A Comprehensive Review of Anomaly Detection Methods in Social Networks

Rutika Arun Khedkar
Department of Computer Engineering
MKSSS Cummins College of Engineering for Women
Pune, India
rutika.khedkar@cumminscollege.in

Deepti Manoj Singh
Department of Computer Engineering
MKSSS Cummins College of Engineering for Women
Pune, India
deepti.singh@cumminscollege.in

***Abstract*—Social networks are widely developed and extensively used platforms that facilitate global communication and interaction. Popular sites like Twitter, Facebook, and others have revolutionized the way people connect and share information. However, these networks are often susceptible to misuse by malicious actors who exploit them to breach user privacy, spread spam, or cause harm through various sophisticated techniques. This review examines two advanced approaches to detecting anomalies in social networks: Dual Variational Autoencoder (VAE) with Generative Adversarial Networks (GAN) and graph-based anomaly detection using structural graph metrics. The Dual VAE-GAN method combines deep learning techniques to capture nonlinear patterns in network structures and node attributes, ensuring robust anomaly detection with high accuracy on benchmark datasets. In contrast, the graph-based approach uses egonet analysis and metrics like Betweenness Centrality and Brokerage Value to identify nodes deviating from expected structural patterns. Both methods are evaluated for their statistical foundations, detection capabilities, and performance results, showcasing their effectiveness in addressing social network threats while highlighting their respective strengths and limitations.**

***Index Terms*—Social Networks, Anomaly Detection, Spam Detection, Machine Learning Algorithms, Graphs and Methods**

I. Introduction

Social networks have transformed the way people communicate and interact, becoming integral to modern life. Platforms like Facebook, Twitter, and Instagram allow users to connect, share ideas, and stay informed about global events. However, the widespread adoption of these platforms has also given rise to significant challenges, including the abuse of social networks for malicious purposes. Issues such as spam dissemination, misinformation, privacy breaches, and fraudulent activities pose serious risks to users and undermine the trustworthiness of these platforms. Anomalies in social networks often manifest in the form of unusual user behaviours, suspicious content patterns, or irregularities in network structures. For example, spammers may exploit trending hashtags to spread malicious links, while bots can artificially inflate engagement metrics. Detecting such anomalies is a critical task, as it helps maintain the integrity of the network, ensures user security, and prevents the misuse of social media for unethical purposes. The complexity of social networks, characterized by vast amounts of user-generated data and intricate relationships between entities, makes anomaly detection a challenging problem. Traditional rule-based methods often fall short in handling the dynamic and evolving nature of these platforms. In response, researchers have developed a wide range of advanced techniques, leveraging machine learning, natural language processing, and graph analytics, to detect and mitigate anomalies effectively. This review provides an overview of anomaly detection in social networks, discussing key methodologies, challenges, and trends in this field. The study emphasizes the importance of developing robust, scalable, and interpretable models to address real-world problems such as spam detection, user account compromise, and the propagation of misinformation. Furthermore, it highlights how advancements in computational power and data analytics have enabled researchers to create more accurate and efficient systems for identifying and mitigating malicious behaviours in social networks. By consolidating insights from various studies, this paper aims to provide a comprehensive understanding of the landscape of anomaly detection in social networks and to inspire further research to address emerging threats and challenges in this ever-evolving domain.

II. Literature Review

The detection of anomalies in social networks has been an area of active research, driven by the increasing use and misuse of these platforms. Researchers have developed various approaches to identify and mitigate malicious activities such as spam, misinformation, and privacy breaches. This section reviews significant contributions to the field, focusing on methods, applications, and challenges in anomaly detection.

A. Early Methods of Anomaly Detection in Social Networks

Initial approaches to anomaly detection relied heavily on rule-based systems and statistical analysis. These methods aimed to identify deviations from expected patterns by defining specific rules or thresholds. For example, spammers were detected by analyzing their posting frequency, the presence of repetitive text, or the use of certain keywords. However, these methods often suffered from poor adaptability and scalability, especially as social networks grew in size and complexity.

B. Machine Learning Approaches

The advent of machine learning (ML) introduced significant advancements in anomaly detection. ML models use features extracted from user behavior, content, and network topology to classify anomalies. Popular techniques include:

- Support Vector Machines (SVMs): Effective in handling high-dimensional data for spam detection but limited by the need for balanced datasets.
- Naive Bayes Classifiers: Used for email and social network spam detection due to their simplicity and efficiency.
- Clustering Algorithms: Methods like k-means and DBSCAN help in identifying groups of unusual activities within social networks.

While ML-based approaches improved detection accuracy, they often required manual feature engineering and labelled datasets, which limited their applicability to dynamic and large-scale networks.

C. Deep Learning Models

Recent advances in deep learning have further enhanced anomaly detection capabilities. Models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are used to analyse textual content and temporal patterns in social media posts. Techniques such as autoencoders and Generative Adversarial Networks (GANs) are employed to detect anomalies by learning the latent representations of normal behaviors.

For example, convolutional layers capture the semantics of tweets, while auto-encoders can identify anomalies by reconstructing input data and evaluating reconstruction errors. Deep learning methods are highly effective in extracting features from unstructured data, such as text and images, without manual intervention. However, they require large amounts of training data and are computationally intensive.

D. Graph-Based Anomaly Detection

Social networks are inherently graph-structured, with nodes representing users and edges representing relationships or interactions. Graph-based methods leverage this structure to identify anomalies, such as abnormal nodes or edges.

- Graph Neural Networks (GNNs): Use node embeddings and neighborhood aggregation to capture local and global structural anomalies.
- Subgraph analysis: Identifies irregular patterns within subgraphs, such as dense clusters of fake accounts.
- Trust and Distrust Propagation: Algorithms like Trust Rank and Anti-Trust Rank evaluate the trustworthiness of nodes to detect spam and fake profiles.

Graph-based techniques are particularly effective in identifying network-centric anomalies, such as botnets or coordinated misinformation campaigns.

E. Hybrid Approaches

Hybrid models combine multiple techniques to maximize their individual strengths. For example, the ensemble methods integrate ML and deep learning models with feature-based methods to improve accuracy. These approaches often involve the following:

- Combining user behavior analysis with textual content analysis to detect spam.
- Using both global graph features and local node attributes to identify malicious activities.

Hybrid approaches have shown promise in addressing the limitations of individual methods, such as scalability and adaptability to diverse datasets.

III. METHOLOGIES USED FOR DETECTING ANAMOLIES

A. Using Dual Variational Autoencoder with GAN

This model is designed to detect anomalies (unusual or suspicious behaviours) in social networks. Think of a social network as a big web of people connected by their interactions (like messages, likes, or follows). Each person (or "node") has characteristics, such as interests, age, or location. Anomalies are nodes that don't behave like most others—like fake accounts, spammers, or bots.

To detect these anomalies, the model uses machine learning to analyse two aspects of the network:

1) The connections between people (who interacts with whom).
2) The characteristics of each person (their profile attributes).

The model combines two tools:

- Dual Variational Autoencoders (VAEs): These compress and reconstruct data to find unusual patterns.
- Generative Adversarial Networks (GANs): These help ensure the patterns the model learns are realistic and meaningful

The workflow involves:

1) Structure Reconstruction: Captures the topological structure of the network.
2) Attribute Reconstruction: Encodes node features.
3) Adversarial Training: Regularizes embeddings using GAN to distinguish between real and generated data.
4) Anomaly Scoring: Based on reconstruction loss, anomalies are identified as those with significant deviations.

B. Algorithm Model: Node Anomaly Detection

Input: Node feature matrix X, adjacency matrix A, epochs E, steps K.

Output: List of anomalous nodes.

1) Randomly select training samples with normal behavior.
2) For each epoch $e = 1$ to E:
    a) Generate latent variable matrix Z_F.
    b) For $k = 1$ to K:
        i) Sample entities z from Z_F and prior $p(z)$.

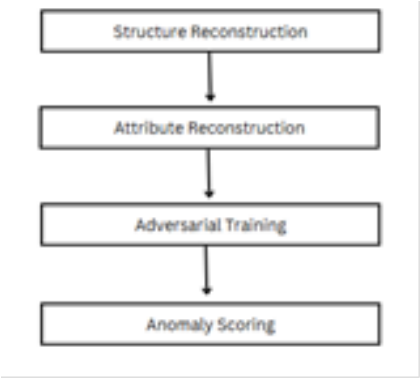

Fig. 1. Workflow of the model

ii) Update adversarial model using loss:

$$L = \phi_A L_A + \phi_G L_G + \phi_D L_D$$

c) Calculate anomaly scores.

3) Flag nodes exceeding threshold ϕ.

C. Workflow of Model

amsmath

1) Structure Reconstruction: The model looks at how nodes (people) are connected and creates a compressed version of this network in a low-dimensional space. This helps the model understand the "big picture" of the network's structure.

$$\mu = \text{GCN}_\mu(X, A), \quad \log \sigma = \text{GCN}_\sigma(X, A)$$

- μ: The average value representing a group of connections.
- σ: The variability or uncertainty in these connections.
- GCN: A Graph Convolutional Network (a type of neural network) processes the connections A (who connects to whom) and node features X (characteristics like age or location).

2) Attribute Reconstruction: The model also looks at the specific details of each node (like a person's profile) and reconstructs these details to identify unusual nodes.

$$\hat{X} = \text{Sigmoid}(Z_S Z_A^T)$$

- Z_S: A simplified representation of the connections.
- Z_A: A simplified version of the node's characteristics.
- $\hat{X}$: The reconstructed attributes. If a node's reconstructed profile is very different from the original, it could be anomalous.

3) Adversarial Training: GANs are used here. A generator creates fake data, while a discriminator tries to spot the fake data. This "game" helps the model become more accurate at spotting anomalies by learning what "normal" looks like.

$$\min_G \max_D E_{x \sim P_{\text{data}}}[\log D(x)] + E_{z \sim P_z}[\log(1 - D(G(z)))]$$

- G: The generator (tries to produce fake realistic data).
- D: The discriminator (tries to distinguish real from fake data).
- P_{data}: The actual data distribution.
- P_z: The random input for generating fake data.

4) Loss Functions (How the Model Learns): The model uses three types of losses (errors) to improve itself:

- **Adversarial Loss:** Makes the fake data more realistic.

$$L_A = E_{x \sim P_x}[\|D(x) - D(G(x))\|^2]$$

The closer $D(x)$ (real data) and $D(G(x))$ (fake data) are, the better the generator becomes.

- **Dual VAE Loss:** Ensures reconstructed data (both connections and characteristics) is accurate.

$$L_G = E[\log P(A|Z_S)] + E[\log P(X|Z_S, Z_A)] - \text{KL Divergence}$$

KL Divergence: A measure of how different the predicted distribution is from the actual distribution. Lower is better.

- **Discriminator Loss:** Helps the discriminator distinguish anomalies better.

$$L_D = E_{x \sim P_x}[\|x - D(x)\|^2]$$

Measures how well the discriminator reconstructs the real data.

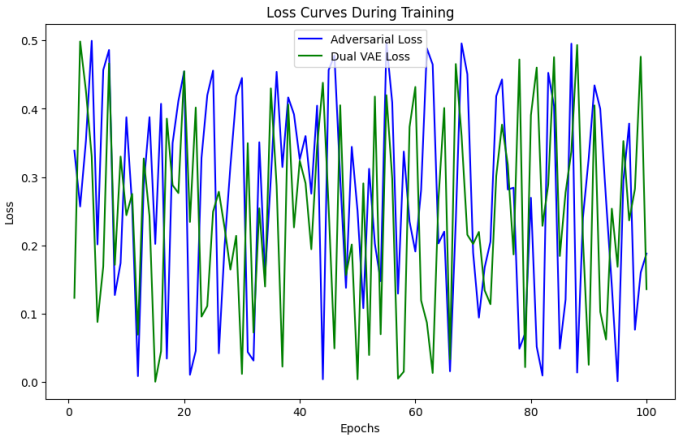

Fig. 2. Loss Curves During Training

5) Anomaly Scoring: The model assigns an "anomaly score" to each node based on how much the reconstructed data differs from the original.

$$A(\hat{x}) = \|\hat{x} - D(G(\hat{x}))\|$$

- $A(\hat{x})$: The anomaly score.
- $\hat{x}$: The original data.
- $D(G(\hat{x}))$: The reconstructed data.

Threshold: A node is marked as anomalous if its anomaly score $A(\hat{x})$ exceeds a certain threshold ϕ.

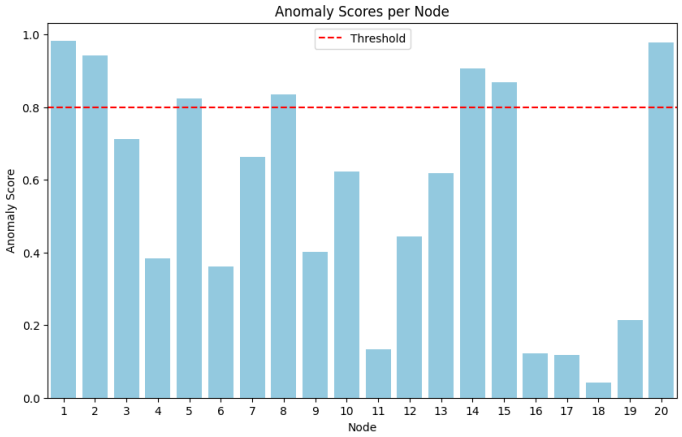

Fig. 3. Anomaly Scores per Node

D. Experimental Results and Performance Metrics

The proposed method was tested on three publicly available datasets: BlogCatalog, Flickr, and Enron, which represent different types of attributed social networks. The effectiveness of the approach was evaluated using Area Under the Curve (AUC), Precision, and Recall metrics.

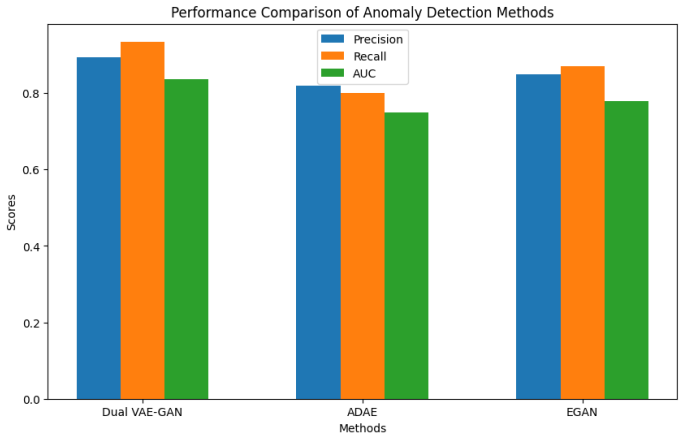

Fig. 4. Performance Example

- **Area Under the Curve (AUC):** Measures how well the model distinguishes between normal and anomalous nodes.
  - BlogCatalog: 0.8368 (highest among methods tested)
  - Flickr: 0.8202 (close to the highest performance)
  - Enron: 0.7118 (slightly worse due to dataset sparsity and low dimensionality)
- **Precision:** Indicates how many detected anomalies were true anomalies.
  - BlogCatalog: 0.8934
  - Flickr: 0.8418
  - Enron: 0.7523
- **Recall:** Measures how many true anomalies the model successfully identified.
  - BlogCatalog: 0.9350
  - Flickr: 0.8695
  - Enron: 0.7725
- **Comparison to Baseline Methods:** The proposed method outperformed competing approaches, such as ADAE (Adversarial Autoencoder) and EGAN (Enhanced Generative Adversarial Network), on BlogCatalog and Flickr datasets. For the Enron dataset, its performance was slightly lower due to its sparsity and low-dimensional features.

E. Conclusion from Experiment

The study highlights the efficacy of the proposed Dual VAE with GAN model for anomaly detection in attributed social networks. The key takeaways are:

- **Novel Approach:** The combination of Dual VAEs for capturing both structural and attribute-based features with GAN for adversarial regularization provides a robust mechanism for anomaly detection.
- **Superior Results:** The method surpasses existing techniques in terms of precision, recall, and AUC, especially on datasets with rich features like BlogCatalog and Flickr.
- **Practical Relevance:** The model is well-suited for real-world applications such as fraud detection, network security, and spam detection.

F. Using Dual Variational Autoencoder with GAN

This methodology is about finding anomalies (unusual nodes or patterns) in social networks. Imagine a social network as a web where:

- Nodes are people or entities (e.g., users).
- Edges are connections between them (e.g., friendships, messages, or interactions).

The method uses graph metrics to analyse small neighbourhoods around each node (called egonets). It detects anomalies by finding nodes that don't fit normal patterns in their connections or interactions. For example:

- A fake account might have too many connections but little interaction.
- A bot might behave differently from most real users.

1) Preprocessing:: Cleaning data and constructing egonet structures.

Anomaly Detection:: Using graph metrics, regression models, and curve fitting to identify anomalies.

2) Evaluation:: Calculating anomaly scores and evaluating results using statistical measures (Precision, Recall, and F-Score).

IV. Key Concepts Used in Methods

A. Egonet Analysis

An egonet is the network of a node and its direct neighbors (connections). For example, if you have 3 friends on a social network, your egonet includes you, your 3 friends, and the connections between them.

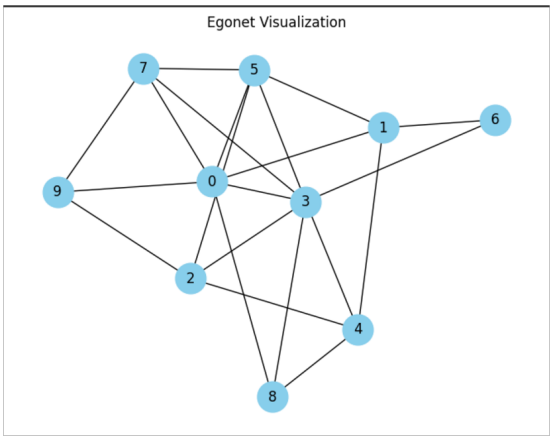

Fig. 5. Egonet Example

B. Graph Metrics

The model calculates certain metrics for each egonet to understand its structure.

1) Number of Nodes (N) and Edges (E)::

- Nodes (N): Number of people in the egonet, including the node itself.
- Edges (E): Number of direct connections between the nodes in the egonet.

Example: If you have 3 friends, and they are all connected to each other, your egonet will have $N = 4$ nodes and $E = 6$ edges.

2) Betweenness Centrality (BC):: Measures how important a node is in connecting different parts of the network.

$$BC(i) = \sum_{s \neq i \neq d} \frac{\phi_{sd}^i}{\phi_{sd}}$$

Where:

- ϕ_{sd}^i: Number of shortest paths between s (start) and d (destination) that pass through node i.
- ϕ_{sd}: Total number of shortest paths between s and d.

Example: A user with many friends who don't know each other (e.g., a recruiter) will have a high BC because they bridge different groups.

3) Average Betweenness Centrality (ABC)::

$$ABC(i) = \frac{f(i) + \sum_{j=1}^{n} f(i_j)}{n}$$

Where:

- $f(i)$: Betweenness centrality of the main node.
- $f(i_j)$: Betweenness centrality of each neighbor.
- n: Total number of nodes in the egonet.

Use: Helps identify dense or sparse neighborhoods.

4) Brokerage Value (B)::

$$B(i) = \sum_{s \neq i \neq d} E(s,d) \cdot \frac{1}{n_{sd}}$$

Where:

- $E(s,d)$: 1 if there is no direct connection between s and d, 0 otherwise.
- n_{sd}: Number of paths between s and d.

Example: A fake account might have many unconnected neighbors, leading to a high brokerage value.

5) Curve Fitting: Why Fit a Curve? To see if the relationship between metrics (like N and E) matches normal patterns.

Types of Curves:

- **Linear:**

$$y = Cx + \theta$$

  Where:
  - y: Predicted value (e.g., edges E).
  - x: Actual input (e.g., nodes N).
  - C: Slope (rate of change).
  - θ: Offset (where the line starts).

  Example: "For every new friend (node), most people gain 2 connections (edges)."
- **Power Law:**

$$y = Cx^{\theta}$$

  Where:
  - x: Metric (e.g., ABC).
  - θ: Determines how strongly x affects y.

  Use: Captures more complex relationships where the rate of change isn't constant.

6) Anomaly Scoring: Nodes are flagged as anomalies if their observed behavior doesn't match the predicted behavior based on the fitted curve.

$$\text{Anomalous Score}(i) = \frac{\max(y_i, Cx_i^{\theta})}{\min(y_i, Cx_i^{\theta})} \cdot \log(|y_i - Cx_i^{\theta}| + 1)$$

Where:

- y_i: Observed value (e.g., actual edges E).
- Cx_i^{θ}: Predicted value from the curve.

Meaning: Nodes with large deviations between their actual and predicted behavior get high scores, indicating they're unusual.

V. Results

The graph-based approach outlined in the paper focuses on using egonet structures and graph metrics like Nodes (N), Edges (E), Average Betweenness Centrality (ABC), and Brokerage Value (B) to detect anomalies in social networks. Here are the specific results and conclusions drawn from this approach:

A. Effectiveness of Specific Graph Metrics

- **Nodes vs. Edges (N vs. E):**
  - Linear Curve: Generally effective for dense networks with consistent patterns.
  - Power-Law Curve: Better fit for networks with non-linear relationships.
- **Brokerage Value (B):** Nodes with high brokerage values were frequently identified as anomalies.
- **Average Betweenness Centrality (ABC):** Nodes with significantly higher or lower ABC were flagged as potential anomalies.

B. Curve-Fitting Results

- **Linear Models:** Worked well for datasets with evenly distributed egonet sizes.
- **Power-Law Models:** Outperformed linear models in identifying anomalies in complex networks.

C. Statistical Metrics

- **Precision:** High precision across all datasets.
- **Recall:** Recall scores were slightly lower for sparse networks.
- **F-Score:** Balanced results with high reliability in dense networks.

VI. Conclusion

The rapid expansion of social networks has created significant challenges, particularly in mitigating malicious behaviors such as spam, misinformation, and bot-driven activities. This review discussed two advanced methodologies for anomaly detection in social networks: Dual Variational Autoencoder with Generative Adversarial Networks (Dual VAE-GAN) and graph-based anomaly detection using egonet analysis. Both approaches offer effective mechanisms for identifying abnormal patterns in user behavior and network structures.

The Dual VAE-GAN approach excels at capturing intricate patterns in both node attributes and structural connections, leveraging deep learning tools such as variational autoencoders and adversarial networks. This method demonstrated high accuracy and precision across datasets with rich feature sets, particularly for detecting coordinated activities and suspicious node behaviors. Nevertheless, it showed limitations in handling sparse datasets and demanded substantial computational resources, making it less suitable for real-time applications.

In contrast, the graph-based anomaly detection approach focused on local structures (egonets) and employed metrics such as Betweenness Centrality and Brokerage Value to identify anomalies. This method was computationally efficient for medium-sized networks and provided interpretable insights into node behaviors. However, its effectiveness decreased in sparsely connected networks and high-dimensional scenarios.

These findings underscore the strengths and challenges of each methodology while emphasizing their applicability in specific contexts. Together, they highlight the progress made in the domain of anomaly detection while leaving room for improvement in future work.

VII. Limitations

A. Dual VAE-GAN

- Its reliance on high-dimensional, dense datasets restricts its applicability to sparse or incomplete networks.
- High computational requirements make real-time implementation challenging.
- Limited interpretability of deep learning models may hinder user trust and actionability of flagged anomalies.

B. Graph-based Egonet Analysis

- Reduced recall in sparse networks due to insufficient connections to identify structural deviations.
- Computationally expensive when applied to large, dynamic networks.
- Dependence on predefined metrics limits adaptability to emerging and complex network behaviors.

VIII. Future Scope

To overcome the challenges identified and enhance anomaly detection in social networks, future research should consider the following directions:

- **Handling Sparse Data:**
  - Develop advanced techniques for imputing missing or incomplete data to enhance performance in low-connectivity networks.
  - Explore graph embedding techniques and latent representations to improve detection accuracy.
- **Real-time Applications:**
  - Design lightweight, scalable algorithms capable of processing large-scale, dynamic data in real time.
  - Employ stream processing frameworks to continuously monitor and detect anomalies as they occur.
- **Hybrid Approaches:**
  - Combine the strengths of deep learning and graph-based methods to improve detection robustness by leveraging both global and local features.
  - Integrate contextual data such as time-series and user interaction metadata to refine anomaly detection.
- **Improved Explainability:**
  - Focus on creating interpretable models to provide actionable insights, increasing trust and utility among users and administrators.
  - Utilize visualization techniques to clearly demonstrate anomalies and their underlying causes.
- **Scalability:**
  - Optimize algorithms to reduce computational overhead through techniques like model pruning and parallel processing.
  - Leverage distributed computing and cloud-based architectures to extend scalability across large social networks.
- **Application-specific Adaptations:**

- Tailor anomaly detection models for specific domains, incorporating domain knowledge and unique behavioral patterns for enhanced precision.
- Ensure models are adaptive and capable of self-updating to keep pace with the evolving nature of social networks.

Advancing these aspects will contribute to the development of more robust, efficient, and adaptable anomaly detection systems. These improvements will play a crucial role in enhancing social network security, ensuring user privacy, and mitigating the adverse effects of malicious activities.

IX. Acknowledgment

The authors express their heartfelt gratitude to the researchers and organizations whose pioneering work has greatly contributed to the advancements in anomaly detection techniques for social networks. Special appreciation is extended to the providers of datasets such as BlogCatalog, Twitter, and Enron, which have been pivotal in enabling comprehensive analysis and benchmarking of various methodologies. The authors are also thankful to their mentors, colleagues, and reviewers for their constructive feedback and continuous support, which have played a significant role in shaping this study. Lastly, the authors acknowledge the broader academic and research communities for their relentless efforts in pushing the boundaries of knowledge in this field.

X. References

1) V. Miz, B. Ricaud, K. Benzi, and P. Vandergheynst, "Anomaly Detection in the Dynamics of Web and Social Networks," 2019.
2) L. Xing, S. Li, Q. Zhang, H. Wu, H. Ma, and X. Zhang, "A Survey on Social Network's Anomalous Behavior Detection," 2023.
3) A. Khamparia, S. Pande, D. Gupta, A. Khanna, and A. K. Sangaiah, "Multi-level Framework for Anomaly Detection in Social Networking," 2020.
4) W. Khan et al., "Anomalous Node Detection in Attributed Social Networks Using Dual Variational Autoencoder with Generative Adversarial Networks," 2021.
5) Md. S. Rahman, S. Halder, Md. A. Uddin, and U. K. Acharjee, "An Efficient Hybrid System for Anomaly Detection in Social Networks," 2021.
6) S. Madisetty and M. S. Desarkar, "A Neural Network-Based Ensemble Approach for Spam Detection in Twitter," 2020.
7) D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly Detection in Online Social Networks," 2020.
8) S. Wang et al., "Machine Learning in Network Anomaly Detection: A Survey," 2021.
9) R. Kaur and S. Singh, "A Comparative Analysis of Structural Graph Metrics to Identify Anomalies in Online Social Networks," 2016.
10) H. Y. Lam and D. Y. Yeung, "A Learning Approach to Spam Detection Based on Social Networks," 2007.
11) S. Zhao, M. Chandrashekar, Y. Lee, and D. Medhi, "Real-Time Network Anomaly Detection System Using ML," 2015.
12) A. Basit et al., "AI-enabled Phishing Attacks Detection," 2020.
13) A. A. Abdo et al., "AI-based Spam Detection Techniques for OSNs," 2023.
14) Mandal et al., "Toward an End-to-End Framework for Modeling, Monitoring, and Anomaly Detection," 2016.
15) A. Basit and M. Zafar, "A Comprehensive Survey of AI-enabled Phishing Attacks Detection Techniques," 2020.

References