World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Graph attention networks for credit card fraud detection: A relational learning approach

Collin Arnold Kabwama [1, *], Pius Businge [1], Curthbert Jeremiah Malingu [1], Jude Innocent Atuhaire [1], Ian Asiimwe Ankunda [1], Joram Gumption Ariho [1], Brian Mugalu [1] and Denis Musinguzi [2]

[1] *Department of Computer Science, Maharishi International University, Fairfield, Iowa, USA.*

[2] *Department of Electrical and Computer Engineering, Makerere University, Kampala, Uganda*

World Journal of Advanced Research and Reviews, 2025, 26(03), 2580-2585

Publication history: Received on 12 May 2025; revised on 18 June 2025; accepted on 20 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.2400

Abstract

Credit cards have proliferated across the financial sector, enhancing accessibility but also creating new targets for fraud. Fraudsters often use subtle, coordinated techniques that are difficult to detect in isolation. This makes credit card fraud detection a suitable task for Graph Neural Networks (GNNs), which can model and analyze the relationships between entities like users, transactions, and devices. In this paper, we apply Graph Attention Networks (GAT) to a simulated credit card transaction dataset to detect fraudulent transactions. We show that they outperform traditional methods by leveraging relational patterns in the data when trained on the same features. Our findings highlight the promise of GNNs for financial fraud detection, particularly in uncovering complex, hidden connections that are not apparent through standalone analysis.

Keywords: GNNs; Gats; Fraud Detection; Deep Learning

1. Introduction

The global integration of information technology into financial systems has played a key role in the rapid development of global economic integration. This has led to an increase in the scale and complexity of financial sector transactions. However, this integration has resulted in an increase in financial fraud risk. According to the Federal Trade Commission (FTC), losses to fraud in 2024 rose to 12.5 billion dollars which represents a 25% jump from the previous year (ftc). Credit card fraud takes up a large percentage of this reported loss as a result of widespread adoption of credit cards for both online and in-person transactions. Financial fraud not only results in financial loss to the victims but also erodes public trust in financial systems and risks losing the progress made towards a cashless economy [1]. This underscores the urgency of developing effective fraud detection systems. Detecting fraudulent transactions at scale requires tools that can analyze massive volumes of data and uncover subtle, hidden patterns—an area where machine learning (ML) has shown great promise.

Machine learning models have been successfully applied to make predictions in multiple fields including real time crime prediction, disaster response optimization and threat detection to improve public safety and emergency management [2]. In fraud detection, classical machine learning models such as Bayesian belief networks, decision trees, and logistic regression [3] have achieved significant success, however, they often fall short due to their reliance on manually engineered features and their difficulty in capturing the dynamic and interconnected nature of fraud. Deep learning methods have improved upon this by enabling automatic feature extraction, offering better generalization and higher detection rates. However, they often treat transactions independently, failing to exploit the relational structure inherent in financial systems—such as shared devices, merchants, or user behavior.

[*] Corresponding author: Collin Arnold Kabwama.

Graph Neural Networks (GNNs) have recently emerged as a powerful tool for fraud detection. GNNs can analyze complex networks represented as graphs such as transaction histories, connections between accounts and links between merchants [4]. By organizing these transactions as graphs where nodes represent accounts and edges represent transactions, GNNs can find patterns that indicate fraud. They are particularly well-suited to this problem for several reasons. First, GNNs can automatically learn expressive representations by aggregating information from a node's neighborhood [5], making them ideal for modeling the interconnected nature of financial data. Second, they can dynamically adapt to evolving fraudulent behaviors by capturing temporal and structural changes in transaction graphs [6]. Third, GNNs offer a natural way to integrate heterogeneous data sources—such as user metadata, transaction histories, and device fingerprints into a unified representation [7]. GNNs have achieved significant progress and exhibited superior performance in fraud detection tasks [8].

In this study, we use the Graph Attention Network (GAT), a type of graph neural network that uses attention to weigh the importance of each node, to classify transactions in a simulated credit card dataset and compare its performance. Our goal is to evaluate how well these models capture the complex dependencies involved in fraudulent activities and to demonstrate their utility in building robust, scalable fraud detection systems.

2. Related work

Recent studies have demonstrated the effectiveness of Graph Neural Networks (GNNs) in financial fraud detection. For instance, CaT-GNN [9] enhances credit card fraud detection by incorporating causal temporal dynamics into GNNs, allowing the model to capture both temporal patterns and causal dependencies in transaction sequences. Similarly, heterogeneous graph representation learning has proven valuable by modeling the multi-dimensional nature of financial systems. This approach has shown strong performance in detecting fraudulent behaviors within complex social networks [10].

One of the central challenges in fraud detection is the class imbalance between fraudulent and legitimate transactions, with fraud cases representing a small minority. To address this, researchers have explored oversampling techniques and specialized loss functions designed to emphasize the minority class during training.

Another major challenge is heterophily [11] , where fraudulent accounts intentionally mimic legitimate behavior by connecting with trusted nodes or conducting normal transactions before fraudulent activity. Standard GNNs struggle with such mixed neighborhoods. To mitigate this, models like CARE- [9] employ a learned neighborhood selector to focus on informative neighbors, while PC-GNN [12] introduces a learnable distance function and a class-balanced sampler to reduce feature dilution and enhance robustness in heterophilic graphs.

A further limitation in the field is the scarcity of labelled real-world data, often due to privacy concerns. To overcome this, Xu et al. [13] proposed FMGAD, a fewshot message-enhanced contrastive learning framework that combines self-supervised learning with message-passing to detect anomalies in low-label settings—an approach particularly suitable for fraud scenarios with limited ground truth data.

As fraud tactics evolve rapidly, the need for dynamic and adaptive GNN models has become increasingly important. These models aim to learn from incoming data in real time, staying ahead of sophisticated fraud strategies. One such effort is the Spatial–Temporal Gated Network [14], which captures both temporal and spatial dependencies in transactional data, enabling better modeling of changing fraud patterns.

3. Methods

3.1. Dataset

We used the IBM credit card transaction dataset [15]. It is a publicly available dataset that contains information about credit card transactions. The dataset includes features such as the amount of a transaction, the type of credit card used, the location of the transaction, the time of the transaction, card number, merchant number, and merchant category code. It includes a label on whether or not a transaction is fraudulent. The dataset is designed to be representative of real-world settings and therefore contains a high level of class imbalance, whereby the number of legitimate transactions outnumber fraudulent transactions by a large margin. It was generated by simulation but the exact process of simulation is not specified. This dataset is widely used to build and evaluate credit card fraud detection models. It contains 24 million unique transactions with 6,000 merchants and 100,000 credit cards. It has 30,000 fraudulent transactions making up 0.1% of the dataset.

3.2. Graph Construction

A suitable graph structure for credit card fraud detection should be heterogeneous, capturing the complex relationships between different types of entities involved in transactions. In our graph, nodes represent merchants and credit card holders, while edges represent transactions characterized by attributes such as time, amount, credit card type, and merchant category code.

Before constructing the graph, we performed pre-processing on the dataset. Missing values in the error column were filled with “no error”, and transaction amounts were normalized using z-score normalization. We encoded categorical features like merchant city and merchant category code to prepare them for graph representation.

We used the NetworkX library to build the graph and opted for a MultiGraph rather than a standard graph, since customers can interact with the same merchant multiple times. To construct the graph, we created unique nodes for each card and merchant, then iterated through the dataset, adding an edge for each transaction. Each edge included the transaction features and its label (fraudulent or not).

The dataset was split into training and test sets using a 4:1 ratio, with masks identifying the partitions. The final graph consisted of 16,317 nodes and 99,757 edges, with 7 features per edge. Of the total edges, 79,805 were used for training and 19,952 for testing.

3.3. Model

The Graph Attention Network (GAT) [16] is a neural network architecture designed for graphstructured data. Its key innovation lies in the use of an attention mechanism to perform node representation learning. GAT allows each node to learn and assign different weights to its neighbors based on their features, thereby enabling a more flexible and expressive representation.

In GAT, each node first applies a shared linear transformation to its feature vector, projecting it into a new feature space. Then, for every neighboring node pair, an attention score is computed using a shared attention mechanism, which takes the concatenation of the transformed features of the source and target nodes, applies a learnable weight vector, and passes the result through a LeakyReLU activation function. These attention scores are then normalized across all neighbors of a node using the softmax function, producing attention coefficients that reflect the relative importance of each neighbor.

The new representation of a node is obtained as a weighted sum of its neighbors transformed features, where the weights are the learned attention coefficients. To enhance model capacity and stability, GAT uses multi-head attention, where multiple attention mechanisms are run in parallel and their outputs are either concatenated for intermediate layers or averaged for the final layer.

3.4. Training

We build a graph neural network based on then graph attention network attention layer to classify the edges. The model has 2 graph attention layers with an elu activation function between them. The graph attention layers had 4 attention heads with a hidden dimension of 64. We used a dropout of 0.3 to prevent overfitting. We used this model for feature extraction and built a multilayer perceptron for classification. We trained the model using the Adam optimizer with a learning rate of 0.001 and a weight decay of $1e - 4$. We trained the model using the cross-entropy loss. Figure 1 shows the training and test loss curves of the model. The loss decreases sharply during the first 100 epochs, indicating rapid initial learning, and then continues to decline more gradually over the remaining training period.

4. Results

We evaluated the performance of our model using precision, recall, F1-score, and accuracy. Given the class imbalance in the dataset, precision, recall, and F1-score are more informative metrics than accuracy.

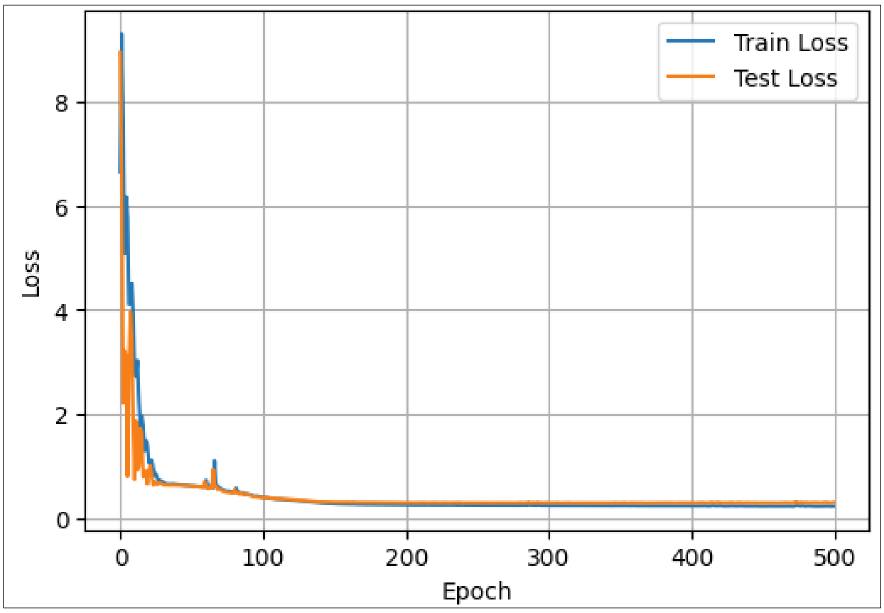

Figure 1 Training and test loss curves of the Graphical Attention

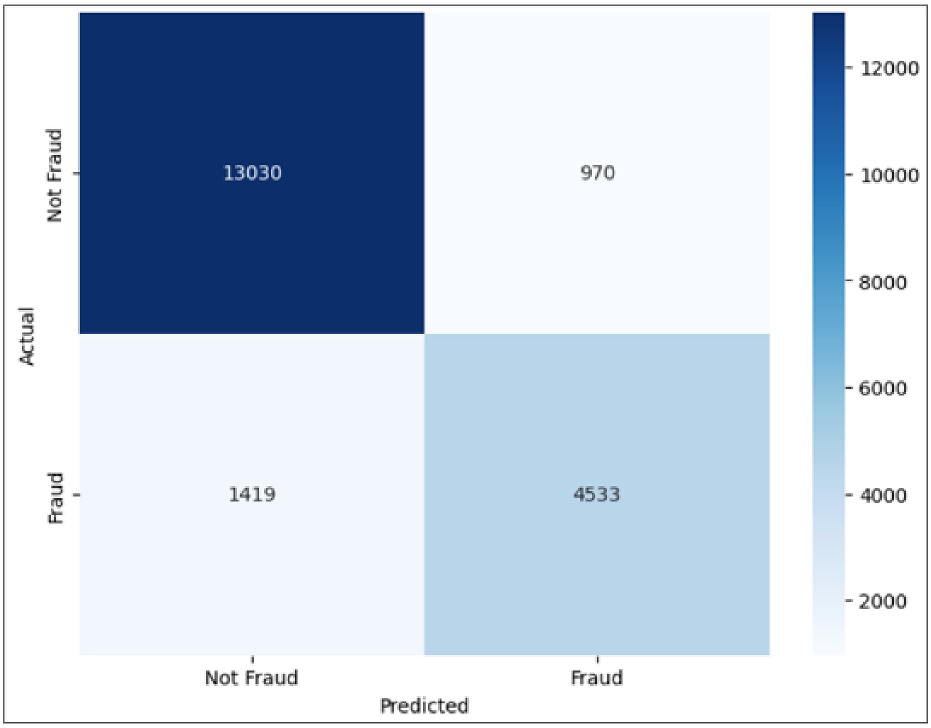

Figure 2 Confusion matrix showing the results of the Graph At-

To benchmark our model, we compared its performance with a random forest classifier trained on the same training set and evaluated on the same test set. The random forest model was configured with 100 estimators. Table 1 shows the results from both models. The GAT model outperforms the random forest model across all the metrics. Figure 2 shows the confusion matrix for the GAT model. The model shows a good balance between fraud detection and minimizing false positive, though the relatively high number of false negatives suggests room for improvement in sensitivity to fraudulent activity.

5. Discussion

Our results demonstrate that the Graph Attention Network (GAT) effectively captures the complex structure of transaction data, achieving superior performance compared to traditional machine learning methods.

Table 1 Performance of the Graphical Attention Network model on the test dataset

Metric	GAT Model	Random Forest
Accuracy	88.0%	86.7%
Precision	86.3%	85.7%
Recall	84.6%	83.8%
F1 score	85.4%	84.7%

The GAT model achieved an accuracy of 88.0%, precision of 86.3%, recall of 84.6%, and an F1-score of 85.4%, outperforming the baseline Random Forest classifier across all metrics. This suggests that leveraging the relational structure between entities—such as merchants and cardholders—significantly improves fraud detection performance.

Nevertheless, several challenges remain. The model's performance is still influenced by the underlying class imbalance and the quality of the transaction graph. Improvements could be made by incorporating temporal dynamics, richer node features, or more advanced sampling techniques to address heterophily and sparsity in the graph. Additionally, explainability remains a challenge for real-world deployment; integrating interpretable GNN methods may help build trust with financial institutions and regulators.

In summary, our study confirms the promise of GNNs specifically GATs for fraud detection in complex financial networks. Future work will explore dynamic GNN models, real-time learning capabilities, and hybrid approaches that integrate GNNs with other anomaly detection techniques to further improve detection accuracy and operational scalability.

6. Conclusion

In this study, we investigated the application of Graph Neural Networks specifically the Graph Attention Network (GAT) for credit card fraud detection using a simulated transaction dataset. By modeling the data as a heterogeneous graph of merchants and cardholders, our approach effectively captured the relational dependencies inherent in financial transactions.

The GAT model outperformed a traditional Random Forest baseline across all major evaluation metrics, including accuracy, precision, recall, and F1-score. These results underscore the potential of attention-based GNNs in identifying subtle patterns of fraudulent behavior that may be missed by conventional classifiers.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] S. Motie and B. Raahemi, Financial fraud detection using graph neural networks: A systematic review. Expert Syst. Appl., 240(C), April 2024.

[2] N. Harriet, A. Brian, T. Evans, Z. Ivan, S. Iga, B. Jimmy and E. Pinyi, "Leveraging AI for real time crime prediction, disaster response optimization and threat detection to improve public safety and emergency management in the US," World Journal of Advanced Research and Reviews, 2024.

[3] I. D. Mienye and Y. Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection.," Applied Sciences, 2023.

[4] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, " Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," IEEE Access, 2022.

[5] F. Scarselli, M. T. A. C. Gori, M. Hagenbuchner and G. Monfardini, "The graph neural network model.," IEEE Transactions on Neural Network, 2009.

[6] Z. Zhang, X. Wang, Z. Zhang, H. Li, Z. Qin and W. Zhu, "Dynamic graph neural networks under spatio-temporal distribution shift," Advances in Neural Information Processing Systems, 2022.

[7] D. Cheng, F. Yang, S. Xiang and J. Liu, "Financial time series forecasting with multi-modality graph neural network," Pattern Recognition, 2022.

[8] D. Cheng, X. Wang, Y. Zhang and L. Zhang, " Graph neural network for fraud detection via spatial-temporal attention," IEEE Transactions on Knowledge and Data Engineering, 2022.

[9] Y. Duan, G. Zhang, S. Wang, X. Peng, W. Ziqi, J. Mao, H. Wu, X. Jiang and K. Wang, "K. Cat-gnn: Enhancing credit card fraud detection via causal temporal graph neural networks," 2024.

[10] B. Wu, K.-M. Chao and Y. Li, "Heterogeneous graph neural networks for fraud detection and explanation in supply chain finance," Inf. Syst., 2024.

[11] Y. Dou, K. Shu, C. Xia, P. S. Yu and L. Sun, "User preference-aware fake news detection," In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2021.

[12] Y. Liu, X. Ao, Z. Qin, J. Chi, J. Feng, H. Yang and H. Q, "Pick and choose: A gnn-based imbalanced learning approach for fraud detection," In Proceedings of the Web Conference, 2021.

[13] F. Xu, N. Wang, X. Wen, M. Gao, C. Guo and X. Zhao, "Few-shot message-enhanced contrastive learning for graph anomaly detection," 2023.

[14] Y. Xie, G. Liu, M. Zhou, L. Wei, H. Zhu, R. Zhou and L. Cao, " A spatial–temporal gated network for credit card fraud detection by learning transactional representations.," IEEE Transactions on Automation Science and Engineering, 2024.

[15] E. R. Altman, "R. Synthesizing credit card transactions," 2019.

[16] P. Velickoviˇc, G. Cucurull, A. Casanova, A. Romero, P. Li´o and Y. Bengio, "Graph attention networks," 2018.