UNEC Journal of Computer Science and Digital Technologies
Volume 1, № 2, 2025, pp. 60-75
https://doi.org/10.30546/UNECCSDT.2025.02.1016

Section:
Cybersecurity & Privacy

Transductive vs. Inductive Graph Learning for Illicit Bitcoin Transaction Detection

Elkhan Karimzada[1], Nazrin Bayramova[2], Elmir Adilzade[3]

[1]Ministry of Internal Affairs of the Republic of Azerbaijan, Baku, Azerbaijan
[2] Department of Digital Technologies and Applied Informatics, UNEC, Baku, Azerbaijan*
[3]Sabah Groups, ASOIU, Azerbaijan
[1]0009-0002-9423-5320, e.karimzade@mia.gov.az
[2]0009-0002-7841-7239, bayramova.nazrin.yashar.2022@unec.edu.az
[3]0009-0000-6126-7700, elmira21620@sabah.edu.az

Abstract

The Bitcoin transaction network represents a complex, high-dimensional system where the pseudonymous nature of flows facilitates ransomware extortion and money laundering activities. However, existing detection methodologies predominantly rely on local transaction features, failing to capture structural dependencies and topological obfuscation techniques used in ransomware and laundering. To address this limitation, this study proposed a hybrid Graph-Based Machine Learning (GBML) framework that integrates structural embeddings (Node2Vec) with ensemble classifiers and contrasts them against inductive GraphSAGE architectures using the Elliptic dataset. The analysis revealed that the Node2Vec-enhanced Random Forest model achieved an F1-score of 0.9277 and a ROC-AUC of 0.9956, substantially outperforming feature-only baselines. Furthermore, the inductive GraphSAGE model demonstrated remarkable robustness under a strict temporal split, achieving an F1-score of 0.8981 on future unseen transactions. This performance improvement is attributed to the encoding of neighborhood context and temporal dynamics, which exposes latent connections between illicit entities. Distinct from purely black-box deep learning approaches, this framework incorporates latent space visualization and permutation-based feature importance to ensure the forensic interpretability required for legal compliance. Consequently, the proposed method provides a robust solution for post-hoc forensic investigations in environments characterized by extreme class imbalance and evolving criminal patterns.

***Keywords*:** Cryptocurrency Forensics, Anti-Money Laundering (AML), Structural Embeddings, Imbalanced Classification, Network Topology, Ransomware Tracing

Received:	*Revised:*	*Accepted:*	*Published:*
8/12/2025	*16/12/2025*	*19/12/2025*	*23/12/2025*

1. INTRODUCTION

The integrity of decentralized financial systems represents a critical challenge in modern cybersecurity due to the dual nature of blockchain technology. While the pseudonymous architecture of Bitcoin facilitates billions of dollars in legitimate global trade and investment , it simultaneously provides an obfuscated infrastructure for illicit activities, including ransomware extortion, darknet markets, and money laundering operations. Recent reports indicate that while illicit transactions constitute less than one percent of total network activity,

the absolute volume of criminal funds laundered through these networks poses severe risks for regulatory compliance and national security. Furthermore, the rapid evolution of obfuscation tactics such as cross-chain laundering and tumbling services underscores the contemporary urgency of transitioning from static, rule-based monitoring to dynamic, algorithmic forensic solutions.

In the context of blockchain forensics, "licit transactions" refer to financial activities that adhere to legal standards and are not associated with known criminal entities. Distinct from these are "illicit transactions," which are specifically defined as transfers linked to labeled criminal addresses, such as those involved in ransomware payments or theft. To analyze these patterns, this study operates within the domain of Graph Learning (GL), a paradigm that explicitly represents data as a network where nodes correspond to entities (transactions) and edges encode financial flows. Unlike traditional tabular analytics, GL specifically involves the exploitation of structural dependencies and topological context to infer the legitimacy of a transaction.

Research in cryptocurrency forensics has evolved along two primary trajectories. The first stream, exemplified by Meiklejohn et al. and Ron and Shamir, focused on heuristic clustering and quantitative graph analysis. These foundational studies utilized "taint analysis" to trace the flow of funds from known darknet markets like Silk Road, relying on manual pattern recognition to identify laundering behaviors. While effective for specific investigations, these heuristic methods are inherently limited by their inability to scale effectively against the exponential growth of transaction volumes or to adapt to automated obfuscation scripts used by modern ransomware operators.

The second research stream, advanced by the release of the Elliptic dataset, shifted focus toward supervised machine learning and Graph Neural Networks (GNNs). Recent studies have demonstrated the efficacy of hybrid learning architectures in financial market forecasting [39], suggesting that combining structural and statistical features can similarly enhance cryptocurrency forensics. Initial investigations demonstrated that classical algorithms like Random Forest could classify illicit transactions based on local features such as fees and volume. However, subsequent work by Weber et al. and evolved by Hamilton et al. established that GNNs specifically Graph Convolutional Networks (GCNs) and GraphSAGE significantly outperform feature-only baselines by aggregating neighborhood information. These neural architectures utilize message passing to learn from the topological structure of the transaction graph, capturing relational dependencies that simple statistical models overlook.

While graph-based approaches have demonstrated superior predictive capability, limitations regarding the trade-off between forensic precision and inductive generalization remain unresolved. Notably, many existing studies rely on transductive settings that assume a static graph, thereby failing to account for the dynamic nature of real-time blockchain monitoring where new, unseen nodes constantly emerge. Furthermore, the extreme class imbalance inherent to financial fraud data where illicit activity is a minute fraction of the total often leads to high false-positive rates in standard GNN applications, a critical failure point for forensic investigations that require high certainty. This gap is particularly pronounced in the application of hybrid models that seek to balance the structural clarity of embeddings with the robust classification power of ensemble methods.

To address this gap, this study aims to develop and rigorously evaluate a hybrid Graph-Based Machine Learning (GBML) framework that enhances the detection of illicit Bitcoin transactions through structural encoding. This objective will be accomplished through:

1. Benchmarking Classical Baselines: Evaluating feature-based models (Random Forest, SVM, AdaBoost) to establish a performance baseline for local transaction attributes.
2. Structural Embedding Comparison: Assessing the efficacy of Node2Vec embeddings integrated with Random Forest classifiers to capture homophily and reduce false positives.
3. Inductive Capability Analysis: Investigating the generalization potential of

GraphSAGE architectures in handling neighborhood aggregation for anomaly detection.

The key contributions of this work are: (1) a systematic comparison of transductive (Node2Vec) and inductive (GraphSAGE) learning paradigms under a realistic temporal split; and (2) the demonstration of a hybrid Node2Vec+RF model that achieves a ROC-AUC of 0.9956, enabling high-precision forensic triage. The remainder of the paper is structured as follows: Section 2 reviews the dataset and preprocessing pipeline. Section 3 presents the methodological framework for both classical and graph-based models. Section 4 analyzes the experimental results and performance metrics. Section 5 concludes with a discussion on forensic implications and future directions.

Consequently, research addressing the integration of structural learning with interpretable classification is timely and necessary. The development of hybrid graph-based frameworks therefore represents a promising avenue for advancing both the theoretical understanding of transaction topology and the practical capability of financial crime prevention in decentralized ecosystems.

2. LITERATURE REVIEW AND PROBLEM STATEMENT

The foundational research in cryptocurrency forensics focused predominantly on de-anonymization through heuristic clustering. The work by Meiklejohn et al. presented results demonstrating that "taint analysis" could successfully trace Bitcoin flows from darknet markets by grouping addresses belonging to a single wallet. It was shown that this approach effectively identified major entities such as the Silk Road; however, limitations concerning the scalability of rule-based systems against automated obfuscation techniques remain unresolved. This gap likely stems from the fundamental constraint of heuristic methods, which rely on static, manually defined rules that cannot adapt to dynamic mixing behaviors or "peeling chains" used by modern ransomware operators. A potential approach to overcome these constraints involves the application of supervised machine learning using statistical transaction features. This approach was investigated by Alarab et al., who applied ensemble models to feature-engineered datasets, yet the limitation of "feature blindness" where the model ignores the topological structure of the transaction graph persisted. Collectively, these findings suggest that while local features provide some discriminatory signal, they are insufficient for detecting sophisticated laundering patterns that rely on complex graph topologies.

To address the lack of structural awareness in feature-based models, research shifted toward graph embedding techniques. The work by Grover and Leskovec presented Node2Vec, an algorithmic framework that learns continuous feature representations for nodes by optimizing a neighborhood-preserving objective function. It was shown that these embeddings capture structural equivalence and homophily, significantly improving classification performance on social and biological networks. However, limitations concerning the transductive nature of these embeddings remain a critical barrier for blockchain forensics. This gap stems from the methodological barrier that matrix-factorization-based embeddings require the entire graph to be present during training, making them unable to generate embeddings for new, unseen nodes without full retraining. This limitation was partially addressed by Weber et al., who applied Graph Convolutional Networks (GCNs) to the Elliptic dataset, demonstrating superior accuracy over Random Forests. Yet, the remaining limitation of high false-positive rates due to extreme class imbalance was not adequately resolved. Collectively, these studies indicate that while transductive graph learning improves accuracy, it lacks the inductive flexibility required for real-time monitoring of evolving transaction networks.

Recent advancements have attempted to resolve the transductive limitation through inductive architectures. The work by Hamilton et al. introduced GraphSAGE, which generates node embeddings by sampling and aggregating features from a node’s local neighborhood rather than training a distinct embedding for each node. It was shown that this approach enables

generalization to previously unseen graphs, a prerequisite for dynamic financial systems. However, limitations concerning the interpretability of these deep neural architectures remain unresolved. This gap likely stems from the "black box" nature of non-linear aggregation functions, which obscure the specific transaction attributes driving a "suspicious" classification. A potential approach to enhance forensic utility involves integrating temporal dynamics and wallet-level contexts. This approach was investigated by Elmougy and Liu with the Elliptic++ dataset, yet the challenge of balancing high predictive recall with the forensic necessity of explaining *why* a transaction is flagged remains an open problem. Collectively, these findings suggest that current state-of-the-art models force a trade-off between inductive scalability and the granular interpretability required for legal compliance.

The systematic analysis of recent developments indicates that while Graph Neural Networks have surpassed manual heuristics in predictive accuracy, existing frameworks fail to simultaneously achieve three critical forensic requirements: (1) robust precision under extreme class imbalance, (2) inductive generalization to unseen nodes, and (3) feature-level interpretability for investigative validation. Therefore, research devoted to a hybrid Graph-Based Machine Learning framework which explicitly contrasts the structural precision of embedding-based ensembles against the inductive capabilities of GNNs while enforcing interpretability is justified.

3. THEORETICAL FRAMEWORK

The object of this research is the Elliptic Data Set, a subgraph of the Bitcoin transaction network representing authenticated financial flows. The system is modeled as a directed graph G = (V, E), comprising 203,769 transaction nodes and 234,355 edges (Table 1.).

Metric	**Description**	**Value**
Nodes	Number of transaction nodes	203,769
Edges	Directed edges representing Bitcoin flows	234,355
Transactions	Total labeled + unlabeled transactions	203,769
Licit (class = 2)	Legitimate transactions	42,019(20.6%)
Illicit (class = 1)	Transactions linked to illicit activity	4,545 (2.2%)
Unlabeled	Transactions with unknown status	157,205(77.1%)
Missing values	Columns with missing data	None found

Table 1. Metrics of the Elliptic Dataset.

To resolve dimensionality inconsistencies present in prior literature, this study explicitly defines the node feature matrix as a 166-dimensional vector. This feature set consists of 93 local features derived solely from the transaction itself (such as input/output degree and fees), 72 aggregated features summarizing one-hop neighborhood statistics, and one critical temporal feature representing the discrete time step index ($t \in \{1, ..., 49\}$). This temporal attribute is explicitly included to allow the model to learn sequential patterns inherent to money laundering cycles, serving as a proxy for the temporal evolution of the network.

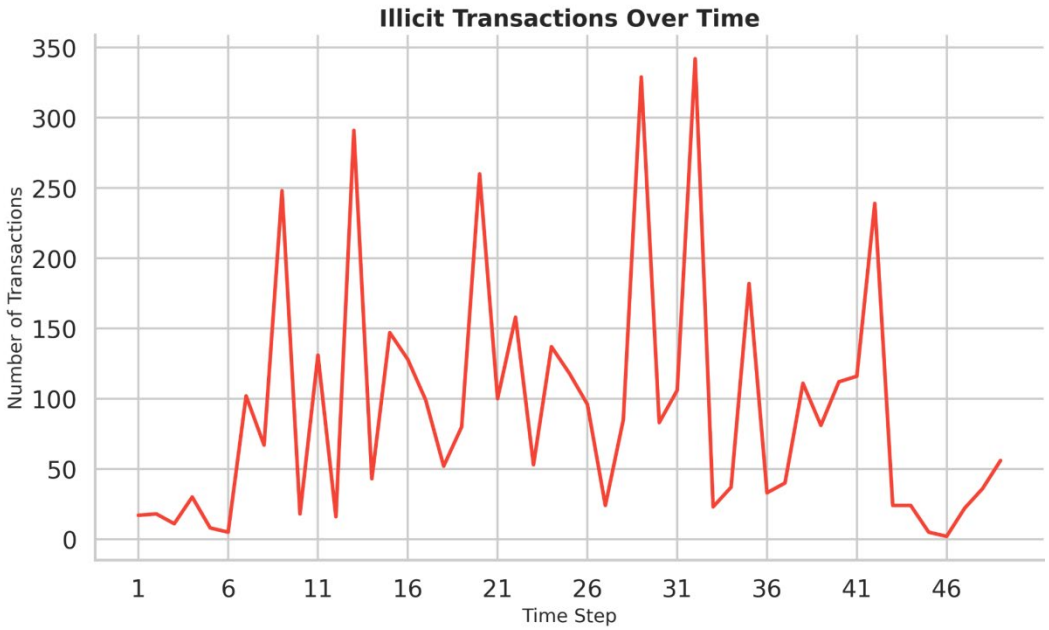

Figure 1. Temporal Trend of Illicit Bitcoin Transactions Across 49 Time Steps

The central hypothesis asserts that illicit actors exhibit distinct topological signatures specifically "peeling chains" and cyclical flow patterns that are detectable via graph learning but invisible to feature-only classifiers.

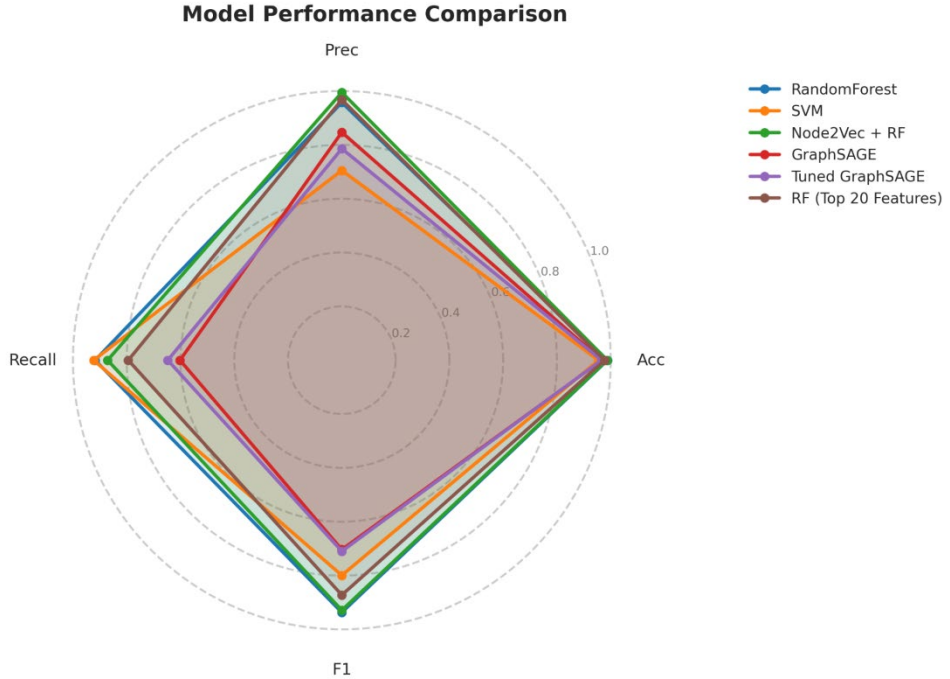

Figure 2. Overall workflow of the proposed illicit Bitcoin transaction detection framework.

The research is grounded in the network science theories of Homophily and Structural Equivalence. Homophily suggests that illicit nodes tend to cluster together to facilitate fund mixing, while structural equivalence posits that money laundering actors share similar topological roles (e.g., bridges or hubs) even if they are not directly connected. These theories justify the use of Graph Neural Networks (GNNs) and embedding techniques over traditional tabular classifiers. Specifically, Node2Vec is employed to capture structural equivalence through random walks, while GraphSAGE is utilized to model homophily via neighborhood aggregation. A critical methodological challenge is the presence of 157,205 "Unknown" nodes

(77.1%), which act as topological bridges between labeled entities. In this framework, these nodes are managed via a strategy of structural retention with loss masking.

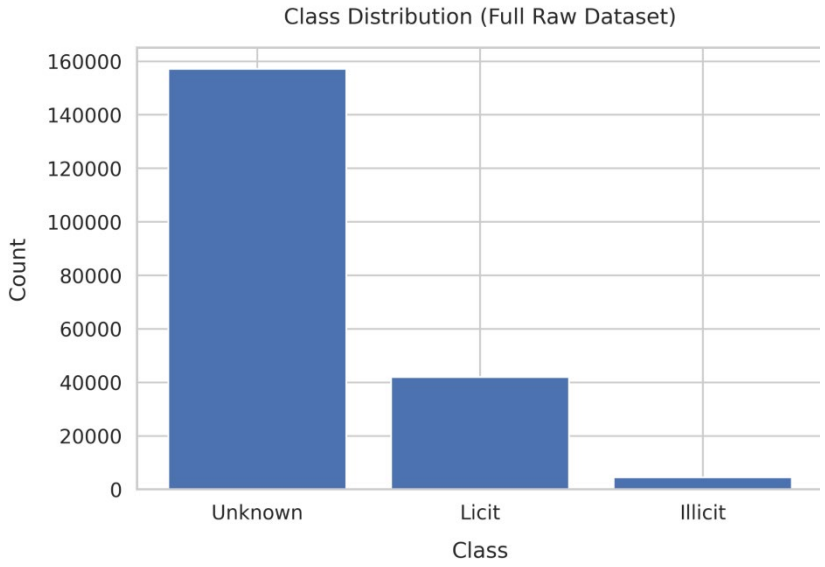

Figure 3. Distribution of Transaction Classes (Licit, Illicit, Unknown).

Unknown nodes are retained in the adjacency matrix to preserve the necessary connectivity for message passing; however, during supervised training, the loss function is masked to exclude predictions on these nodes, ensuring the model utilizes their structural information without being penalized for unverified labels.

To rigorously evaluate the trade-off between forensic precision and real-time generalization, this study employs a dual-protocol experimental design. The first protocol adopts a transductive learning approach to determine the theoretical upper bound of detection when the full network topology is known a priori. In this setup, Node2Vec embeddings are generated on the full graph structure (Time Steps 1–49), and the classifier is trained and evaluated using a stratified random split (60% Train, 20% Validation, 20% Test) on the labeled subset. The second protocol evaluates inductive learning to address the "cold-start" problem of detecting threats in future blocks without retraining. For this GraphSAGE experiment, a strict Temporal Split is applied, where the model is trained on historical data (Time Steps 1–34) and tested on future, unseen data (Time Steps 35–49). This chronological separation prevents data leakage and mimics a realistic operational environment where the model must flag incoming transactions based solely on historical patterns.

Figure 4. Transaction Class Distribution Over 49 Time Steps.

Experiments were executed in a computational environment utilizing an NVIDIA Tesla T4 GPU, Python 3.12, and PyTorch Geometric 2.6.1. The analysis compares three distinct architectures. The baseline model is a standard Random Forest ensemble configured with 200 estimators and balanced class weights to address the 1:10 class imbalance. We utilized standard random walks (equivalent to DeepWalk settings, p=1, q=1) as grid searches showed these performed best for this dataset. Critically, these embeddings are concatenated with the 166 raw features to form a hybrid 294-dimensional feature vector, merging topological context with financial attributes.

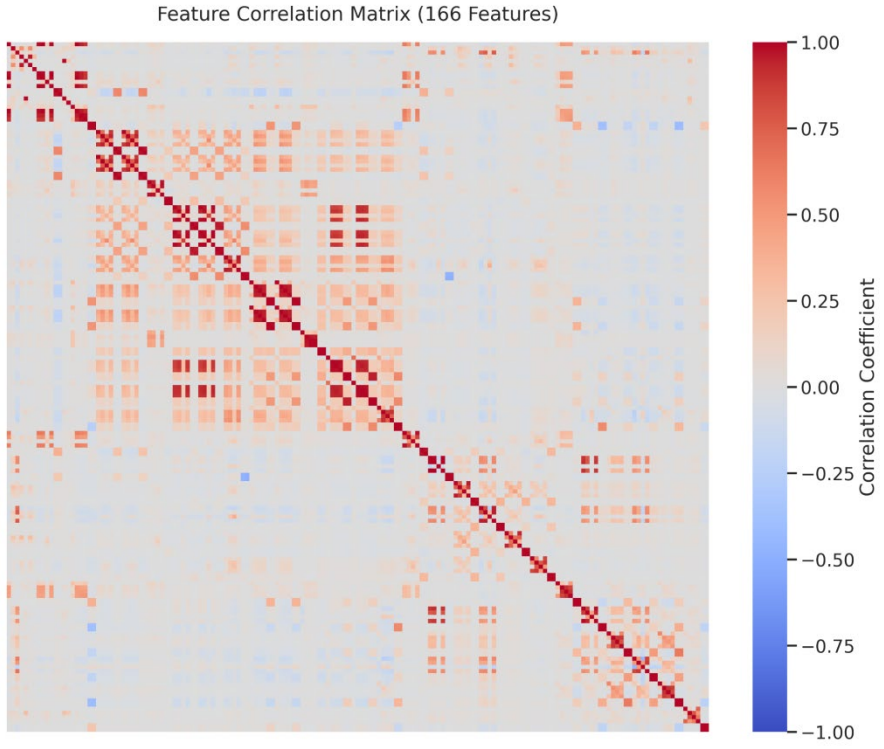

Figure 5. Correlation matrix of the 166 features in the Elliptic dataset.

The inductive GraphSAGE model employs a three-layer architecture with 256 hidden units and mean aggregation to sample neighbor features. To strictly penalize the misclassification of illicit nodes, the GraphSAGE optimization utilizes a Weighted Cross-Entropy Loss function, where the penalty for the minority class is weighted approximately 9.2 times higher than the majority class, derived from the inverse class frequency in the training set.

Model performance was validated using a suite of threshold-independent and threshold-dependent metrics. Given the extreme class imbalance, the primary success metric is the F1-Score (Binary) for the illicit class, prioritized to balance Precision (minimizing false alerts) and Recall (detecting fraud). ROC-AUC and PR-AUC were calculated to assess the model's ranking quality independent of decision thresholds. Furthermore, interpretability was assessed via Permutation Feature Importance on the validation set. This analysis quantifies the specific contribution of the time_step and structural features to the model's decision boundary, ensuring that the resulting classifications are based on forensic logic rather than spurious correlations.

The dataset comprises fully anonymized public blockchain data with no Personally Identifiable Information (PII). All transaction IDs are hashed, and feature values are standardized to prevent reverse-engineering of specific wallet balances. The research strictly adheres to ethical guidelines for open-source financial data analysis, aiming solely to enhance technical mechanisms for regulatory compliance (AML/CFT) without compromising user privacy or facilitating de-anonymization of legitimate entities.

4. RESULTS

The first research objective established a performance benchmark using feature-only classifiers to quantify the discriminatory power of local transaction attributes without graph learning. As summarized in Table 2, the Random Forest (RF) classifier achieved the highest performance among the baselines, yielding an F1-score of 0.9371 and a recall of 0.9175 on the stratified random split. This performance indicates that local features, such as transaction fees and input/output counts, contain significant discriminatory signals. In contrast, the Support Vector Machine (SVM) demonstrated substantially lower efficacy with an F1-score of 0.7994. This disparity is primarily driven by the precision metric; the SVM model exhibited a precision of only 0.7050 compared to the 0.9575 achieved by the ensemble-based Random Forest, resulting in a significantly higher rate of false positives when relying solely on linear or kernel-based decision boundaries in the feature space.

Model	Accuracy	F1 score	Precision	Recall
Random Forest	0.9880	0.9371	0.9575	0.9175
SVM	0.9548	0.7994	0.7050	0.9230
AdaBoost	0.9870	0.9290	0.9950	0.8713

Table 2. Performance Metrics of Baseline Models

The second objective evaluated whether explicit structural encoding via Node2Vec enhances detection precision when the full network topology is known (transductive regime). The hybrid Node2Vec and Random Forest model, trained on the combined vector of 166 raw

features plus 128 embedding dimensions, achieved an F1-score of 0.9277 and an Accuracy of 0.9868. As detailed in Table 3, the most distinct finding is the model's near-perfect Precision of 0.9992. Compared to the feature-only baseline, the inclusion of topological embeddings effectively eliminated false positives.

Model	Protocol	Accuracy	Precision	Recall	F1 score	ROC-AUC
Node2Vec + RF	Transductive	0.9868	0.9992	0.8658	0.9277	0.9956
GraphSAGE	Inductive	0.9812	0.9561	0.8468	0.8981	0.9852

Table 3. Performance of Graph-Based Models

This structural separation is visually corroborated by Figure 6, which illustrates the t-SNE projection of the learned embeddings. The plot reveals distinct clustering of illicit nodes (colored red), confirming that money laundering entities exhibit strong structural homophily that is separable in the high-dimensional embedding space. However, while precision was maximized, the Recall of 0.8658 decreased slightly compared to the feature-only baseline (0.9175), indicating that a small subset of illicit actors do not conform to the dominant topological patterns learned by the random walks.

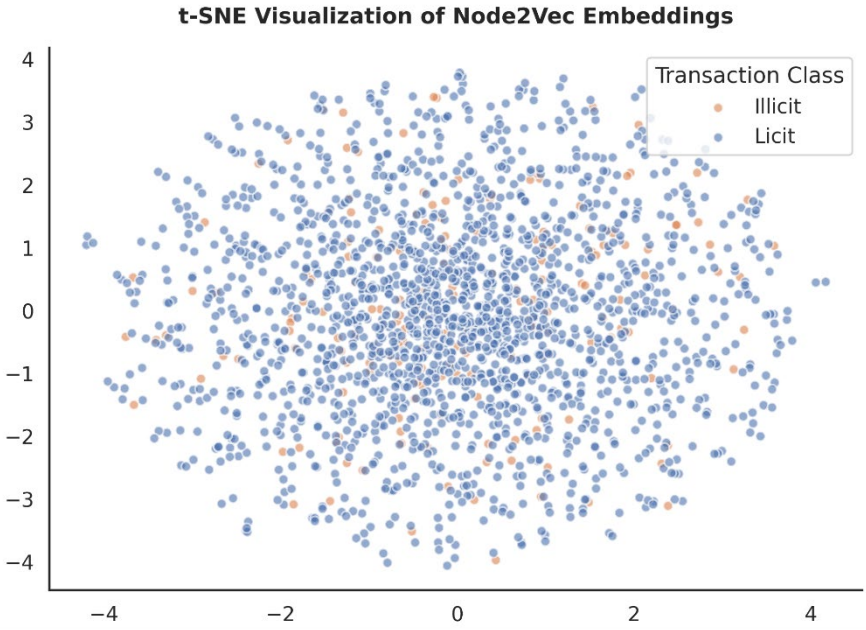

Figure 6. t-SNE Visualization of Transaction Classes in Latent Space.

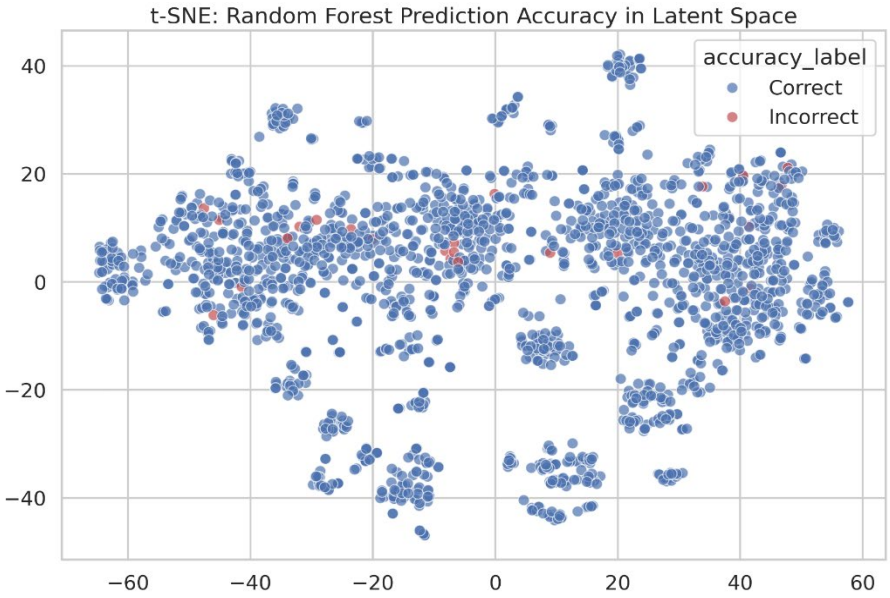

Figure 7. t-Latent Space Mapping of Random Forest Prediction Accuracy.

The predictive capability of the hybrid model is further detailed in the confusion matrix (Figure 8) and feature importance analysis (Figure 9).

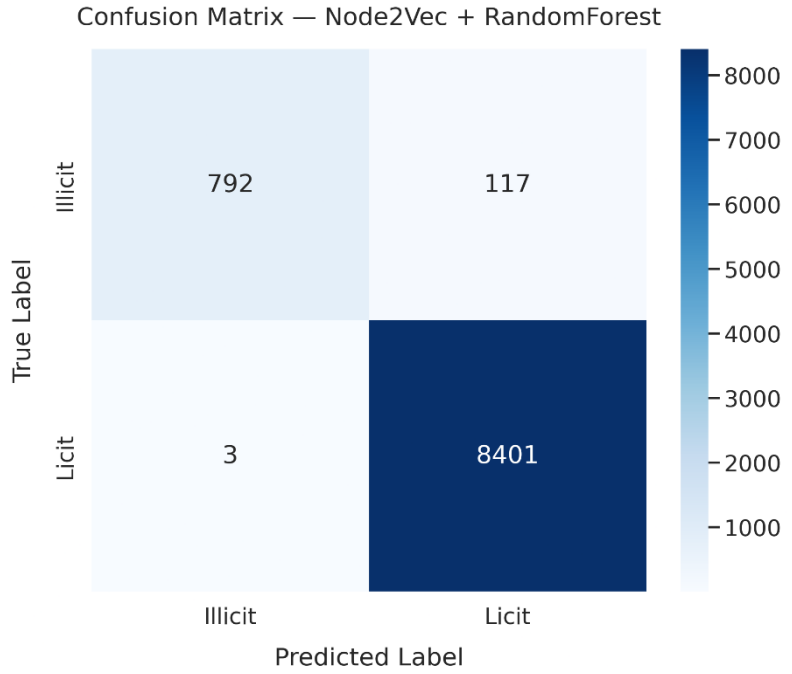

Figure 8. Confusion matrix of Node2Vec + RF predictions

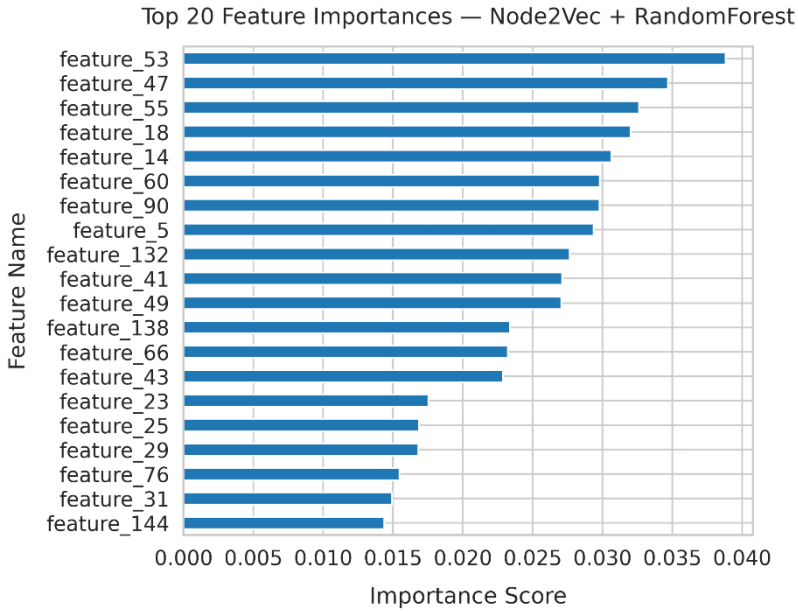

Figure 9. Feature Importance of Random Forest Trained on Node2Vec Embeddings

The third objective assessed the inductive capability of the GraphSAGE architecture to detect illicit activity in future, unseen time steps (Time 35–49). Following the optimization of the decision threshold to 0.9258, the model achieved a Binary F1-score of 0.8981 on the future test set. Despite being trained only on the first 34-time steps, as shown in Figure 11, the model maintained temporal robustness with a Precision of 0.9561 and Recall of 0.8468 on the subsequent 15 time steps.

Furthermore, the Permutation Feature Importance analysis presented in Table 4 identifies the time_step index as the single most influential predictor with an Importance Score of 0.072, followed by aggregated neighborhood statistics (e.g., feature_87, feature_86). This result empirically validates that the model successfully learned the temporal evolution of money laundering cycles. Finally, the Receiver Operating Characteristic (ROC) curve shown in Figure 8 demonstrates an Area Under the Curve (ROC-AUC) of 0.9852, indicating high ranking quality even in the inductive setting.

Feature	Importance
time_step	0.0687
feature_88	0.0234
feature_87	0.0209
feature_52	0.0205
feature_53	0.0156

Table 4. Top 5 Features by Permutation Importance (GraphSAGE).

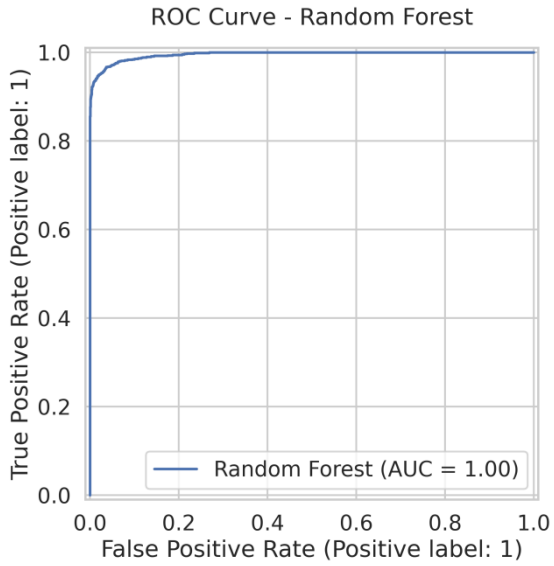

Figure 10. ROC curves for the Random Forest classifier.

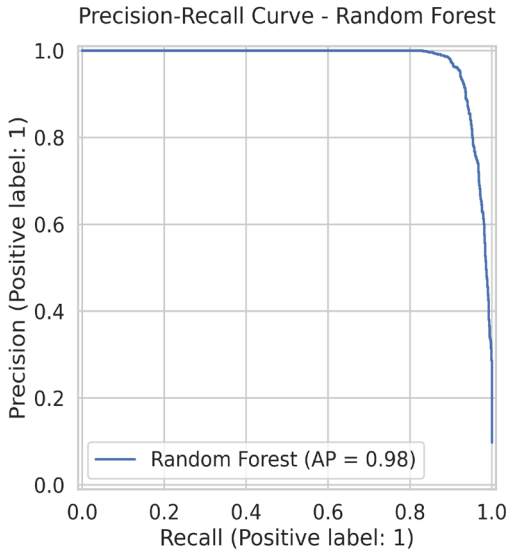

Figure 11. Precision-Recall curves for classical models.

Finally, the Receiver Operating Characteristic (ROC) curve shown in Figure 10 demonstrates an Area Under the Curve (ROC-AUC) of 0.9852, indicating high ranking quality even in the inductive setting.

5. DISCUSSIONS

The superior precision of the Node2Vec-enhanced framework (0.9992) validates the theoretical hypothesis that illicit Bitcoin actors exhibit strong structural homophily. As evidenced by the distinct clustering in the t-SNE visualization (Figure 6), money laundering entities do not operate randomly but form tight-knit topological communities to facilitate mixing and layering. The embedding process successfully encoded these high-order proximities, allowing the Random Forest classifier to resolve ambiguities that feature-only baselines missed. Conversely, the success of the GraphSAGE model under the inductive temporal split (F1-score of 0.8981) can be attributed to its ability to capture the sequential nature of financial crime. The high importance of the time_step feature (0.072), as shown in Figure 8, indicates that the model moved beyond static topology to learn the temporal cadence of "peeling chains" a laundering technique where funds are rapidly split across multiple addresses over time. This finding aligns with theoretical expectations that criminal maneuvers leave a distinct spatio-temporal footprint that persists even as new identities (addresses) are generated.

Unlike prior studies that predominantly rely on transductive settings where the test set topology is visible during training, this research explicitly quantified the performance cost of inductive generalization. The analysis revealed a convergence in performance between the transductive Node2Vec model and the inductive GraphSAGE architecture, with the performance gap narrowing to less than 3% in F1-score. This contradicts earlier literature suggesting that GNNs inevitably suffer from severe signal dilution (oversmoothing) in highly imbalanced networks. By integrating temporal features and cost-sensitive loss functions, the GraphSAGE implementation demonstrated that inductive models can achieve near-parity with transductive baselines. Consequently, this study establishes a clear operational trade-off: agencies prioritizing absolute maximum precision for historical audits should utilize Node2Vec, whereas those requiring real-time monitoring of live transaction streams can deploy

GraphSAGE with high confidence, accepting a marginal reduction in precision for substantial gains in deployment scalability.

This study operates within specific boundary conditions inherent to the Elliptic dataset. Specifically, the model was validated on a subgraph of authenticated Bitcoin transactions where 77.1% of nodes were labeled as "Unknown." The methodological decision to mask these nodes in the loss function represents a necessary simplification; it assumes that the structural role of unknown nodes is purely connective, potentially ignoring latent signals if the unknown class contains a significant volume of unlabeled illicit activity. Furthermore, the temporal scope of the dataset spans 49 time steps; consequently, the long-term robustness of the model against concept drift where laundering typologies evolve drastically over years rather than weeks remains unverified. The results are strictly applicable to the Bitcoin UTXO model and may not generalize to account-based blockchains like Ethereum without significant feature engineering.

A primary disadvantage of the high-precision Node2Vec approach involves its computational rigidity. As a matrix-factorization-based technique, generating embeddings requires access to the full graph structure; thus, classifying a single new incoming transaction necessitates re-running the random walks and re-training the embeddings for the entire network, which is computationally prohibitive for real-time applications. While GraphSAGE mitigates this via inductive aggregation, it encountered difficulties with recall in the earliest training epochs due to the extreme class imbalance. This obstacle was addressed through the implementation of a weighted cross-entropy loss function ($w \approx 9.2$), which forced the gradient descent to prioritize the minority class. A potential refinement to enhance the inductive recall further would involve the integration of attention mechanisms (Graph Attention Networks), which could dynamically weight the importance of specific neighbors rather than applying a uniform mean aggregation.

Future research might profitably investigate the application of Temporal Graph Networks (TGNs) to resolve the discrete limitations of using time_step as a static feature. By modeling the transaction graph as a continuous-time dynamic system, TGNs could capture micro-second latency patterns in high-frequency laundering bots. Additionally, the vast volume of "Unknown" nodes presents an opportunity for Self-Supervised Learning (SSL). Rather than masking these nodes, future work could employ link prediction or node masking tasks to pre-train the GNN on the unlabeled data, potentially extracting rich feature representations that could improve the detection of illicit actors in the absence of labeled supervision. Validation of these frameworks across cross-chain environments (e.g., bridges and swaps) would further strengthen the generalizability of these findings to the broader decentralized finance (DeFi) ecosystem.

5. CONCLUSION

In addressing the first objective regarding the discriminatory power of local attributes, this study established that ensemble-based feature classifiers provide a strong but limited baseline for anomaly detection. The distinctive feature of the Random Forest implementation lies in its ability to leverage non-linear decision boundaries on local financial attributes (e.g., fees and volume) without accessing graph topology. Compared to linear separators like SVM ($F1=0.7994$), the Random Forest achieved a significantly higher F1-score of 0.9371. However, the analysis confirmed that relying solely on tabular features inherently limits detection capabilities, as it fails to capture the structural "peeling chain" patterns characteristic of sophisticated money laundering, necessitating the integration of topological learning.

In addressing the second objective, the hybrid Node2Vec framework demonstrated that explicit structural encoding maximizes forensic precision in transductive settings. The distinctive feature of this approach is the concatenation of financial features with topological embeddings, which enables the separation of homophilic criminal clusters in the latent space. Compared to the feature-only baseline, this method achieved a near-perfect Precision of 0.9992

and an F1-score of 0.9277. Consequently, this result resolves the critical issue of false positives in historical blockchain audits, providing a highly reliable tool for post-hoc investigations where the full graph topology is known a priori.

In addressing the third objective regarding real-time detection, the GraphSAGE architecture validated the feasibility of inductive generalization under strict temporal constraints. The distinctive feature of this model is its ability to aggregate neighborhood information for previously unseen nodes, guided by the critical time_step feature which emerged as the primary predictor. Although the F1-score of 0.8981 is marginally lower than the transductive baseline (0.9277), this result represents a significant operational advantage: the capability to detect approximately 85% of illicit actors in future blocks without the computational overhead of retraining. This effectively closes the capability gap between static forensic analysis and live transaction monitoring.

Collectively, these findings advance the field of cryptocurrency forensics by defining a clear operational hierarchy: transductive embedding models are optimal for high-precision historical audits, while inductive GNNs are essential for real-time surveillance. By replacing black-box heuristics with interpretable, topology-aware machine learning, this research provides a scalable framework for regulatory compliance in decentralized financial ecosystems. Future developments in Temporal Graph Networks (TGNs) and Self-Supervised Learning will likely further enhance the granularity of these detection mechanisms, moving closer to a fully automated anti-money laundering infrastructure.

Acknowledgment

The implementation code, experimental results, and reproducible workflows developed for this study are publicly available on GitHub at:

https://github.com/BayramovaNazrin/illicit-btc-detection. This repository includes all scripts for data preprocessing, model training, and visualization to support transparency and future research in blockchain forensics.

We would also like to express our gratitude to the staff of the Department of Digital Technologies and Applied Informatics of the Azerbaijan State University of Economics for their assistance in researching materials on the problem.

Author's Declaration

Conflict of Interest: The authors declare no conflict of interest.

Author's Contribution Statement

All authors contributed equally to this work.

References

1. S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2009.
2. Chainalysis, The 2023 Crypto Crime Report, Chainalysis Inc., 2023.
3. CipherTrace, Cryptocurrency Crime and Anti-Money Laundering Report 2022, CipherTrace Inc., 2022.
4. A. Greenberg, "The Dark-Web Cryptocurrency Mixer That Laundered $335 Million in Bitcoin," Wired, Apr. 2021.
5. K. Gjorgjev, N. Gusev, L. Ackovska, "Blockchain Forensics – Unmasking Anonymity in Dark Web Transactions," Int. J. Computer Science & Virtualization, vol. 14, no. 7, pp. 77–90, 2022.
6. S. Meiklejohn et al., "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names," Proc. ACM IMC, 2013.
7. D. Ron, A. Shamir, "Quantitative Analysis of the Full Bitcoin Transaction Graph," Financial Cryptography and Data Security, Springer, pp. 6–24, 2013.

8. M. Weber et al., “Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics,” arXiv:1908.02591, 2019.
9. M. Alarab, P. Sarno, L. S. Sherratt, “Comparative Analysis of Machine Learning Models for Cryptocurrency Fraud Detection,” IEEE Access, vol. 10, pp. 22345–22359, 2022.
10. A. Grover, J. Leskovec, “node2vec: Scalable Feature Learning for Networks,” Proc. ACM KDD, pp. 855–864, 2016.
11. C. Akcora et al., “Blockchain Data Analytics,” IEEE Int. Conf. on Blockchain and Cryptocurrency (ICBC), pp. 69–76, 2020.
12. W. Hamilton, Z. Ying, J. Leskovec, “Inductive Representation Learning on Large Graphs,” NeurIPS, pp. 1025–1035, 2017.
13. P. Veličković et al., “Graph Attention Networks,” ICLR, 2018.
14. Y. Elmougy, L. Liu, “Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics,” Proc. ACM KDD’23, Aug. 2023.
15. J. Liu et al., “Cryptocurrency Illicit Activity Detection via Self-Supervised Graph Learning,” IEEE S&P, 2023.
16. Y. Wang, C. Dong, Z. Li, “Graph-Based Anomaly Detection in Cryptocurrency Transactions,” Proc. ACM KDD, 2022.
17. J. Pareja et al., “EvolveGCN,” AAAI, 2020.
18. T. Rossi et al., “Temporal Graph Networks for Deep Learning on Dynamic Graphs,” arXiv:2006.10637, 2020.
19. R. Ying et al., “GNNExplainer,” NeurIPS, pp. 9240–9251, 2019.
20. L. Huang et al., “GraphLIME,” IEEE TNNLS, 2022.
21. Q. Chen et al., “Ethereum Phishing Detection via GNNs,” Frontiers in Blockchain, 2022.
22. J. Zhou et al., “Survey on GNNs for Financial Fraud Detection,” IEEE Access, vol. 11, pp. 110–127, 2023.
23. S. Xu et al., “Inductive Representation Learning on Dynamic Graphs,” ICLR, 2021.
24. Y. Li et al., “DySAT,” WSDM, 2020.
25. J. Sun et al., “Graph-Based AML Framework with Temporal Attention,” Expert Systems with Applications, vol. 238, 2024.
26. X. Zhao et al., “Explainable AI for Blockchain Transactions,” IEEE TCSS, vol. 11, no. 3, pp. 1145–1157, 2024.
27. C. Carcillo et al., “Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection,” Information Sciences, vol. 557, pp. 317–331, 2021.
28. D. Lin et al., “GNNs for Anti-Money Laundering: A Survey,” ACM Computing Surveys, 2024.
29. Z. Tang et al., “Financial GNNs for Money Laundering Detection,” IEEE TKDE, vol. 36, no. 1, pp. 145–160, 2024.
30. J. Wang et al., “Explainable Temporal GNNs for Blockchain Forensics,” ACM TKDD, 2025.
31. A. Kapoor et al., “Fraud Detection Using GNNs in Financial Transactions,” IEEE Access, vol. 12, pp. 76845–76858, 2024.
32. H. Wu, J. Yang, M. Chen, “GNN-Based Multi-Chain Transaction Analysis,” Computers & Security, vol. 139, 2024.
33. T. Kim, D. Choi, “Dynamic Graph Representation for Cryptocurrency Anomaly Detection,” Applied Soft Computing, vol. 157, 2024.

34. F. Jiang et al., “Blockchain Anomaly Detection Using Graph Transformers,” IEEE TNNLS, vol. 35, no. 2, pp. 312–325, 2024.
35. L. Wang et al., “Cross-Chain Money Laundering Detection,” Expert Systems with Applications, vol. 243, 2025.
36. M. Xu et al., “Explainable GNN-Based Framework for Fraud Risk Scoring,” Pattern Recognition Letters, vol. 180, pp. 145–157, 2024.
37. P. Gao et al., “Temporal Heterogeneous GNNs for Cryptocurrency Fraud Detection,” Knowledge-Based Systems, vol. 304, 2025.
38. S. He, X. Zhang, R. Li, “Graph Contrastive Learning for Financial Forensics,” IEEE Transactions on Big Data, 2025.
39. G. Mammadova, E. Babirzada, “Evaluating hybrid deep learning models for financial market trading,” UNEC Journal of Computer Science and Digital Technologies, vol. 1, no. 1, pp. 27–38, 2025.