International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 1, Jan-Feb 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Enhancing Fraud Detection using Scalable Graph Neural Networks GNN

Dasari Veera Venkata Nooka Surya Saran*, Korivi Dhilly Srikanth Reddy*, Ridhi*
Pabolu Syama Sahtya Sri Phani Pradeep*, Bhukya Rohith*
*Department of Computer Science and Engineering
Apex Institute of Technology, Chandigarh University
Mohali, Punjab, India
Email: dasarisaran2005@gmail.com, korividhillysrikanthreddy@gmail.com, ridhiaggarwal603@gmail.com,
pradeeppabolu6816@gmail.com, rohithbhukya758@gmail.com

***Abstract*—The growing amount and sophistication of financial. deals have rendered the process of detecting fraud as a challenge to. modern financial systems. Conventional machine learning techniques. do not capture transactions often and analyze them individually. relational patterns that can seek to denote coordinated fraudulent. activities. GNNs are effective. solution modeled as an interrelation of transactions, enabling learning feature-level as well as structural in- formation. Nevertheless, the use of GNNs on massive transactions. graphs is highly scalable because of the high challenges. computational and memory requirements. This paper proposes a graph-based framework of fraud detection that is scalable. SAGE architecture that makes use of neighbor sampling and mini- batch training to facilitate effective learning on huge transactions. networks. The approach presented is a representation of transactions as. graph nodes shares with captures relational dependencies. transaction attributes. IEEE- Experimental evaluation. The scalable GNN has been shown to have succeeded in detecting CIS fraud in dataset. Framework attains a ROC-AUC score of 0.723 and gets better. fraud recall versus full-graph GNN training and significantly lowering the computational needs. The results high- light the performance of scalable graph-based fraud. identify and prove that it can be successfully implemented. in massive financial economies.**

***Index Terms*—Fraud Detection, Graph Neural Networks (GNN), GraphSAGE, Transaction Graph, Financial Transactions, Scalable Deep Learning, Neighborhood Sampling, Mini-batch Training, Imbalanced Classification, Machine Learning**

I. INTRODUCTION

The blistering development of online and digital financial services payment platforms has highly grown the volume and partition of transactional information created globally. These developments have made life easy and more convenient accessibility, they have too created novel opportunities of financial losses were made due to fraud of this nature financial institutions and financial customers [1], [2]. Fraudulent transactions are usually complex strategies that are predatory. clandestine connections between various objects including users, de vices, merchants, and accounts, which makes them more and more difficult to detect demanding the conventional methods.

The classical fraud detection tools are based mostly on rule-based detection tools traditional machine learning models and based techniques such as logistic regression, decision trees and ensemble methods [3], [4]. Albeit these methods may work well identify familiar fraud types, they normally study transactions in isolation and they do not grasp the complex relational financial network dependencies. As fraudsters they keep varying their strategies, these are feature-based and stagnant methods usually have difficulties in generalizing to emerging fraud scenarios, which results in more false negatives and less detection reliability [2], [5].

In a bid to overcome such shortcomings, the latest studies have examined. Financial trans-modeling techniques in the form of graph-based learning actions as networks of each other. In this representation, Transactions, accounts and merchants are the entities represented as nodes, their relationships being represented as edges. Graph Neural Networks (GNNs) have become one of them effective learning framework based on such relational structures This is because aggregating information of neighboring nodes is done by them both structural and attribute-level pattern capturing [6], [7]. This feature allows GNNs to detect low-level fraud that can be invisible to the conventional independent feature analysis.

Nevertheless, the use of GNNs in fraud, regardless of their effectiveness detection poses serious computational problems. There are usually millions of nodes in world transaction graphs edges, and thus full-graph training is computationally costly and memory-intensive [8]. This is a scaling constraint that is limiting the introduction of the GNN-based fraud detection systems into practical conditions, where massive data can be processed on time is essential. Recent advances such as neighborhood sampling and mini-batch training have shown promise in addressing these scalability challenges by enabling efficient training on large graphs without requiring full graph processing [8], [9].

Motivated by these challenges, this work focuses on developing a scalable Graph Neural Network–based fraud detection

framework capable of efficiently processing large-scale transaction networks. Specifically, we model transaction data as a graph and employ a sampling-based GraphSAGE architecture to enable scalable learning while preserving detection performance. The proposed approach aims to improve the practicality of graph-based fraud detection by reducing computational overhead and enabling efficient training on large transaction datasets.

Contributions: The main contributions of this paper are summarized as follows:

- We construct a transaction graph representation that captures relational dependencies among financial transactions.
- We implement and evaluate a scalable Graph Neural Network model using neighbor sampling and mini-batch training.
- We compare the proposed scalable approach with traditional machine learning and full-graph GNN models.
- We demonstrate that scalable GNN training improves fraud detection recall while enabling efficient processing of large transaction graphs.

Organization of the paper: The remainder of this paper is organized as follows Section II provides a literature review of related work. graph-based learning and fraud detection. Section III describes the experiment design and measures of evaluation. Section IV introduces the offered scalable GNN methodology. V discusses the analysis of the results and scalability. Finally, VI sums up the paper and identifies the future research direction.

II. Related Work

A variety of techniques has been widely applied in the detection of fraud of machine and statistical learning. Early fraud detection systems used the rule-based methods mainly defined based on professional skills and preset targets These systems worked well in the detection of known fraud patterns, they were not flexible and did not identify patterns of fraud never witnessed before [1], [3]. As financial Machine learning techniques were made more complicated, and processes became more complex added to enhance accuracy of detection and automation of pattern identification of transactional data.

Conventional machine learning models include logistic regression, decision trees, support vector machines and ensemble. Various techniques have been broadly used in fraud detection exercises. [2], [4]. These models make use of handcrafted features extracted between relation attributes to distinguish between fraudulent and honest transactions. Random Forest is one of the ensemble techniques. Boosting, especially Gradient Boosting has shown good performance because of their capability to predict non linear relationships and reduce overfitting [5]. Nonetheless, these methods usually treat.as an independent observation and they are not taken connections among objects, restricting their performance in identifying fraud synergies. The shortcomings of independent feature-based to fill in the deficiencies of independent feature-based network-based fraud models Network-based fraud has been studied by researchers representation of transactional data as graphs in their tectonic approaches.

In graph-based methods, users, accounts, and nodes are expressed as transactions, and their interactions are represented as edges. This representation mod-allows the mod representation relational dependencies and interaction patterns which are usually signs of fraudulent conduct [10]. Graph-based there have been improvements in the ability of an anomaly detector in the determination of suspicious patterns that contain more than two inter-connected entities.

Graph Neural Networks (GNNs) have become more recently was discovered as a powerful model of learning through graph-structured data. GNNs are extensions of neural networks through integration of neighborhood aggregation mechanisms that allow nodes to acquire representations on the basis of their local and global graph structure [6], [7]. This capability makes GNNs are especially applicable in detecting fraud, particularly where the linkages among entities can be of value on contextual terms information. The effectiveness of several studies has been proved social ness of GNN-based models in detecting financial fraud, social network analysis, and anomaly detection tasks [11], [12].

GraphSAGE introduced is one of the GNN architectures a scaling represented inductive learning method. sample learning on large graphs at random by sampling a set number Neighbouring node during training [8]. This sampling-based strategy enables GNN models to work well on large datasets without necessarily being required to do full graph traversal. Similarly, Cluster-GCN suggested methods of partitioning graphs demonstrate computational efficiency and training efficiency [9], With these developments, it is now possible to use GNN dynamics to large-scale real world datasets. In spite of such developments, scalability will continue to be one of the critical challenges human ¿leading in the implementation of GNN-fraud cases.

The graph training methods need a lot of memory and computational resources, and thereby impractical on large transaction networks. Recent literature has discussed sampling- mini-batch processing methods and based training methods to enhance scalability and preserve model performance [8], [9].Scalable GNN has not been empirically evaluated remains the case of models in the context of financial fraud detection limited.

In this we extend these advances and apply them in this effort a scalable Graph Neural Network-based using neighbor mini-batch and sampling based fraud detection. Unlike the proposed approach lays emphasis on traditional full-graph approaches on how to enhance the efficiency of training without losing the capacity to identify graph patterns of big transaction graphs. This methodology makes NN-based fraud practicable. Large scale financial detection systems.

III. Experimental Setup

The IEEE-CIS Fraud was used as the experiments, which is publicly available through the Kaggle platform [13] detection dataset financial transactions records of the type of fraud or the type of lemishness. To remain computationally feasible and yet maintain representative of 200,000 realistic characteristics transactions has been chosen to be constructed as a graph. In the resulting transaction graph G (V, E), each node is a transactions and directed edges between transactions with common ones card identifier, address identifier, or product are some of the attributes. This model allows the re- to be captured correlations between transactions that can be coordinated or suspicious activity. There is a feature attached to every node, vector derived as a result of transaction attributes, where numerical Standardization and encoding of features. Standardization features are categorical and are encoded into numerical form. The data was categorized into training and testing sets with a stratified split of 80:20 in order to retain the original distribution of classes.

The data was categorized into training and testing sets with a stratified split of 80:20 in order to retain the original distribution of classes.

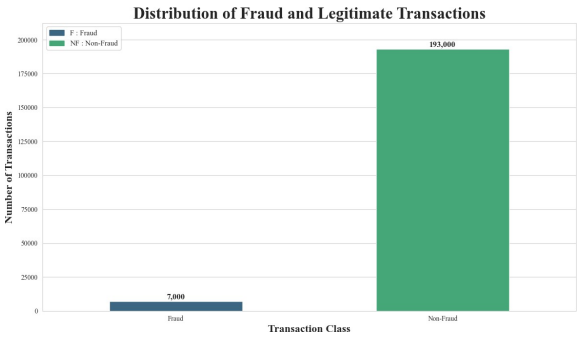

Fig. 1: Distribution of fraud and legitimate transactions in the IEEE-CIS dataset subset used for graph construction. The dataset shows significant class imbalance, with fraudulent transactions representing a small proportion of the total transactions.

As shown in Fig. 1, the dataset is highly imbalanced, with fraudulent transactions being significantly fewer than legitimate transactions. This imbalance reflects real-world financial transaction data and justifies the use of weighted loss and recall-focused evaluation metrics.

The suggested fraud detection system is put in place with two graph convolution using the GraphSAGE architecture layers, one of which is a hidden layer of 64 units and ReLU that is, after activation, there is an output layer to binary classification. Adam optimizer was used to train the model using the learning rate equal to 0.001 and weighted cross-entropy loss to address class imbalance neighbor To enhance scalability, neighbor sampling was used, in which 15 neighbors were sampled at the first layer and 10 neighbors on the second layer, which is made possible for mini-batch training with 1024 batch-size. This sampling strategy has a much lower memory demand than that of the competitors to full-graph training without relational information loss. PyTorch Geometric and Implementation of the model In the model, PyTorch Geometric was used to trained in a cloud environment based on a GPU. Performance was measured by the conventional measures such as Accuracy, With specific attention to Precision, Recall, F1-score and ROC-AUC rewarded by focusing on recall and ROC-AUC because of the imbalance type of fraud detection activities.

IV. Proposed Methodology

This study proposes a scalable Graph Neural Network–based framework for fraud detection in large-scale transaction datasets. The methodology focuses on representing transactions as a graph and applying a sampling-based GraphSAGE model to enable efficient and scalable training. The overall workflow consists of four main stages: data preprocessing, graph construction, scalable Graph Neural Network training, and fraud classification.

A. Data Preprocessing and Feature Preparation

The transaction dataset is first preprocessed to prepare node features and labels for graph learning. Each transaction is associated with a fraud label indicating whether it is fraudulent or legitimate. Numerical and categorical features such as transaction amount, card identifier, address identifier, and product category are selected as node attributes. Categorical features are converted into numerical representations using encoding techniques, and feature normalization is applied to improve training stability.

To ensure computational feasibility while maintaining realistic graph structure, a large subset of transactions is used to construct the graph. This allows evaluation of model performance under scalable conditions.

B. Transaction Graph Construction

The transaction data is represented as a graph, where each node corresponds to a transaction. Edges are created between transactions that share common attributes such as card identifier, address information, or product category. These connections enable the model to capture relationships between transactions that may indicate coordinated or suspicious activity.

This graph representation allows the model to utilize both individual transaction features and relational information between transactions, which is essential for detecting complex fraud patterns.

C. Scalable Graph Neural Network Model

To learn meaningful representations from the transaction graph, this work employs the GraphSAGE architecture. GraphSAGE is specifically designed for scalable learning on large graphs by aggregating information from neighboring nodes.

Unlike traditional graph convolution methods that require full graph processing, GraphSAGE supports inductive learning and efficient computation.

The GraphSAGE model consists of multiple graph convolution layers that update node representations by combining node features with information from neighboring nodes. These learned representations capture structural patterns and contextual information that help distinguish fraudulent transactions from legitimate ones.

D. Mini-Batch Training Using Neighbor Sampling

One of the key contributions of this work is the use of neighbor sampling to enable scalable training. Instead of processing the entire graph at once, a fixed number of neighboring nodes are sampled during each training iteration. This allows the model to be trained using mini-batches, significantly reducing memory usage and computational cost.

Mini-batch training makes it possible to scale the model to large transaction graphs while preserving important relational information. This approach enables efficient training without requiring full graph loading, which is often impractical for large financial datasets.

E. Fraud Classification

The learned node representations are used to classify each transaction as fraudulent or legitimate. A classification layer is applied on top of the GraphSAGE model to produce fraud predictions. To address class imbalance in fraud detection, weighted loss is used during training to ensure that fraud cases receive appropriate importance.

The proposed framework enables scalable fraud detection by combining graph-based learning with efficient sampling techniques. This approach improves the model's ability to detect fraud while maintaining computational efficiency for large-scale transaction data.

V. Results and Discussion

In this part, the results of the experiment are discussed. composed scalable Graph Neural Network framework and posed scalable Graph Neural Network framework compares its results to conventional machine learning models and full-graph Graph Neural Network models. The evaluation pays attention to performance in detecting frauds and also to the performance of detecting frauds. Neighbor sampling and scaling achieved by the benefits of neighbor sampling.

A. Baseline Model Performance

To set up a performance standard, conventional machine learning methods such as Logistic Regression and Random. The same dataset and feature was used to evaluate forest representation.

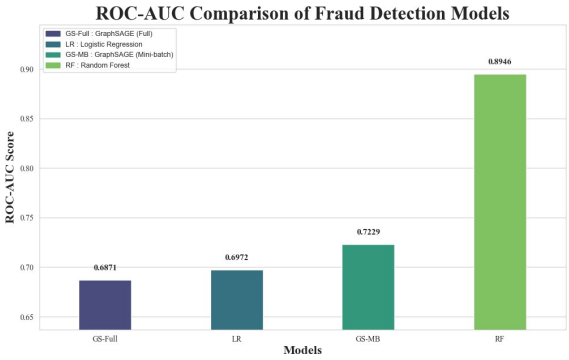

Fig. 2: ROC-AUC comparison of fraud detection models.

Logistic Regression was the one that attained a ROC-AUC score of 0.697 as shown in Fig. 2. This result indicates moderate capability in distinguishing fraudulent and legitimate transactions. However, the fraud recall was extremely low at 0.02, as illustrated in Fig. 3. This shows that the Logistic Regression could not be effective in identifying fraudulent transactions. This weakness can be explained by the fact that Logistic Regression works with each transaction separately and does not take into account the data on how transactions relate to each other.

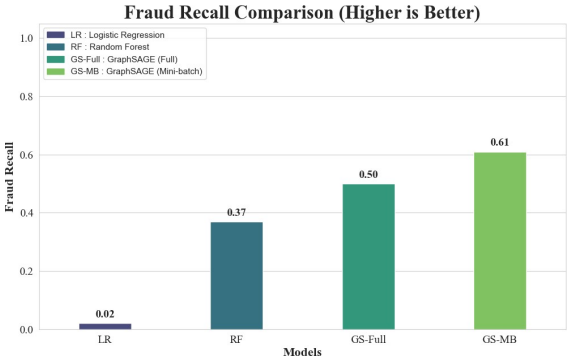

Fig. 3: Fraud recall comparison across different models.

B. Full-Graph Graph Neural Network Performance

Random Forest had much higher performance, and with ROC-AUC score of 0.895 and better fraud recall in comparison with Logistic Regression, as shown in Fig. 2. The fraud recall improved to 0.37, as illustrated in Fig. 3. The ensemble nature of Random Forest allows the improvement of modeling nonlinear feature. interactions, which leads to enhanced classification. Nevertheless, Random Forest remains an independent trans- operant is action-oriented and does not capitalize on relational information existing in transaction networks.

C. Full-Graph Graph Neural Network Performance

The effectiveness of graph-based learning can be measured by the use of the GraphSAGE model has been trained by full-graph propagation. This strategy had a ROC-AUC score of 0.687 and 0.50, which is represented in Fig. 2 and Fig. 3 through fraud recall. These evidence indicates that with the addition of graph structure it is possible to the model to define relational dependencies between transactions, detecting frauds better than traditional machine learning models.

Full-graph training however has serious scalability issues. The full graph is computationally and memory intensive to process which is not feasible in case of large scale transaction data. This weakness limits the applicability of full-graph Graph Neural Networks in fraud detection in real-world.

D. Scalable Graph Neural Network Performance

The suggested scalable Graph Neural Network model employs neighbor sampling and mini-batch training to enhance scalability and ability to learn relationally. The scalable GraphSAGE model obtained ROC-AUC of 0.723, better than the full-graph GraphSAGE model.

The scalable GraphSAGE model achieved a ROC-AUC score of 0.723, outperforming the full-graph GraphSAGE model, as shown in Fig. 2. More importantly, the fraud recall improved to 0.61, as illustrated in Fig. 3, which is the highest among all evaluated models.

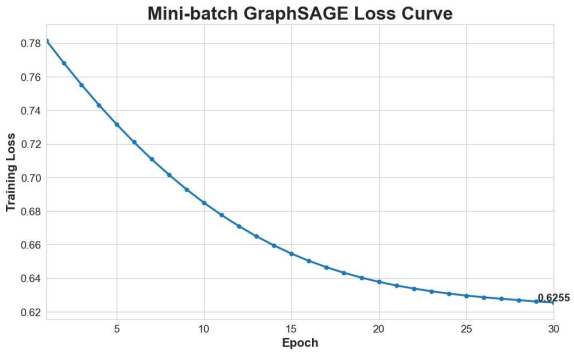

Fig. 4: Training loss convergence of the scalable GraphSAGE model.

The training loss curve shown in Fig. 4 demonstrates stable convergence of the scalable GraphSAGE model. The loss consistently decreases across training epochs, indicating effective learning of transaction patterns.

But even though the percent accuracy 2 of the scalable in general classification is high, Random Forest had a higher model than Random Forest, so this is ex-anticipated because of the gross disproportion of fraud detection datasets recall is more important in fraud detection programs important measure than accuracy since the ultimate aim is to achieve is to properly detect fraudulent transactions.

E. Comparison of Model Performance

The results of the performance of the assessed mod are in Table I else, such as Logistic Regression (LR), Random Forest (RF),GraphSAGE (1) Full Graph model, and GraphSAGE (2) Graph Neural Networks Mini-Batch model can be scaled.

TABLE I: Performance Comparison of Fraud Detection Models

Model	ROC-AUC	Accuracy	Fraud Recall	Precision
LR	0.697	0.96	0.02	0.39
RF	0.895	0.98	0.37	0.89
GraphSAGE (1)	0.687	0.79	0.50	0.08
GraphSAGE (2)	0.723	0.73	0.61	0.08

These findings show that the GraphSAGE (2) scalable mini batch model has the highest recall of frauds as compared to all evaluated models. This proves the efficiency of relational learning and sample of neighbors in detecting fraudulent transaction patterns. The ROC-AUC comparison shown in Fig. 2 and the fraud recall comparison shown in Fig. 3 clearly.

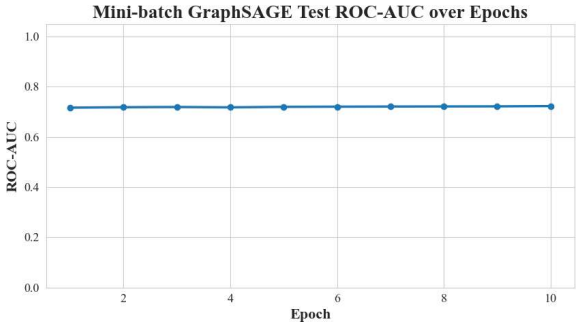

Fig. 5: ROC-AUC performance of the scalable GraphSAGE model across epochs.

F. Scalability Analysis

The mini-batch training and neighbor sampling are significantly scalable to full-graph training. Rather than working on the entire graph at each training iteration, the model only processes a fraction of the nodes and their sampled neighbors. This minimizes the consumption of memory and facilitates effective training of large transaction graphs.

Graph Neural Networks are able to predict through the scalable framework can be used on large-scale tasks of detecting frauds where full-graph is used the training would be computationally infeasible.

VI. Conclusion and Future Work

In this paper, a scalable Graph Neural Network was shown. structure of fraud detection in huge transaction networks. The

proposed approach helps the model to capture financial transactions by modeling them in the form of a graph structure. relational dependencies among transactions which tend to be of a frequent nature. is suggestive of a fraudulent nature. Unlike traditional machine procedures of learning by analyzing transactions separately, the graph-based method takes into consideration both feature-level and structural data to enhance the capacity of detecting fraud.

To overcome the limitations of scaling of traditional Graph. This work used a neighbor, and this was done through Neural Network training. mini-batch training with sampling and the GraphSAGE architecture. This is a scalable learning scheme which eliminates memory and computation costs by considerable amounts which is appropriate when training large transaction graphs. The scalable Graph Neural was proven to work experimentally. Network realized a ROC-AUC score of 0.723 and had an improvement. fraud recall versus full-graph Graph Neural Networktraining. The effectiveness is indicated in the improved fraud recall. of relational learning in fraudulent transactions detection.

The traditional machine learning models including the Random Forest, although demonstrating better overall accuracy, are entirely based on relying on one factor. Work on single transaction characteristics and fail to enumerate relational. fraud patterns. On the contrary, the suggested scalable Graph Neural. Network framework offers better detection of fraud due to its graph structure capability and without compromising. computational efficiency.

The findings validate the scalable Graph Neural Networks. provide a viable and working solution to detect fraud in. massive financial systems. The use of neighbor sampling allows achieving efficient training without needing to fully process the graph, which makes the method appropriate to deploy in practice. where the volumes of transactions keep on growing.

The proposed framework can be further improved in the future work. in several directions. First, detecting performance can be enhanced by adding other features of transactions and identity information. Second, the temporal graph modeling is able to. be examined to trace time varying fraud trends. Third, such advanced sampling strategies and graph architectures. since Graph Attention Networks can enhance representation. learning. Lastly, incorporation of real-time streaming feature. would allow using in online identification of frauds. where a prompt identification is necessary. The effectiveness can also be enhanced by these improvements. and scalability of Graph Neural Network based fraud detection frameworks in real-life financial settings.

REFERENCES

[1] E. W. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[2] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.

[3] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*, vol. 34, pp. 1–14, 2010.

[4] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

[5] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection using machine learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784–3797, 2015.

[6] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *International Conference on Learning Representations (ICLR)*, 2017.

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020.

[8] W. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[9] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, "Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks," in *ACM SIGKDD*, 2019.

[10] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[11] M. Weber, G. Domeniconi, J. Chen, D. K. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson, "Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics," in *ACM SIGKDD*, 2019.

[12] Z. Liu, Y. Dou, and P. S. Yu, "Heterogeneous graph neural networks for malicious account detection," *ACM Conference on Information and Knowledge Management*, 2018.

[13] IEEE Computational Intelligence Society and Vesta Corporation, "Ieee-cis fraud detection dataset," https://www.kaggle.com/competitions/ieee-fraud-detection/data, 2019, accessed: 2026-02-14.

