# 1. WOX7001 RESEARCH METHODOLOGY 1/2025/2026

# A Comparative Analysis and Research Gap Identification of Graph Neural Networks for Financial Anomaly Detection

# 2. ASSIGNMENT 1

| | |
|---|---|
| Matric Number: | 24201617 |
| Name | ZOU TING |
| Occurrence: | Occ 1 (MCS/MCS (AC)/MSE (ST)) |
| Lecturer Name: | DR MUMTAZ BEGUM PEER MUSTAFA |

## Table of Contents

## Abstract

The rapid expansion of financial transaction networks and the increasing complexity of fraudulent activities have made graph neural networks (GNNs) a key focus in anomaly detection research.Literature has confirmed that GNNs excel at modeling complex multi-hop relationships among financial entities and outperform traditional machine learning in detecting organized fraud patterns.However, major challenges have hindered their practical deployment, including severe class imbalance, concept drift in dynamic environments, model interpretability lacking regulatory compliance, and scalability issues in real-time processing.This study aims to systematically analyze the application of various GNN models in financial risk scenarios, such as fraud detection and anti-money laundering.It will summarize the characteristics and advantages of their methods and overcome the limitations through a critical review of recent literature.The main achievement is to identify the core research gaps and bottlenecks that hinder the practical application of financial institutions.Subsequently, this study will formulate precise research questions and goals, providing a clear direction for future model development and empirical comparisons, and establishing a basic framework for the advancement of this field.

## 1. Introduction (Research Background)

The digital transformation of the financial industry, while enhancing efficiency, has also intensified the risks and complexity of fraud and money laundering activities.Traditional detection systems typically rely on predefined rules or classic machine learning models and often fail to identify complex collusive frauds, which manifest as subtle anomalies in vast and interrelated transaction networks[5].This restriction has promoted the exploration of advanced deep learning techniques, especially graph neural networks (GNNs), which offer a paradigm shift in analyzing relational data.

GNNs provides a natural and powerful framework for financial anomaly detection by representing the financial ecosystem graphically.In this structure, nodes typically represent entities such as users, accounts or merchants, while edges represent transactions, relationships or shared attributes.This representation enables GNNs to leverage node features and the topological structure of the network, allowing them to capture complex relational dependencies and multi-hop impacts, which are all signs of organized fraudulent activities[1].For instance, if a seemingly legitimate transaction is connected through the path of a previously marked account, it might be marked as suspicious.

In recent years, research on applying various GNN architectures, including Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), to key financial tasks such as credit card fraud detection, anti-money laundering (AML), and cryptocurrency transaction monitoring has surged.These innovations include methods for integrating domain knowledge, such as leveraging known suspicious entities (" strong nodes ") to enhance model performance [2].The investigation in this field highlights a rapidly developing landscape and has achieved promising results[6].Despite this passionate academic pursuit and the potential demonstrated in controlled experiments, there remains a significant gap between the research prototype and the robust and reliable system deployed in the production environment of financial institutions.Operational challenges related to data, model adaptability and compliance remain largely unresolved.Therefore, this study aims to clearly clarify the unsolved practical problems by reviewing the current most advanced technologies and establish a direct approach for meaningful future research to bridge this gap, thereby providing a timely and structured foundation.

## 2. Research Problem Statement

The practical deployment of graph neural networks (GNNs) in financial anomaly detection is hindered by the significant gap between their potential demonstrated in research and their operational robustness in real financial institutions.Although they have outstanding capabilities in modeling complex transaction relationships, some interrelated challenges remain unresolved, which brings significant pain points for industry adoption.

The core issue is that the existing GNN methods are difficult to overcome the specific operational limitations in the financial field.Firstly, the inherent extreme class imbalance in financial data makes GNN models tend towards the majority of classes, seriously damaging their recall rate and robustness.Although some studies have addressed this issue through sampling techniques [3], others have proposed new architectural changes to guide the focus of the model [2].Secondly, the dynamic nature of financial networks leads to concept drift, and trading models evolve over time.Many GNN models designed for static graphs cannot adapt effectively, resulting in rapid performance degradation [4].Thirdly, the lack of model interpretability remains the main obstacle for financial institutions to comply with regulatory requirements.The "black box" nature of many GNNs makes it difficult to provide actionable explanations for marking anomalies[1].Finally, due to the use of private or synthetic datasets, there is a lack of unified benchmarks in this field, making it difficult to fairly compare methods and ensure universality [1].

Therefore, the existing research has not enabled practitioners to fully possess the ability to overcome these core operational obstacles.

## 3. Research Questions

3.1. How to effectively adapt to the GNN architecture to mitigate the performance degradation caused by extreme class imbalance in financial transaction data?

3.2. Which GNN variants and methods are most suitable for handling concept drift and the temporal evolution of financial transaction graphs?

3.3. To what extent do the existing GNN interpretability technologies meet the interpretability requirements of financial regulatory authorities in real audit scenarios?

3.4. What are the main scalability bottlenecks of the most advanced GNN models when dealing with large-capacity, streaming financial data?

## 4. Research Objectives

To systematically address the research questions outlined in Section 3, this study established the following specific research objectives (ROs), which are designed to be concrete, measurable, achievable, relevant and time-bound (SMART) :

4.1. RO1: To identify and systematically compare at least three advanced techniques for handling imbalanced data (e.g., cost-sensitive learning, advanced sampling) by the end of the literature review phase.Their integration with the GNN architecture will be evaluated to determine whether they can achieve at least a 15% improvement in f2 scores on the baseline GNN model and report on at least two different benchmark datasets.

4.2. RO2: To analyze and classify dynamic GNN methods (e.g., time graph networks) from the literature.This analysis will generate a conceptual framework (completed before the key analysis writing stage), which outlines their applicability, key mechanisms and performance in detecting anomalies in the evolving financial transaction graph.

4.3. RO3: To critically evaluate at least four GNN explainability methods (such as gnexplainer, PGExplainer) by formulating a set of financial compliance standards (e.g., operability, fidelity, and understandability).This assessment will generate an evaluation matrix (as the main output of critical analysis), clearly comparing its advantages and disadvantages with established standards.

4.4. RO4: Based on the performance reported on large-scale datasets, analyze the computational costs (e.g., inference latency, memory usage) of three prominent GNN architectures (such as GCN, GAT, and GraphSAGE).This analysis will identify key bottlenecks and ultimately propose at least two advanced architectural optimizations for real-time reasoning, which will be outlined in the research abstract.

## 5. Research Scope and Contribution

### 5.1. Research Scope

To address the challenges identified in actual deployment, this study systematically analyzed and critically reviewed the academic literature on GNN in financial anomaly detection.The investigation will focus on peer-reviewed journal articles published after 2021, covering models designed for tasks such as payment fraud detection and anti-money laundering.It is worth noting that this study does not involve the main development, training or empirical testing of new algorithms or models.All analyses, conclusions and identified gaps will be derived from the synthesis and evaluation of the results reported in the selected literature.

### 5.2. Contribution

The positioning of this research is multi-faceted contributions.Theoretically, it will integrate the scattered and rapidly growing working entities into a unified overview, clearly depicting the progress and the persistent gaps.It aims to propose a structured framework to understand core challenges such as dynamic learning and interpretability.In fact, the research results will provide valuable references for fintech practitioners and developers. By conducting a clear comparative analysis of the methods and their limitations, they will inform the design and selection of more powerful and deployable GNN-based systems.In terms of methods, this study presents a rigorous documentary-based gap identification approach, laying the foundation for a systematic review in this field and providing a clear roadmap for future preliminary research, aiming to bridge the established gap between academic innovation and industrial application.

## 6. Critical Analysis

### 6.1. Literature Review

To reflect the necessity of this project, a large number of literatures have revealed a strong concern for solutions to the core operational challenges of development, including class imbalance, dynamic graphs, and interpretability.However, research strategies and technical methods vary greatly in different studies.To systematically obtain these dimensions, Table 1 provides a detailed technical comparison of methods, datasets, and results.It provide a comprehensive foundation for the subsequent critical analysis.

## 6.2. Critical Analysis

A critical synthesis of literature on financial anomaly detection based on artificial intelligence reveals a field characterized by rapid algorithmic innovation, but hindered by significant methodological and practical limitations.These limitations constitute the main obstacles to the transition from academic research to robust practical deployment.

### **6.2.1.** Methodological Limitations and Generalizability Concerns

The main weakness of the entire field lies in the limited generalizability and reproducibility.As shown in the works of Chen[2], the overwhelming reliance on private or synthetic datasets makes it impossible to independently verify dominant claims or conduct fair benchmarking.This fosters an environment in which models may overfit the characteristics of a single, inaccessible dataset, thereby casting doubt on their performance in other financial settings.Furthermore, although methods for handling dynamic graphs show promise [4], they typically introduce significant computational complexity[5,6,7,8].In a high-throughput financial environment, the trade-off between model complexity and actual low-latency processing is rarely quantified or discussed, which poses a major obstacle to production systems.

### **6.2.2.** Disconnect Between Evaluation Standards and Industrial Requirements

There is a profound disconnection between academic assessment indicators and the actual demands of the industry. Firstly, verification is almost universally carried out in offline settings, ignoring real-world challenges such as concept drift, operational data pipelines, and integration with legacy banking infrastructure[9].Secondly, there has always been a lack of systematic analysis of computational efficiency, including indicators such as large-scale inference latency, memory usage, and training time[9].Models that achieve marginal accuracy gains at the cost of tenfold increase in inference time are impractical for real-time fraud detection.Finally, the evaluation of interpretability techniques remains superficial.Research typically employs technical indicators such as fidelity, but does not allow domain experts to assess whether the provided explanations are truly understandable and feasible for key decisions (such as submitting a suspicious activity report [1]).

### **6.2.3.** Technical Convergence without Standardized Assessment

A key positive trend is the widespread adoption of dual-purpose attention mechanisms in models such as Graph Attention Networks (GATs) [8,10].These mechanisms allow nodes to weigh the influence of neighbors and identify suspicious local structures as a form of inherent interpretability, thereby simultaneously improving performance [8,10], a technique also employed in models like HHLN-GNN for feature aggregation [3].However, a key divergence emerged: although these studies utilized the focus on explainability, none of them established a standardized framework to assess the reliability of these explanations or to transform them into actionable insights for risk analysts.This highlights the consensus on the utility of these technologies, but there is a serious lack of a consistent, human-centered assessment of their explanatory power.

## 6.3. Research Gap

The above critical analysis reveals several obvious gaps in the current research field, which represents a direct opportunity for future work to bridge the gap between academic exploration and industrial application.

### 6.3.1. Lack of Standardized and Realistic Benchmarks

Currently, there is a serious lack of public, large-scale and realistic benchmark datasets for financial anomaly detection, especially those that contain dynamic time graphs.This gap fundamentally hinders repeatable research, fair model comparisons, and reliable evaluations of the universality of models among different financial institutions and types of fraud.

### 6.3.2. Human-centered interpretability evaluation

Although explainable AI (XAI) methods for GNNS are under active development, there are significant gaps in strictly evaluating their practical utility.Future research must test these explanations in collaborative studies with domain experts (for example, risk analysts) to determine whether they truly enhance audit and compliance workflows by providing reliable and actionable insights.

### 6.3.3. Lightweight and Streaming GNN Architectures

The focus in this field is mainly on enhancing the accuracy of complex models, which often comes at the expense of computational efficiency.Therefore, there is a significant gap in the dedicated design and comprehensive evaluation of lightweight GNN architectures specifically optimized for high-frequency, low-latency stream transaction processing in real-world production environments.

### 6.3.4. Integrated Frameworks for Data Dynamics

Most studies address the challenges of class imbalance and conceptual drift in isolation.A promising research gap lies in developing a unified GNN framework that can jointly and adaptively handle the severe class imbalance in the native evolutionary graph structure, thereby going beyond the limitations of the static batch processing paradigm.

### 6.3.5. Cross-Domain Generalization and Transfer Learning

The potential to transfer GNN models trained on the data of one financial institution to another or across different types of fraudulent activities (for example, from credit card fraud to anti-money laundering) remains largely unexplored.Systematic research on cross-domain and cross-institutional transfer learning of GNNs is a key gap, which has the potential to solve the problem of data scarcity and significantly increase the adoption rate of models.

## 7. Summary

This literature review integrates the latest progress of GNNs in financial anomaly detection, highlighting a field that is rich in methodological innovations but faces significant practical obstacles. Analysis confirms that although complex models have been developed to address imbalance, dynamics, and interpretability, their evaluations often lack rigor in terms of repeatability, computational efficiency, and real-world practicality. The main limitations are the reliance on non-standardized private data, the neglect of real-time performance constraints, and the insufficient verification of interpretability for end users. These strictly identified shortcomings directly map and verify the research questions expounded in Part A.They emphasize the necessity of shifting the research focus from pure academic performance indicators to solutions that are not only accurate but also repeatable, scalable, and truly applicable to financial institutions, thereby clearly demonstrating the proposed research questions and objectives.

## 8. Reference List

1. Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. Expert Systems with Applications, 240, 122156. https://doi.org/10.1016/j.eswa.2023.122156
2. Chen, J., Chen, Q., Jiang, F., Guo, X., Sha, K., & Wang, Y. (2024). SCN_GNN: A GNN-based fraud detection algorithm combining strong node and graph topology information. Expert Systems with Applications, 237, 121643. https://doi.org/10.1016/j.eswa.2023.121643

3. Tong, G., & Shen, J. (2023). Financial transaction fraud detector based on imbalance learning and graph neural network. Applied Soft Computing, 149, 110984. https://doi.org/10.1016/j.asoc.2023.110984
4. Nguyen, T. T., Phan, T. C., Pham, H. T., Nguyen, T. T., Jo, J., & Nguyen, Q. V. H. (2023). Example-based explanations for streaming fraud detection on graphs. *Information Sciences*, *621*, 319-340. https://doi.org/10.1016/j.ins.2022.11.119
5. Han, B., Wei, Y., Wang, Q., Collibus, F. M. D., & Tessone, C. J. (2024). MT 2 AD: multi-layer temporal transaction anomaly detection in ethereum networks with GNN. *Complex & Intelligent Systems*, *10*(1), 613-626. https://doi.org/10.1007/s40747-023-01126-z
6. Qian, J., & Tong, G. (2025). Metapath-guided graph neural networks for financial fraud detection. *Computers and Electrical Engineering*, *126*, 110428. https://doi.org/10.1016/j.compeleceng.2025.110428
7. Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal heterogeneous graph contrastive learning for fraud detection in credit card transactions. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3599787
8. Khosravi, S., Kargari, M., Teimourpour, B., & Talebi, M. (2025). Transaction fraud detection via attentional spatial–temporal GNN. *The Journal of Supercomputing*, *81*(4), 537. https://doi.org/10.1007/s11227-025-06983-8
9. Deprez, B., Wei, W., Verbeke, W., Baesens, B., Mets, K., & Verdonck, T. (2025). Advances in Continual Graph Learning for Anti-Money Laundering Systems: A Comprehensive Review. *Wiley Interdisciplinary Reviews: Computational Statistics*, *17*(3), e70040. https://doi.org/10.1002/wics.70040
10. Li, E., Ouyang, J., Xiang, S., Qin, L., & Chen, L. (2025). Efficient relation-aware heterogeneous graph neural network for fraud detection. *World Wide Web*, *28*(5), 55. https://doi.org/10.1007/s11280-025-01369-5

Table 1: Comparative Analysis Table

| Author(s) & Year | Research Focus | Methodology | Dataset/Context | Key Findings | Strengths | Weaknesses/Limitations |
|---|---|---|---|---|---|---|
| Soroor Motie, Bijan Raahemi (2024) [1] | Systematic review of GNNs in financial fraud detection, proposing new GNN classification and identifying research trends. | Systematic Literature Review (SLR) | 33 selected papers covering cryptocurrency, credit card, etc | GNN is good at capturing complex transaction relationships;<br>Most studies focus on static graphs and supervised learning; | New GNN classification;<br>Identify comprehensive future research directions | Search restrictions may miss relevant papers;<br>Exclude gray literature; |
| Jing Chen et al. (2024)[2] | SCN_GNN algorithm for fraud detection in sparse multi-relation graphs using strong nodes and topology. | GNN algorithm development with structured similarity-aware module combines up-sampling, down-sampling, and improved reinforcement learning modules | YelpChi dataset and Amazon dataset | SCN_GNN outperforms state-of-the-art methods, effectively handling graph sparsity and camouflage behaviors | Innovative combination of topology and strong node information;<br>Effectively handle imbalanced datasets | Limited improvements to already dense graphs;<br>Higher algorithmic complexity;<br>requires hyperparameter tuning |
| Tong & Shen (2023)[3] | HL-GNN model for fraud detection handling class imbalance and heterogeneous/homogeneous graph connections | imbalance processor, join classifier, self-concerned aggregator, and prototype-based discriminator | Elliptic (Bitcoin transaction fraud), YelpChi (review fraud), Amazon (review fraud) | HL-GNN Significantly outperforms baselines in F1-macro, AUC, and GMean metrics.Ablation studies have confirmed the necessity of imbalanced treatment and homogeneous/heterogeneous modules. | 1. Comprehensively addresses imbalance and heterogeneity;<br>2. Clear model design with detailed algorithms. | 1. Train relying on labeled data;<br>2. Do not simulate the temporal dynamics of transactions;<br>3. It does not contain unstructured information (for example, text) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Nguyen et al. (2023)[4] | Interpretable AI for fraud detection by retrieving similar historical fraud subgraphs | Query-by-example framework with graph embedding and optimized flow settings | Real world: Amazon, YelpCh, Books.<br>Synthetic: Generated using infection (SI, SIS, SIR) and influence (IC, LT) models | 1. Provides meaningful explanations in <1 second<br>2. Embedding is much faster than traditional methods | 1.Model-agnostic;<br>2. Balances similarity and diversity;<br>3. Efficient for real-time use | 1. Post-hoc explanations depend on historical data quality;<br>2. Requires parameter tuning. |
| Han et al. (2024)[5] | Anomaly detection in the Ethereum transaction network | A Graph Attention Network models the dynamic multi-layer transaction graph for detection. | Ethereum token transaction networks (BNB, USDT, LNK) | MT$^2$AD achieved better performance in identifying abnormal transaction patterns compared with the baseline method | Effectively capture cross-cryptocurrency transaction patterns and temporal dynamics | 1. A large amount of labeled data is required for training<br>2. High computational complexity |
| Qian & Tong (2025)[6] | Financial fraud detection using Graph Neural Networks | Metapath-based subgraph generation, attention mechanism, semi-supervised learning | YelpChi, Amazon, Elliptic, T-Finance | Handles class imbalance and hidden patterns to outperform best baselines, boosting F1 by up to 11.33% and AUC by 3.54%. | 1. Combines local and global graph information<br>2. Uses semi-supervised learning to reduce labeling cost | 1. High computational cost<br>2. relies on predefined metapaths and hyperparameter tuning |

| | | | | | | |
|---|---|---|---|---|---|---|
| Wang et al. (2025)[7] | Credit card fraud detection using temporal heterogeneous graphs and contrastive learning | Proposes TH-GCL framework integrating temporal GNNs, heterogeneous graphs, and dual-view contrastive learning | Real-world IEEE-CIS and Credit Card Fraud Detection datasets with extreme class imbalance | Achieves SOTA performance (96.1% AUC-ROC) and superior zero-day fraud detection capability | Effectively handles temporal dynamics, data sparsity, and complex multi-entity relationships | 1. High computational complexity<br>2. High dependency on quality graph construction |
| Khosravi et al. (2025)[8] | Transaction fraud detection based on Attention spatio-temporal graph neural network | An AST-GNN with graph autoencoder reconstruction, temporal coding, spatial aggregation (internal/interrelationship) and transformation layer was proposed | YelpChi , Amazon ; Real-world: Iranian bank transactions | AUC on Amazon and YelpChi Outperforms 13 baselines including GCN, GAT, CIES-GNN | 1. Handle node camouflage<br>2. Capture long path dependencies<br>3. Effectively process imbalanced data<br>4. Integrate global and local information | 1. High computational complexity, 2. Requires careful hyperparameter tuning<br>3. Limited real-time application potential |

| | | | | | | |
|---|---|---|---|---|---|---|
| Deprez et al. (2025) [9] | Evaluate the continuous learning strategies for graph neural networks under real-world anti-money laundering constraints | Evaluates replay (GEM), regularization (EWC, MAS), and architecture-based (LwF, TWP) CL methods on GCN backbone | IBM AML (synthetic), Elliptic (Bitcoin transactions) | 1. Replay (GEM) & topology-aware (TWP) methods reduce forgetting<br>2. Wider/deeper models increase overfitting & forgetting | 1. Addresses real-world constraints (data retention, computation)<br>2. comprehensive CL method comparison | 1. Strong task similarity leads to forgetting<br>2. Hyperparameter sensitive;<br>3. Only for specific graphic CL Settings |
| Li et al. (2025)[10] | Heterogeneous graph fraud detection based on relation-aware GNN and knowledge distillation | Proposes RHGNN (relation-aware preprocessing, hybrid propagation) & DRHGNN (distilled MLP for efficiency) | T-Finance, Yelp, Amazon (financial/social media fraud detection) | RHGNN achieves SOTA accuracy (e.g., 0.9734 AUC on Amazon); DRHGNN improves inference speed 2-4× with minimal accuracy loss | 1. Handles graph heterogeneity effectively;<br>2. Balances accuracy and efficiency;<br>3. Comprehensive evaluation on multiple datasets | 1. Complex model design<br>2. Requires careful hyperparameter tuning<br>3. Distillation may slightly reduce peak performance |