and smart healthcare. He has published more than 70 research papers in reputed journals and international conference proceedings. He holds seven patents in his domain and is further working as an editor in several reputed journals. He is a member of CSI, CRSI, ACM, IAENG, CSTA, IACSIT, and ICSE.

MOHAMMED SHUAIB (Member, IEEE) received the M.Tech. degree in computer engineering from Aligarh Muslim University, India, and the Ph.D. degree in computer engineering from the Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Malaysia. He is currently with the Department of Computer Science and Engineering, Jazan University, Saudi Arabia, and a Visiting Research Fellow with the Faculty of Information Technology, INTI International University, Malaysia. He holds four patents in his field and has contributed to numerous educational workshops and seminars. He has authored 21 scientific articles published in internationally recognized journals, including Elsevier, IEEE, and Springer, all indexed by the Web of Science. His research interests include self-sovereign identity, blockchain, information security, the IoT, identity management, cloud computing, cyber-physical systems, and sensor networks. His research extends to various editorial responsibilities, serving as an Associate Editor for *Sustainable Computing* (Elsevier), an Academic Editor for *PLOS ONE*, and an Editorial Board Member for *Discover Internet of Things* (Springer), and *Frontiers in Blockchain*. He is actively involved with several editorial boards, including the *Journal of Economy and Technology* (Elsevier) and the *International Journal of Advances in Applied Sciences* (IAES).

RAJAN JOHN (Member, IEEE) is currently an Assistant Professor with the Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia. He has more than 24 years of experience in teaching and research. He has presented and published several papers in conferences and journals. His research interests include networking, artificial intelligence, data mining, software agents, and decision support systems. Apart from IEEE, he holds membership in various professional bodies, such as Life Member of IACSIT, ISTE, and IAENG. He was the Former Chair of the IEEE Ghana Section, Ghana, West Africa.