International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Survey on Graph Encoder for Node Anomaly Detection

Priyansha Tiwari[1], Shweta Dubey[2]

[1]Shri Shankaracharya Professional university, Bhilai India
Email: *priyansha.tiwari191[at]gmail.com*

[2]Shri Shankaracharya Professional university, Bhilai India
Email: *dubeyshweta18111984[at]gmail.com*

Abstract: ***Money laundering is the practice through which criminals disguise the origins of illegally obtained funds, enabling them to integrate these funds into the legitimate financial system. The United Nations estimates that annually, between 2% and 5% of global GDP—approximately $0.8 to $2 trillion is laundered worldwide. This staggering figure underscores the critical need for effective identification and enforcement of anti-money laundering (AML) measures. A variety of techniques have been proposed to detect money laundering, primarily through the analysis of transaction graphs that illustrate money transfers between bank accounts. These methods often focus on the structural and behavioral dynamics of dense subgraphs associated with these transactions. However, a significant limitation of many of these techniques is their failure to account for the common practice of high-volume fund flows that traverse multiple interconnected bank accounts in a chain-like manner. Moreover, many current AML approaches tend to either achieve lower detection accuracy or incur high computational costs, rendering them less effective and practical for real-world financial systems. Consequently, this results in only a small fraction of money laundering activities being successfully detected and prevented, highlighting a critical gap in the ability to combat this global issue effectively. Addressing these challenges is essential to enhance the reliability and efficiency of money laundering detection efforts.***

Keywords: Graph Anomaly Detection, Attribute Encoder, Graph Connection

1. Introduction

Money laundering represents a substantial challenge to the global financial system, allowing criminals to obscure the origins of illicit funds and facilitating further illegal activities. According to estimates from the United Nations, a staggering $0.8 to $2 trillion—equivalent to 2% to 5% of global GDP—is laundered each year. This alarming figure highlights the pressing need for effective strategies to detect and prevent money laundering. In response, researchers and practitioners have explored various techniques to identify money laundering activities, often focusing on the analysis of transaction graphs that illustrate the movement of money between bank accounts. These methods typically investigate the structural and behavioral characteristics of these financial networks. However, many existing approaches struggle to accurately detect money laundering due to the complexity of high-volume transactions and the intricate relationships among multiple accounts. Additionally, challenges related to computational efficiency and the occurrence of false positives can limit the practical applicability of these techniques in real-world settings. As a result, a considerable portion of money laundering goes undetected, posing significant risks to economic stability and security. This survey paper seeks to provide a detailed examination of the current landscape of money laundering detection research. We will analyze various methodologies, assessing their strengths and weaknesses, and explore emerging trends and technologies in the field. By synthesizing these findings, we aim to highlight gaps in the existing literature and contribute to the development of more effective anti-money laundering frameworks capable of addressing the complexities of modern financial systems.

1.I. Graph anomaly Detection Method

Graph anomaly detection methods aim to identify unusual patterns or behaviors within graph-structured data. One common approach is statistical methods, which analyze metrics like node degree, clustering coefficients, or community structures to flag deviations from expected behavior. Another popular technique is machine learning, where algorithms such as autoencoders, one-class SVMs, or graph neural networks are trained on normal data to recognize anomalies based on learned representations. Additionally, spectral methods leverage the graph's eigenvalues and eigenvectors to detect structural anomalies. Hybrid methods that combine these techniques are also emerging, offering improved accuracy and robustness. Ultimately, the choice of method depends on the specific characteristics of the graph and the nature of the anomalies being investigated. Graph anomaly detection methods encompass a variety of techniques tailored to different types of graphs and anomaly characteristics.

Statistical Approaches: These methods rely on the distribution of graph metrics, such as the degree distribution or clustering coefficients. By establishing baseline statistical models, they can identify nodes or edges that significantly deviate from expected patterns, suggesting potential anomalies.

Machine Learning Techniques: Supervised and unsupervised learning algorithms are increasingly popular for anomaly detection in graphs. Supervised methods require labeled training data and can include decision trees or support vector machines, while unsupervised approaches, like clustering algorithms (e.g., k-means), help discover anomalies without prior labels. Autoencoders and graph

neural networks (GNNs) are particularly effective, as they can learn complex patterns and relationships in high-dimensional graph data.

Spectral Methods: These techniques analyze the graph's Laplacian matrix and its eigenvalues to identify anomalies based on structural properties. By examining how the graph's topology changes, these methods can highlight unusual substructures or node behavior.

Random Walks and Graph Embeddings: Random walk-based methods assess node behavior by simulating walks across the graph, allowing for the identification of nodes that behave differently than expected. Graph embeddings transform nodes into continuous vector spaces, facilitating the application of traditional anomaly detection algorithms.

Hybrid Approaches: Combining multiple methods can enhance detection capabilities. For instance, integrating statistical techniques with machine learning can improve robustness and accuracy, as different methods may capture distinct aspects of anomalies.

Temporal Graph Anomaly Detection: For graphs that evolve over time, temporal anomaly detection techniques focus on identifying anomalies in dynamic graph data. This can involve analyzing changes in node connectivity, edge weights, or attributes over time, allowing for the detection of unusual spikes in activity, like sudden surges in communication between specific nodes.

Graph-Based Outlier Detection: This method explicitly defines outliers based on their relationships within the graph. Techniques like Local Outlier Factor (LOF) can be adapted to graphs, measuring how isolated a node is compared to its neighbors. Nodes that have significantly fewer connections or those that belong to sparse regions of the graph are flagged as potential anomalies.

Subgraph Detection: Sometimes, anomalies can be characterized by the presence of unusual subgraphs or motifs. Algorithms can search for specific patterns or structures that deviate from the expected norm, enabling the detection of complex anomalies, such as fraudulent groups in social networks or unexpected interactions in biological networks.

2. Various Methods

Graph encoders have gained significant attention in the field of graph representation learning. They transform graph-structured data into a format suitable for machine learning tasks, enabling effective anomaly detection, node classification, and link prediction. Here's a summary of key approaches and developments in this area:

1) Graph Neural Networks (GNNs):

GNNs are a prominent class of graph encoders that leverage node features and structural information. They update node representations by aggregating features from neighboring nodes, allowing for context-aware embeddings. Notable variants include:

- Graph Convolutional Networks (GCNs): Introduced by Kipf and Welling (2017), GCNs use convolutional layers adapted to graph structures, allowing for efficient learning of node embeddings based on local neighborhoods.
- Graph Attention Networks (GATs): GATs, proposed by Veličković et al. (2018), introduce attention mechanisms to weigh the importance of neighboring nodes, improving the focus on relevant connections.

2) Graph Autoencoders (GAEs):

Graph autoencoders are unsupervised methods designed to learn node embeddings by reconstructing the graph structure. Variants include:

- Variational Graph Autoencoders (VGAEs): Building on GAEs, VGAEs incorporate a probabilistic framework to capture uncertainty in node representations, enhancing the model's ability to generalize across unseen data.

3) DeepWalk and Node2Vec:

These pioneering algorithms use random walks to generate node sequences, which can be treated like sentences in natural language.

- DeepWalk: Introduced by Perozzi et al. (2014), DeepWalk applies skip-gram models to learn embeddings from these sequences.
- Node2Vec: This extension, by Grover and Leskovec (2016), introduces a biased random walk approach that balances between breadth-first and depth-first exploration, allowing for more flexible embedding capture.

4) GraphSAGE:

Proposed by Hamilton et al. (2017), GraphSAGE (Graph Sample and Aggregation) learns to generate embeddings by sampling and aggregating features from a node's neighbors. This approach is particularly beneficial for large graphs where full neighborhood aggregation is computationally expensive.

5) Graph Isomorphism Networks (GINs):

GINs, introduced by Xu et al. (2019), focus on distinguishing graph structures by using a powerful aggregation function that enables the model to capture the expressiveness of graph isomorphism, improving performance on various graph-based tasks.

6) Spatial and Temporal Graph Encoders:

Recent works have started to explore spatial-temporal graphs, which capture both structural and temporal dynamics. Models like Spatio-Temporal Graph Convolutional Networks (ST-GCNs) integrate temporal information to enhance the representation of evolving graphs, useful in applications like traffic prediction and event detection.

7) Applications and Benchmarks

Graph encoders have been applied across diverse domains, including social network analysis, recommendation systems, and bioinformatics. Several benchmark datasets, such as Cora, Citeseer, and PubMed, have been established to evaluate the performance of different graph encoding techniques, facilitating comparative studies.

3. Related Work on Graph Encoder

These papers collectively highlight the advancements in graph encoding techniques, each contributing unique methodologies and insights that have shaped the field. The evolution from basic random walk-based methods to sophisticated neural architectures underscores the growing importance of graph representation learning across various applications, including anomaly detection, social network analysis, and recommendation systems.

1) Title: Realistic Synthetic Financial Transactions for Anti-Money Laundering Models
Author: Erik Altman, Jovan Blanuša, Luc von Niederhäusern
Publication Year: NeurIPS 2023

Method:
The study utilizes generative modeling techniques to create realistic synthetic financial transaction data. Methods such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs) may be employed to mimic the distribution of real financial transactions, capturing complex relationships between transaction features.

Key features that are indicative of normal and suspicious behavior are carefully crafted. These may include transaction amounts, frequencies, types, and sender/receiver information. The model aims to ensure that the synthetic data reflects the nuanced patterns found in actual financial datasets.

Finding: The research demonstrates that synthetic data can effectively replicate the statistical characteristics of real-world transaction data, making it suitable for training AML models without compromising sensitive information. Models trained on synthetic data show comparable performance to those trained on real data, particularly in detecting fraudulent transactions. This suggests that synthetic datasets can be a viable alternative, especially in scenarios where access to real data is limited due to privacy concerns.

2)Title: PREM: A Simple Yet Effective Approach for Node-Level Graph Anomaly Detection
Author: Junjun Pan, Yixin Liu
Publication Year: 2023

Method: PREM employs a simple yet powerful approach to learn node representations based on the local graph structure. It utilizes techniques such as node embeddings to capture the relationships and characteristics of nodes within their neighbourhood. The model extracts features from nodes, considering attributes and structural properties (e.g., degree, clustering coefficient) that may indicate anomalous behavior. This feature set is designed to highlight deviations from typical patterns.

Findings: PREM demonstrates strong performance in identifying node-level anomalies across various graph datasets, showing competitive results compared to more complex models. The simplicity of the approach contributes to its effectiveness. The method is noted for its computational efficiency, making it suitable for large-scale graphs. This efficiency arises from its straightforward feature extraction and scoring mechanism, allowing for quick detection processes.

3)Title: ANEMONE: Graph Anomaly Detection with Multi-Scale Contrastive Learning
Author: Ming Jin.
Publication Year: 2021

Method: ANEMONE employs a multi-scale approach to contrastive learning, where it generates different scales of graph representations. This allows the model to capture both local and global features effectively. By contrasting these representations, the model learns to distinguish between normal and anomalous patterns.

he method uses graph neural networks (GNNs) to encode the graph structure into embeddings. These embeddings incorporate both node features and the relational structure of the graph, enhancing the model's ability to identify anomalies based on contextual information.

Findings: ANEMONE outperforms existing graph anomaly detection methods across multiple benchmark datasets, demonstrating its effectiveness in identifying anomalies. The multi-scale contrastive learning approach significantly enhances detection accuracy. The model exhibits strong robustness against noisy data and perturbations, maintaining high performance even in challenging scenarios where anomalies might be subtle or obscured by noise. By leveraging contrastive learning, the model provides interpretable results, allowing users to understand why certain instances are classified as anomalous based on their learned representations.

4) Title: DenseFlow: Spotting Cryptocurrency Money Laundering in Ethereum Transaction Graphs
Author: Dan Lin
Publication Year: 2024

Method: DenseFlow begins by constructing a transaction graph from Ethereum blockchain data. Each node represents an address (account), and each edge represents a transaction between addresses. This graph-based representation allows for a comprehensive analysis of transactional relationships. The method employs flow aggregation techniques to capture the flow of funds across the graph. It aggregates transactions over time, allowing for the identification of dense subgraphs where funds may be laundered. This approach highlights patterns of behavior that indicate potential money laundering activities.

Findings: DenseFlow demonstrates a high level of effectiveness in detecting money laundering activities within Ethereum transaction graphs. The combination of flow aggregation and anomaly detection algorithms results in improved identification of suspicious transactions. The approach successfully identifies dense subgraphs that may indicate laundering activities, showcasing how funds are concentrated among certain addresses over time. This is crucial for understanding the networks of potentially illicit activities. The temporal analysis reveals patterns of behavior that correlate with known money laundering

schemes. This insight helps law enforcement and regulatory bodies understand the dynamics of cryptocurrency transactions more thoroughly.

5) Title: DGraph: A Large-Scale Financial Dataset for Graph Anomaly Detection
Author: Xuanwen Huang
Publishing Year: 2023

Method: DGraph is designed as a large-scale dataset specifically for graph anomaly detection in financial contexts. The dataset incorporates real-world financial transactions and relationships, structured as a graph where nodes represent entities (such as accounts or transactions) and edges represent relationships or transactions between them. To facilitate supervised learning, the dataset includes labeled anomalies based on expert-defined criteria and real-world fraud patterns. This allows for clear differentiation between normal and anomalous instances, essential for training and evaluating detection algorithms.

Finding: DGraph serves as a comprehensive resource for researchers and practitioners in the field of graph anomaly detection. Its large scale and detailed labeling make it a valuable benchmark for developing and evaluating detection methods. Experiments conducted using DGraph reveal that various state-of-the-art anomaly detection algorithms can effectively identify fraudulent behavior in financial networks. The dataset provides insights into the strengths and weaknesses of different approaches. The findings illustrate the complexity of detecting anomalies in financial graphs, highlighting that certain fraudulent behaviors may manifest in subtle ways that require sophisticated detection mechanisms to uncover.

6) Title: GADBench: Revisiting and Benchmarking Supervised Graph Anomaly Detection
Author: Jianheng Tang
Publishing Year: 2023

Method: GADBench establishes a standardized benchmarking framework for supervised graph anomaly detection. It revisits existing datasets, ensuring that they are suitable for contemporary machine learning techniques and addressing previous limitations in the datasets used. The study includes a curated selection of benchmark datasets specifically designed for graph anomaly detection tasks. Each dataset is chosen based on its relevance, complexity, and diversity of anomaly types, including real-world scenarios from domains like social networks and finance. A comprehensive evaluation protocol is introduced, incorporating various metrics such as precision, recall, F1-score, AUC, and computational efficiency. This multi-faceted approach allows for a thorough assessment of different algorithms under consistent conditions.

Finding: The results reveal significant variations in performance among different algorithms, highlighting the strengths and weaknesses of each approach in various scenarios. Some algorithms excel in specific types of anomalies while struggling with others. GADBench underscores the necessity for robust and well-annotated datasets in the field of graph anomaly detection. The study points out that many existing datasets have limitations that can impact the generalizability of detection algorithms. The findings indicate that the choice of evaluation metrics can greatly influence perceived performance. Certain algorithms may perform well on specific metrics while underperforming on others, suggesting that a comprehensive evaluation is critical.

7) Title: Towards Self-Interpretable Graph-Level Anomaly Detection
Author: Yixin Liu
Publishing: 2023

Method: The approach combines graph neural networks (GNNs) with self-explanatory mechanisms to enhance interpretability. The model is designed to detect anomalies at the graph level while providing insights into the reasoning behind its predictions. The model learns rich embeddings for graphs by utilizing both structural and feature-based information. It captures the relationships and properties of nodes and edges, allowing for effective anomaly detection in complex graph structures. The self-explanatory capabilities enhance user interaction with the model. By providing clear explanations of detected anomalies, the framework supports decision-making processes, particularly in critical domains such as finance and security.

Finding: The model successfully demonstrates a significant improvement in interpretability compared to traditional anomaly detection methods. Users can understand the rationale behind detected anomalies, making it easier to trust and validate the model's outputs. The approach shows strong performance in detecting anomalies across various benchmark datasets. It effectively identifies graph-level anomalies while maintaining a balance between detection accuracy and interpretability. The model proves to be robust against variations in graph structure and noise, maintaining high detection performance in challenging scenarios. This robustness is crucial for practical applications in dynamic environments.

8) TitleComGA: Community-Aware Attributed Graph Anomaly Detection
Author: Xuexiong Luo
Publishing 2022

Method: ComGA begins by identifying communities within the attributed graph using community detection algorithms. This step helps in understanding the inherent structure of the graph and establishes a foundation for localized anomaly detection. The method constructs representations that capture both the structural and attribute information of nodes. Each node's features, combined with its community context, create a rich embedding that reflects its role within the graph. ComGA employs a hybrid framework that combines unsupervised learning with community-aware techniques. It analyzes the node embeddings and their relationships within the community to identify deviations from normal behavior.

Finding: ComGA demonstrates improved detection accuracy compared to baseline methods, particularly in scenarios where community structures significantly influence node behavior. The community-aware approach allows for

better identification of anomalies that may be overlooked in global analyses. The framework is robust to noise and variations in the data, effectively maintaining performance even when faced with incomplete or noisy attributes. This robustness is crucial for practical applications in real-world settings. The method provides insights into detected anomalies by highlighting the community context of each anomalous node. This contextual information helps users understand why certain nodes are classified as anomalies, enhancing trust in the model.

9) Title: Rethinking Graph Neural Networks for Anomaly Detection
Author: Jianheng Tang
Publishing 2022

Method: The study begins by critically analyzing existing graph neural network (GNN) architectures used for anomaly detection. It identifies limitations, such as insufficient focus on local neighborhood information and suboptimal aggregation techniques. The proposed method introduces an enhanced message-passing mechanism that improves how information is exchanged between nodes. This mechanism emphasizes both local and global context, allowing for a more nuanced understanding of node relationships. A new objective function is formulated to explicitly optimize for anomaly detection. This involves refining the loss function to focus on distinguishing between normal and anomalous nodes more effectively.

Finding: The proposed method shows significant improvements in anomaly detection performance across various benchmark datasets. The enhanced message-passing and optimized objective function contribute to better identification of anomalous nodes. The approach offers insights into detected anomalies, providing explanations for why certain nodes are classified as anomalous. This interpretability is valuable for practical applications where understanding the rationale behind decisions is critical. The method is designed to be scalable, capable of handling large graphs without compromising performance. This scalability is essential for real-world applications in areas such as fraud detection and network security.

10) Title: Principal Neighbourhood Aggregation for Graph Nets.
Author: Gabriele Corso
Publishing 2020

Method: The study introduces a novel framework for aggregating information from a node's neighbors, focusing on principal component analysis (PCA) techniques. This method allows for efficient summarization of the local neighborhood, capturing essential structural and feature information while reducing dimensionality. The approach involves extracting principal components from the feature representations of neighboring nodes. This helps in identifying the most significant directions of variation in the data, enabling the model to focus on relevant features during the aggregation process. The proposed method is integrated into a graph neural network (GNN) architecture, where the principal neighborhood aggregation replaces traditional aggregation methods. This modification aims to enhance the model's ability to learn robust representations that are sensitive to the underlying graph structure.

Finding: C The method demonstrates significant improvements in performance across multiple graph-related tasks. By utilizing principal neighborhood aggregation, the model achieves better accuracy in capturing the relationships between nodes. The approach shows robustness to noise in the data, as the dimensionality reduction helps filter out irrelevant features while preserving crucial information. This robustness is beneficial for practical applications in noisy environments. The principal component extraction process is computationally efficient, allowing for scalable implementations in large graphs. This efficiency is critical for applications involving extensive datasets.

4. Graph Applications

Graph applications are widespread and span various domains due to the ability of graph structures to represent complex relationships and interactions. Here are some notable areas where graph applications are prevalent:

Social Network Analysis
Community Detection: Identifying groups of closely connected users.
Influencer Identification: Recognizing key individuals who can sway opinions or behaviors within the network.
Recommendation Systems: Suggesting friends, content, or connections based on user interactions and relationships.

Fraud Detection
Financial Transactions: Analyzing transaction graphs to identify patterns indicative of fraud, such as money laundering or credit card fraud.
Insurance Claims: Detecting suspicious claims by examining relationships between claimants and service providers.

Biological and Healthcare Networks
Protein-Protein Interaction: Mapping interactions between proteins to understand biological processes.
Disease Propagation Models: Studying how diseases spread through networks, which can inform public health strategies.

Transportation and Logistics
Route Optimization: Utilizing graphs to find the most efficient routes for delivery and transportation.
Traffic Flow Analysis: Analyzing road networks to improve traffic management and reduce congestion.

Knowledge Graphs
Information Retrieval: Structuring knowledge in a graph format to enhance search capabilities and data retrieval.
Semantic Search: Enabling more intuitive searches by understanding the relationships between concepts.

Recommendation Systems
Collaborative Filtering: Using user-item interaction graphs to recommend products, movies, or services based on user preferences and behavior.

Telecommunications
Network Optimization: Analyzing network graphs to optimize data flow and improve service quality.
Fault Detection: Identifying failures or performance issues in network infrastructure.

Financial Market Analysis
Portfolio Management: Analyzing relationships between assets to inform investment strategies.
Market Trend Analysis: Studying connections among stocks, commodities, or other financial instruments.

Cybersecurity
Intrusion Detection: Monitoring network traffic graphs to identify anomalous behavior indicative of security breaches.
Vulnerability Analysis: Assessing connections in software systems to identify potential vulnerabilities.

Recommendation and Personalization
User Behavior Analysis: Leveraging graph structures to understand user preferences and tailor recommendations in e-commerce and media platforms.

5. Conclusion

In conclusion, the versatility of graph applications across diverse domains underscores their significance in understanding and navigating complex relationships and interactions. From social network analysis to fraud detection, and from healthcare to cybersecurity, graph-based approaches provide powerful tools for modeling, analyzing, and interpreting data. As advancements in graph theory and machine learning continue to unfold, the potential for innovative applications grows, paving the way for more efficient solutions and deeper insights. Embracing these technologies will enable industries to harness the full potential of their data, driving informed decision-making and fostering a more connected understanding of intricate systems.

The integration of advanced graph algorithms and machine learning techniques amplifies their effectiveness, enabling more precise predictions and insights. As the field continues to evolve, the potential for innovative applications will expand, fostering improvements in efficiency and decision-making across industries. By leveraging the power of graphs, organizations can unlock new opportunities for growth, enhance their analytical capabilities, and navigate the complexities of modern data landscapes, ultimately leading to more informed strategies and better outcomes.

References

[1] Altman, Erik, et al. "Realistic synthetic financial transactions for anti-money laundering models." *Advances in Neural Information Processing Systems* 36 (2024).
[2] Pan, Junjun, et al. "PREM: A Simple Yet Effective Approach for Node-Level Graph Anomaly Detection." *2023 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2023.
[3] Jin, Ming, et al. "Anemone: Graph anomaly detection with multi-scale contrastive learning." *Proceedings of the 30th ACM international conference on information & knowledge management*. 2021.
[4] Lin, Dan, et al. "DenseFlow: Spotting Cryptocurrency Money Laundering in Ethereum Transaction Graphs." *Proceedings of the ACM on Web Conference 2024*. 2024.
[5] Xuanwen, Huang, et al. "DGraph: A Large-Scale Financial Dataset for Graph Anomaly Detection." *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*. 2022.
[6] Tang, Jianheng, et al. "Gadbench: Revisiting and benchmarking supervised graph anomaly detection." *Advances in Neural Information Processing Systems* 36 (2023): 29628-29653.
[7] Liu, Yixin, et al. "Towards self-interpretable graph-level anomaly detection." *Advances in Neural Information Processing Systems* 36 (2024).
[8] Luo, Xuexiong, et al. "Comga: Community-aware attributed graph anomaly detection." *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022.
[9] Tang, Jianheng, et al. "Rethinking graph neural networks for anomaly detection." *International Conference on Machine Learning*. PMLR, 2022.
[10] Corso, Gabriele, et al. "Principal neighbourhood aggregation for graph nets." *Advances in Neural Information Processing Systems* 33 (2020): 13260-13271.