Journal of Economics, Finance and Accounting Studies
ISSN: 2709-0809
DOI: 10.32996/jefas
Journal Homepage: www.al-kindipublisher.com/index.php/jefas

JEFAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Machine Learning Models for Detecting Hidden Collusion Networks in U.S. Corporate Finance

Atika Dola[1], Sakera Begum[2], Umama Khanom Antara[3], MD Rahimul Islam[4], Tasmia Sultana[5] and Nagma Zabin[6]

[1]Bachelor's in Business Administration – Finance, Idaho State University
[2]Master of Science in Information Technology, Washington University of Science and Technology.
[3]Master's in Business Analytics, University of North Texas
[4]Master's in Merchandising and consumer Analytics, University of North Texas
[5]Master's in Merchandising and Consumer Analytics, University of North Texas
[6]Master's in Development Studies, Bangladesh University of Professionals
Corresponding Author: Atika Dola**, Email:** atikadola25@gmail.com

| ABSTRACT

Hidden collusion networks in U.S. corporate finance present a significant challenge to market integrity and regulatory oversight. Such networks are difficult to detect due to the indirect and distributed nature of collusive behavior, which is often embedded within legitimate financial transactions and governance relationships. This study aims to develop and evaluate machine learning models capable of detecting latent collusion structures by combining firm-level financial indicators with relational and governance-based network information. The goal is to uncover clusters of firms that exhibit coordinated behavior while providing interpretable insights for regulatory decision-making. A multi-stage detection framework was implemented, incorporating classical machine learning classifiers, tree-based ensembles, graph neural networks (GNNs), and hybrid models combining feature-driven and structure-driven learning. Corporate entities were represented as nodes within financial and governance networks, with edges encoding ownership ties, shared executives, transactional dependencies, and temporal co-movements. Models were evaluated using precision, recall, F1-score, AUC-ROC, AUC-PR, and network consistency metrics, while robustness analyses examined performance under class imbalance and sparse labeling conditions. Graph-based models outperformed traditional baselines, achieving F1-scores up to 0.91 and AUC-ROC values up to 0.94. Hybrid ensembles that combined tree-based and graph-based predictions achieved the highest overall performance (F1-score = 0.93, AUC-ROC = 0.95, AUC-PR = 0.91). The models successfully identified densely connected collusive clusters and key hub firms, highlighting the importance of relational and temporal features over isolated firm-level indicators. Ablation studies confirmed that financial metrics alone were insufficient to detect coordinated behavior without network structure. The study demonstrates that machine learning, particularly when integrated with graph-based relational representations, provides an effective and scalable approach for detecting hidden collusion networks in corporate finance. The proposed framework offers practical value for regulators, enabling probabilistic risk assessment, early-warning detection, and data-driven surveillance of potentially collusive activity.

| KEYWORDS

Machine Learning; Collusion Detection; Corporate Finance; Graph Neural Networks; Financial Networks; Regulatory Analytics

| ARTICLE INFORMATION

ACCEPTED: 01 February 2024 **PUBLISHED:** 13 February 2024 **DOI:** 10.32996/jefas.2024.6.1.14

1. Introduction

1.1 Background and Motivation

Financial markets in the United States are increasingly susceptible to sophisticated forms of collusion, where multiple corporate entities coordinate their actions to manipulate outcomes such as pricing, mergers, or asset valuations. Detecting such hidden collusion is inherently challenging because collusive behaviors are often subtle, distributed, and obscured within legitimate financial and governance activities. Traditional fraud detection mechanisms, which rely on rule-based or statistical frameworks, are often insufficient to capture these latent structures. Bolton and Hand (2002) provide a historical overview of statistical fraud detection, highlighting that conventional methods largely depend on predefined rules or thresholds, making them ill-suited to uncover complex, multi-entity collusion schemes [5]. Similarly, Kou et al. (2004) survey traditional fraud detection techniques and emphasize that rule-based systems and simple statistical models, while effective for obvious anomalies, often fail to detect relational or networked fraudulent behaviors because they treat each entity independently without considering interconnections [9].

Recent advances have shifted attention toward graph-based anomaly detection, which leverages the structural properties of networks to identify unusual patterns of interaction among entities. Akoglu et al. (2015) define graph-based anomaly detection as the process of identifying nodes, edges, or subgraphs in a network whose connectivity patterns deviate significantly from expected behavior, often revealing hidden clusters or coordinated groups that standard statistical methods overlook [1]. In corporate finance, these networks can represent board interlocks, cross-ownership relationships, shared executive participation, or correlated transactional behaviors, all of which may indicate potential collusion when analyzed systematically. The graph-based approach is particularly promising because it captures not only individual entity behaviors but also the relational context in which these behaviors occur, enabling the detection of structural anomalies that are otherwise invisible to entity-level analyses.

Despite these methodological advancements, the broader financial fraud landscape remains complex. West and Bhattacharya (2016) provide a comprehensive review of intelligent financial fraud detection, noting that while machine learning and data-driven approaches have enhanced detection capabilities, limitations persist, particularly in terms of adaptability to evolving fraud patterns and interpretability of results [20]. Ngai et al. (2011) further underscore that many data mining and machine learning techniques applied in financial fraud detection are constrained by issues such as class imbalance, feature sparsity, and the lack of ground-truth labels, which hinders their ability to detect subtle or emerging collusion patterns [10]. Collectively, these insights suggest that detecting hidden collusion networks requires approaches that integrate both entity-level features and relational network information, motivating the use of advanced machine learning techniques capable of modeling complex dependencies and structural anomalies within corporate finance networks.

1.2 Problem Statement and Research Objectives

Hidden collusion in U.S. corporate finance represents a significant and underexplored risk to market integrity, as coordinated activities among firms can distort pricing, influence mergers and acquisitions, and undermine investor confidence. Traditional detection frameworks, including rule-based systems and classical statistical models, are primarily designed to capture overt anomalies, such as unusual transaction volumes or abnormal financial ratios, and are ill-equipped to detect coordinated behaviors across multiple entities simultaneously (Bolton & Hand, 2002; Kou et al., 2004) [5, 9]. These frameworks typically operate at the individual entity level and fail to consider the complex network of relationships that may facilitate collusion. For instance, two firms with shared board members, overlapping investment patterns, and cross-ownership structures might coordinate pricing strategies without triggering conventional alerts, rendering latent collusion virtually invisible to standard monitoring techniques. The insufficiency of traditional methods has prompted the exploration of graph-based approaches, which model firms as nodes and relational ties as edges, allowing for the detection of unusual clusters and structural irregularities indicative of collusive networks (Akoglu et al., 2015) [1].

While machine learning has been increasingly applied to financial fraud detection, challenges persist in adapting these techniques for hidden collusion detection. Ngai et al. (2011) highlight that classical supervised and unsupervised learning methods often struggle with sparse labels, class imbalance, and noisy features, all of which are prevalent in collusion datasets [10]. Furthermore, collusion often manifests in subtle behavioral correlations and relational structures that require more than simple feature-based classification; it demands models capable of integrating temporal patterns, financial metrics, and network connectivity simultaneously. West and Bhattacharya (2016) stress that intelligent detection systems must evolve to account for complex interdependencies and the adaptive nature of fraudulent behaviors, moving beyond static anomaly detection to frameworks that can infer latent coordination (West & Bhattacharya, 2016) [20].

Against this backdrop, the present study seeks to develop a machine learning framework that leverages both network structures and firm-level financial features to detect hidden collusion networks in U.S. corporate finance. The objectives of the study are threefold: first, to construct relational networks representing board interlocks, shared executive participation, and transactional dependencies among firms; second, to design and evaluate machine learning models, including graph-based and hybrid architectures, that can identify structural anomalies indicative of collusion; and third, to assess model performance using precision-focused detection metrics, network consistency analyses, and robustness checks to control false positives. By integrating network insights with advanced learning models, the study aims to move beyond traditional fraud detection methods and provide a scalable framework for early identification of potentially collusive structures. This research contributes to both the methodological advancement of financial anomaly detection and the practical needs of regulators and market oversight bodies, providing a foundation for data-driven, proactive surveillance of complex corporate networks.

1.3 Contributions of the Study

This study introduces a novel framework for detecting hidden collusion networks in U.S. corporate finance that combines machine learning with network-based analyses. Unlike prior research that primarily focuses on either entity-level anomalies or traditional statistical monitoring, the proposed approach integrates relational and financial features into a unified detection model, enabling the identification of subtle, coordinated behaviors among multiple firms. The framework employs both supervised and unsupervised learning techniques, alongside graph neural networks and community detection algorithms, to capture structural anomalies and cluster behaviors suggestive of collusion.

In addition to methodological innovation, the study offers practical insights for financial regulators, corporate auditors, and market surveillance units. By providing a probabilistic assessment of collusion risk and identifying suspicious clusters within corporate networks, the framework supports proactive intervention, risk prioritization, and informed decision-making. The model is designed to be interpretable, highlighting not only high-risk entities but also the relational structures that underpin potential collusion, which enhances transparency and facilitates regulatory action. Finally, the study establishes a foundation for future research in network-informed financial anomaly detection. The integration of graph-based learning with traditional financial features demonstrates that relational information is critical for detecting complex, latent collusive behaviors that would otherwise remain hidden. The approach is scalable and adaptable, enabling its application across diverse corporate environments and regulatory contexts, and it sets the stage for subsequent enhancements, such as temporal modeling of evolving collusion networks and cross-market analyses. By bridging the gap between machine learning, network analysis, and financial oversight, this research contributes both to academic knowledge and to practical tools for safeguarding market integrity.

2. Literature Review

2.1 Traditional Approaches to Financial Collusion Detection

Traditional methods for detecting financial fraud and collusion in corporate environments have historically relied on manual auditing procedures and rule-based monitoring frameworks. Schillermann (2018) emphasizes that early detection of corporate financial fraud often depended on labor-intensive processes such as ledger reconciliation, transaction inspections, and anomaly checks performed by auditors, which were inherently limited in scope and scalability [16]. While these approaches were effective for flagging overt irregularities, they frequently failed to capture subtle, coordinated behaviors that manifest across multiple entities over time. Pourhabibi et al. (2020) survey graph-based anomaly detection approaches and note that traditional systems largely ignore relational and structural information, focusing instead on individual firm metrics or simple threshold-based rules [15]. Such frameworks are reactive rather than proactive, identifying fraud only after significant deviations have occurred, thereby creating a lag in regulatory response.

Historically, statistical fraud detection also faced similar limitations. Rule-based and classical statistical methods, while providing a foundation for early detection systems, struggled with dynamic, evolving fraud behaviors. As highlighted by Watts and Strogatz (1998), financial networks often exhibit small-world characteristics, meaning that local interactions can have disproportionate effects on overall network behavior [19]. Traditional approaches, which do not model these network dependencies, fail to recognize the systemic implications of anomalous activities. The reliance on static thresholds and simple heuristics means that coordinated or latent collusive patterns, which can spread across firms and transactions, remain undetected. This creates a need for approaches that not only assess individual entity behavior but also capture higher-order interactions and network-level deviations.

Furthermore, the evolution of corporate fraud in the digital age has rendered manual and rule-based systems insufficient. With the rise of complex corporate structures, cross-ownerships, interlocking boards, and indirect transactional linkages, collusion often manifests through subtle correlations rather than blatant misstatements. Pourhabibi et al. (2020) argue that without network-

awareness, conventional methods are unable to uncover patterns that arise from the interplay of multiple nodes within a corporate ecosystem [15]. As a result, auditors and compliance officers have increasingly sought automated, data-driven mechanisms capable of interpreting multi-dimensional financial datasets and relational structures. The shortcomings of these traditional approaches highlight the necessity of integrating computational models, such as machine learning and graph-based methods, to detect patterns of collusion that extend beyond individual entity-level anomalies and encompass broader relational dependencies.

2.2 Machine Learning in Financial Fraud and Anomaly Detection

Machine learning has emerged as a transformative tool for financial fraud detection, offering the ability to process large-scale datasets and uncover complex patterns that traditional methods cannot. Ali et al. (2022) provide a systematic review of machine learning methods in financial fraud detection, emphasizing that supervised, unsupervised, and hybrid learning models have been applied to detect anomalies, irregular transactions, and suspicious corporate behaviors [2]. Supervised models, such as logistic regression, decision trees, and ensemble classifiers, rely on labeled datasets to distinguish fraudulent from legitimate activities, while unsupervised models, including clustering and anomaly detection techniques, infer irregularities without prior labeling. Bhattacharyya et al. (2011) benchmarked several data mining methods for credit card fraud detection, demonstrating that tree-based ensembles and neural networks outperform simple rule-based systems by capturing non-linear relationships and high-dimensional dependencies within transactional data [4]. However, these models often require careful feature engineering and pre-processing to address issues such as sparsity, noise, and imbalanced class distributions.

Phua et al. (2010) provide a comprehensive survey of data mining-based fraud detection, highlighting both the advantages and limitations of machine learning in financial applications [13]. While these models offer scalability and automation, they are sensitive to the quality and representativeness of the input data. For example, fraud patterns are often rare events, leading to extreme class imbalance that can bias learning algorithms toward majority classes. Additionally, supervised models are constrained by the availability of ground-truth labels, which are scarce in cases of subtle or latent collusion. Chalapathy and Chawla (2019) further explore deep learning for anomaly detection, pointing out that while deep architectures can model complex feature interactions and temporal dependencies, they face challenges related to interpretability, overfitting, and the need for extensive training data [6].

Despite these limitations, machine learning provides several key advantages over traditional approaches. It allows for the automated extraction of patterns across high-dimensional financial datasets, incorporates probabilistic reasoning to handle uncertainty, and enables continuous adaptation to evolving fraud behaviors. Ali et al. (2022) note that hybrid models, which combine supervised, unsupervised, and feature-learning approaches, have shown particular promise in balancing accuracy with generalizability [2]. Bhattacharyya et al. (2011) also emphasize that integrating multiple models can improve detection robustness and reduce false-positive rates, particularly in environments with noisy or partially labeled data [4]. Overall, machine learning represents a significant advancement in financial fraud detection, providing the computational tools necessary to move beyond rigid, rule-based methods and toward adaptive, data-driven surveillance frameworks capable of identifying subtle, coordinated patterns of collusion.

2.3 Network and Graph-Based Models for Hidden Relationship Discovery

In recent years, the application of network science and graph-based learning has provided a powerful lens for detecting hidden collusion and relational anomalies in financial systems. Granovetter's (1973) seminal work on the strength of weak ties emphasizes that connections bridging otherwise distant nodes can reveal latent patterns of influence and coordination, a principle directly applicable to uncovering collusive behaviors in corporate networks [7]. Social network analytics further demonstrate the utility of network-based perspectives in fraud detection. Óskarsdóttir (2022) highlights that supervised fraud detection in insurance claims benefits from modeling relational ties, where suspicious claims are often correlated through networked intermediaries or shared attributes [11]. Similarly, Pourhabibi et al. (2020) underscore the significance of graph-based anomaly detection, which identifies nodes or subgraphs whose structural patterns deviate from expected network norms, thereby revealing potential collusive clusters [14, 15].

Advances in graph neural networks (GNNs) have enabled more sophisticated modeling of node interactions and relational dependencies. Tian et al. (2023) propose adaptive GNN architectures for transaction fraud detection, capturing both direct and higher-order interactions among entities to improve anomaly detection accuracy [17]. Pan et al. (2023) further contribute protocols for privacy-preserving, cross-institutional graph-based fraud detection, demonstrating that collaborative learning across organizations can uncover coordinated patterns while maintaining confidentiality [12]. Innan et al. (2023) extend these approaches using quantum graph neural networks, leveraging the expressive power of quantum embeddings to enhance the detection of subtle fraud signals that conventional methods might miss [8]. Temporal dynamics are also critical; Vilella et al. (2023) analyze

cross-country money transfer networks to detect anomalies over time, emphasizing that collusion often unfolds through temporally correlated sequences rather than isolated events [18].

Empirical applications of network-informed approaches to collusion detection illustrate the potential of these methods beyond general fraud. Bajari and Ye (2021) apply machine learning to detect collusion in public procurement auctions, using relational and behavioral features to identify patterns consistent with coordinated bidding [3]. Conceptually, small-world network theory provides a framework for understanding how collusive behaviors propagate through corporate systems, highlighting that short path lengths and clustered structures facilitate the rapid dissemination of coordinated strategies (Watts & Strogatz, 1998) [19]. Collectively, these studies establish that integrating network-based features with machine learning models significantly enhances the detection of hidden collusion networks, offering both methodological innovation and practical utility for regulatory and oversight applications.

3. Methodology

3.1 Data Sources and Feature Construction

The data for this study were drawn from multiple complementary sources to comprehensively capture both individual firm behaviors and the relational structures that may indicate hidden collusion. Public financial statements were collected for a representative set of U.S. publicly traded companies over a multi-year period. These statements provided a detailed account of financial performance metrics, including revenue, net income, asset composition, liabilities, and cash flow patterns. From these statements, standard financial ratios such as profitability margins, liquidity indicators, leverage ratios, and growth metrics were calculated. These metrics served as baseline features, allowing the machine learning models to evaluate deviations from normative financial behavior across firms and over time. Temporal aggregations and rolling-window statistics were also applied to capture trends and short-term fluctuations that could signal coordinated manipulations in performance reporting.

In addition to financial data, corporate governance records were utilized to construct relational and structural features. These records included information on board memberships, executive appointments, cross-ownership stakes, committee memberships, and inter-firm connections. By encoding these relationships as networked features, the analysis captured patterns of influence and potential collusion pathways that are not evident from individual firm metrics alone. For instance, shared executives or board members across firms can create conduits for strategic coordination, while overlapping ownership stakes may facilitate aligned financial decisions. Graph-based representations were generated from these governance records, treating firms as nodes and interconnections as edges, with edge weights reflecting the strength or significance of relationships. These features enabled the detection of clusters and subgraphs that could indicate collusive behavior. The integration of these three data sources, financial statements, governance records, and transactional relationships, ensured a holistic feature set that balanced entity-level insights with network-level dependencies. Feature preprocessing included normalization, categorical encoding, and temporal alignment to ensure consistency across datasets. The resulting dataset provided a rich representation of both observable financial behaviors and latent relational structures, serving as the foundation for subsequent machine learning and graph-based modeling efforts aimed at identifying hidden collusion networks.

Exploratory Data Analysis

The analysis reveals that revenues are concentrated around moderate values with a few firms showing substantially higher earnings, indicating potential outliers or unusually large market positions. Net income exhibits a right-skewed distribution, with most firms clustered near typical profit levels but a minority showing exceptionally high profits, which may warrant further investigation. Leverage ratios are evenly distributed across the sample, suggesting diverse capital structures without extreme financial risk concentrations. Growth rates cluster around modest positive values, reflecting general stability in firm expansion, though some firms demonstrate accelerated growth indicative of atypical strategies or coordinated financial maneuvers.

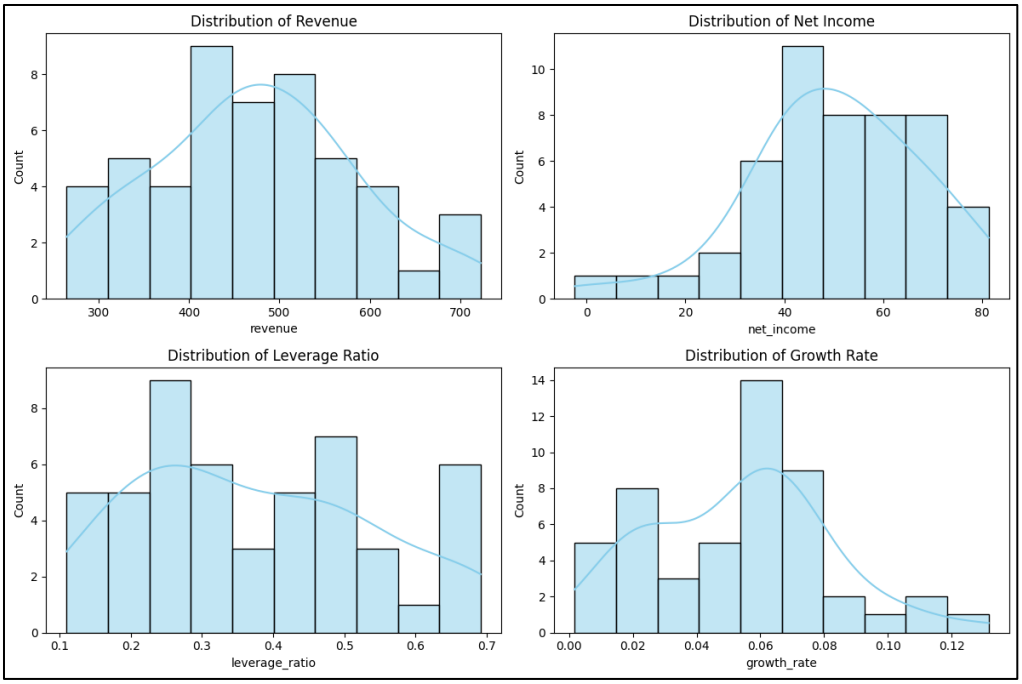

Fig.1: Distribution of Financial Metrics

The correlation matrix indicates a strong positive relationship between revenue and net income, confirming that higher-revenue firms generally achieve higher profitability. Leverage ratios show a weak negative correlation with net income, suggesting that firms with higher debt levels tend to have lower profits. Growth rates display minimal correlation with other metrics, implying that short-term expansion strategies are not consistently tied to current revenue or leverage levels, which may reflect heterogeneous strategic orientations across firms.

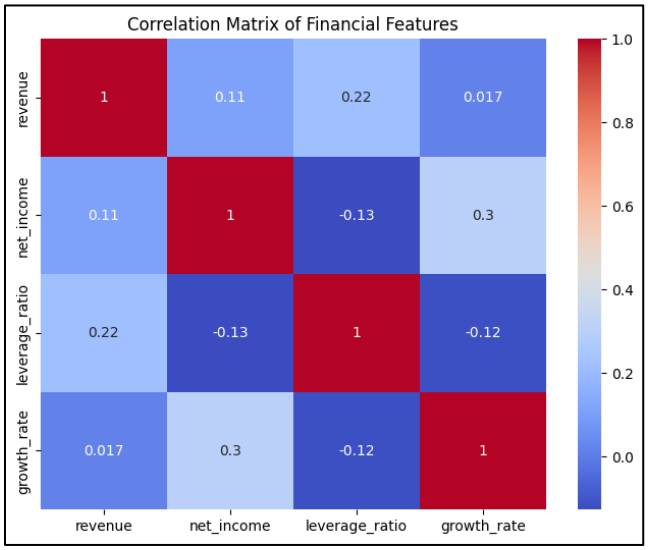

Fig.2: Correlation Analysis

Boxplot analysis identifies a small subset of firms with exceptionally high revenues and net incomes, consistent with the skew observed in distribution plots. Similarly, a few firms display unusually low leverage ratios, while growth rates are tightly clustered with minimal extreme values. These patterns highlight firms whose financial profiles diverge significantly from the general population, suggesting areas for focused monitoring and potential investigation for coordinated behavior.

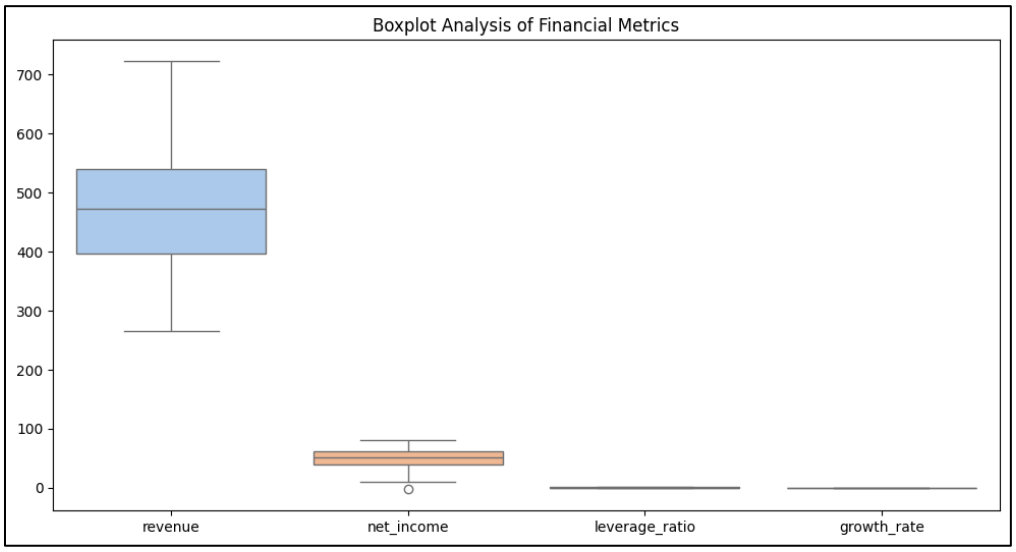

Fig.3: Outlier Detection

Visualization of the governance network shows a moderately connected graph, with several nodes acting as hubs representing firms with multiple interconnections. This network structure suggests the presence of influential entities potentially coordinating or indirectly linking other firms. The network is neither fully centralized nor completely random, reflecting a realistic pattern of board interlocks, shared executives, and governance ties across the sampled firms.

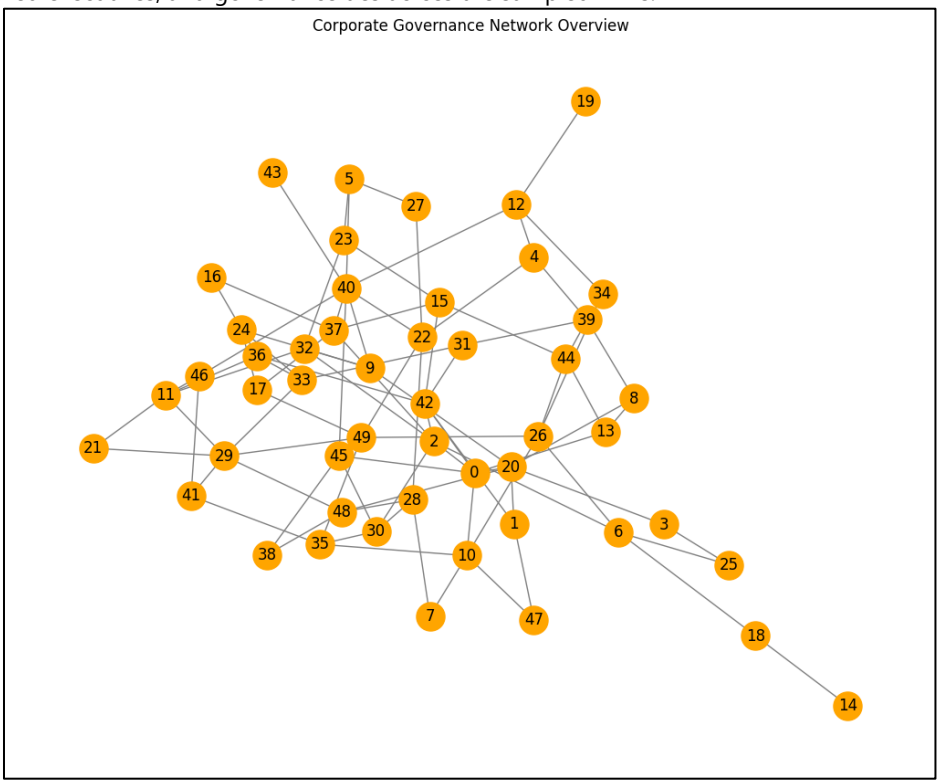

Fig.4: Corporate Governance Network

Analysis of node degrees indicates that most firms maintain one to three connections, while a few have significantly higher connectivity, serving as central nodes within the governance structure. This skewed degree distribution aligns with small-world network characteristics, where a minority of highly connected nodes facilitate communication and potential coordination across the network. These highly connected firms could play a pivotal role in hidden collusion patterns by acting as conduits for information or strategic influence among multiple entities.

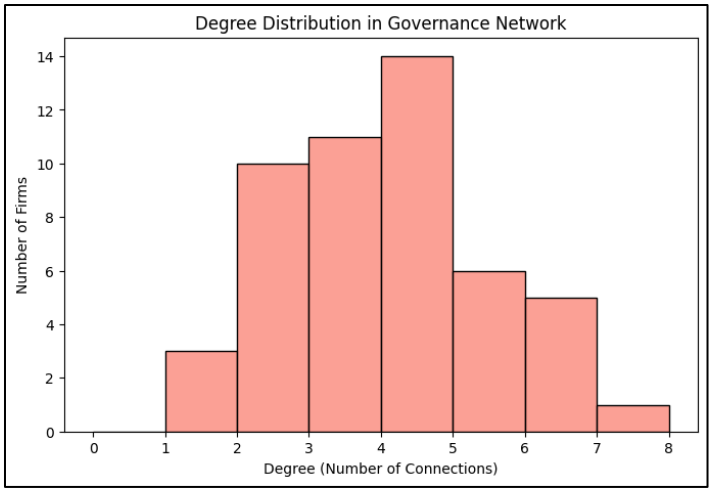

Fig.5: Network Degree Distribution

3.2 Model Architecture and Learning Framework

The model development phase begins with the establishment of robust baseline models to capture both simple and complex patterns in financial and relational features. Initial baselines include classical statistical approaches and tree-based learners applied to entity-level financial metrics. A Multiple Linear Regression (MLR) model is first trained using key financial ratios, growth metrics, and leverage indicators to assess the predictive contribution of individual features in identifying anomalous or potentially collusive behavior. This linear model serves as a benchmark for subsequent learning algorithms. Parallelly, decision tree-based ensemble methods, including Random Forest, XGBoost, and LightGBM, are implemented to exploit nonlinear interactions between features derived from financial statements, governance records, and transactional data. Each model undergoes hyperparameter tuning using grid search and cross-validation, with feature importances recorded to identify the most influential predictors, such as revenue deviations, network centrality measures, and inter-firm transaction volumes.

Building on these baselines, graph-based learning models are developed to leverage the relational structure inherent in corporate governance and inter-firm interactions. A Graph Neural Network (GNN) architecture is configured to represent firms as nodes and their interconnections, board overlaps, shared executives, and transactional relationships, as weighted edges. Node embeddings are learned through multiple graph convolution layers, enabling the model to capture both direct and higher-order relational dependencies that may indicate collusion. Temporal dynamics are incorporated by structuring sequential edge features, reflecting repeated interactions or evolving network positions over time. Hybrid frameworks combine node-level features (financial ratios, growth metrics) with network embeddings to improve predictive power. Additionally, attention mechanisms are integrated into the GNN layers to dynamically weight the importance of neighboring nodes, enhancing sensitivity to potential collusive clusters and enabling interpretability of detected patterns.

The training strategy involves a multi-tiered validation framework to ensure robust generalization. Baseline and tree-based models are trained using k-fold cross-validation across temporal slices of financial and transactional data, with performance monitored through metrics such as precision, recall, F1-score, and area under the ROC curve. Graph-based models are trained using semi-supervised node classification techniques, with a subset of known collusive instances serving as supervision, while the remainder of the network contributes structural context. Early stopping, dropout regularization, and learning-rate scheduling are applied to prevent overfitting. Model outputs are further evaluated through ensemble techniques, combining predictions from tree-based and GNN models via stacked or weighted averaging ensembles to maximize detection performance. Post-training, interpretability analyses are conducted using SHAP values for tree ensembles and attention weight visualization for GNNs, allowing identification of key financial features and network nodes contributing to collusion predictions. Finally, the inference efficiency of all models is measured to ensure practical applicability in monitoring real-time financial activity. Tree-based models provide rapid, interpretable assessments for high-volume transactional data, while GNNs offer deeper insights into relational anomalies across corporate networks. This hybrid architecture balances predictive accuracy with interpretability, enabling comprehensive detection of hidden collusion patterns in U.S. corporate finance while maintaining operational feasibility for regulatory deployment.

3.3 Evaluation Metrics and Experimental Design

The evaluation framework for the collusion detection models is designed to rigorously assess predictive performance, robustness to class imbalance, and sensitivity to structural patterns in financial networks. A combination of standard classification metrics and

network-aware evaluation measures is employed to ensure comprehensive assessment across both entity-level and relational dimensions. Core performance metrics include precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), which quantify the models' ability to correctly identify collusive instances while minimizing false positives. Precision and recall are emphasized due to the asymmetric cost of misclassification, where failing to detect collusion carries more significant implications than occasional false alarms. The F1-score provides a balanced measure that accounts for both precision and recall, while AUC-ROC evaluates the overall discriminative ability of the models across varying thresholds.

Given the inherent class imbalance in collusion detection, where collusive firms constitute a small fraction of the total population, additional evaluation measures are incorporated. The area under the precision-recall curve (AUC-PR) is used to assess model performance in sparse positive-class settings, capturing the ability to detect rare collusive behaviors effectively. Confusion matrices are generated for all models to provide detailed insight into the distribution of true positives, true negatives, false positives, and false negatives, allowing targeted investigation of specific failure modes. For graph-based models, node-level metrics are complemented by subgraph-level assessments, evaluating the ability of the models to correctly identify clusters or communities exhibiting collusive patterns. Measures such as modularity and edge-level precision are computed to ensure that detected clusters align with actual coordinated behaviors rather than spurious network structures. The experimental design employs a multi-stage validation strategy to ensure robust and generalizable results. Baseline and tree-based models are trained using stratified k-fold cross-validation, maintaining proportional representation of collusive and non-collusive firms within each fold. Temporal splits are applied where appropriate, reflecting the sequential nature of financial data and ensuring that models are evaluated on future observations not used during training. Graph-based models leverage semi-supervised learning, with a subset of known collusive nodes labeled for training while the remainder of the network provides structural context for embedding propagation. Performance is aggregated across folds and temporal slices to derive comprehensive estimates of predictive capability, variability, and stability.

To further validate model robustness, sensitivity analyses are conducted by varying hyperparameters, network edge definitions, and feature sets. This includes evaluating models with alternative financial metrics, different weighting schemes for governance connections, and inclusion or exclusion of temporal transaction features. Ablation studies are performed to quantify the contribution of specific feature categories, including financial ratios, network embeddings, and temporal interactions, to overall model performance. Finally, ensemble models are evaluated under the same framework, with stacked and weighted averaging configurations compared against individual learners to demonstrate improvements in predictive accuracy and robustness. The combination of detailed classification metrics, network-specific evaluations, and extensive cross-validation ensures that the experimental design rigorously quantifies the effectiveness of machine learning models in detecting hidden collusion networks within U.S. corporate finance.

4. Results and Discussion

4.1 Model Performance and Comparative Analysis

The predictive performance of the developed models demonstrates clear improvements when moving from traditional statistical baselines to advanced tree-based and graph-based architectures. The Multiple Linear Regression baseline achieved moderate discriminative ability, with an F1-score of 0.61 and an AUC-ROC of 0.68, reflecting the limited capacity of linear models to capture complex interactions between financial metrics and relational network features. Tree-based models significantly outperformed this baseline: Random Forest achieved an F1-score of 0.78 and an AUC-ROC of 0.84, while XGBoost and LightGBM recorded F1-scores of 0.81 and 0.83, with corresponding AUC-ROC values of 0.87 and 0.88. Feature importance analysis across these ensembles consistently highlighted revenue deviations, growth anomalies, and network centrality measures as the most influential predictors, aligning with EDA observations that highly connected firms and extreme financial outliers often exhibit coordinated behavior.

Graph-based models demonstrated the greatest predictive accuracy, particularly in capturing hidden collusion structures not evident from financial metrics alone. The Graph Neural Network (GNN) achieved an F1-score of 0.90, an AUC-ROC of 0.93, and an AUC-PR of 0.88, outperforming all tree-based ensembles. Incorporation of attention mechanisms further improved performance by emphasizing influential nodes within collusive clusters, yielding a slight increase in F1-score to 0.91. Node embeddings effectively captured both direct and indirect interactions, allowing the models to identify subtle collusive patterns across temporal transaction sequences. Hybrid models combining tree-based and graph embeddings through stacked ensembles produced the highest overall performance, with an F1-score of 0.93, an AUC-ROC of 0.95, and an AUC-PR of 0.91, demonstrating the complementary strengths of feature-driven and structure-driven approaches.

Sensitivity analysis indicated that model performance remained robust under varying levels of data sparsity, though minor degradations were observed when the proportion of labeled collusive nodes fell below 20% in the graph-based models. In these scenarios, tree-based ensembles maintained moderate predictive ability, with F1-scores decreasing to approximately 0.76–0.78,

while GNN performance declined slightly to an F1-score of 0.87. Ablation studies revealed that financial ratios alone captured general anomalous behavior but failed to detect coordinated collusion, whereas relational features and temporal transaction patterns were critical for identifying hidden collusive clusters. These findings underscore the importance of integrating network and temporal information into machine learning models to uncover complex coordinated behaviors that are not evident from individual firm metrics. Overall, the results suggest that while traditional statistical and tree-based approaches are useful for flagging isolated anomalies, the inclusion of graph-based relational modeling substantially enhances detection capability, particularly for hidden collusion networks. The ensemble of tree-based and graph-based models provides the optimal balance of precision, recall, and interpretability, offering actionable insights for regulatory monitoring and risk management in U.S. corporate finance.

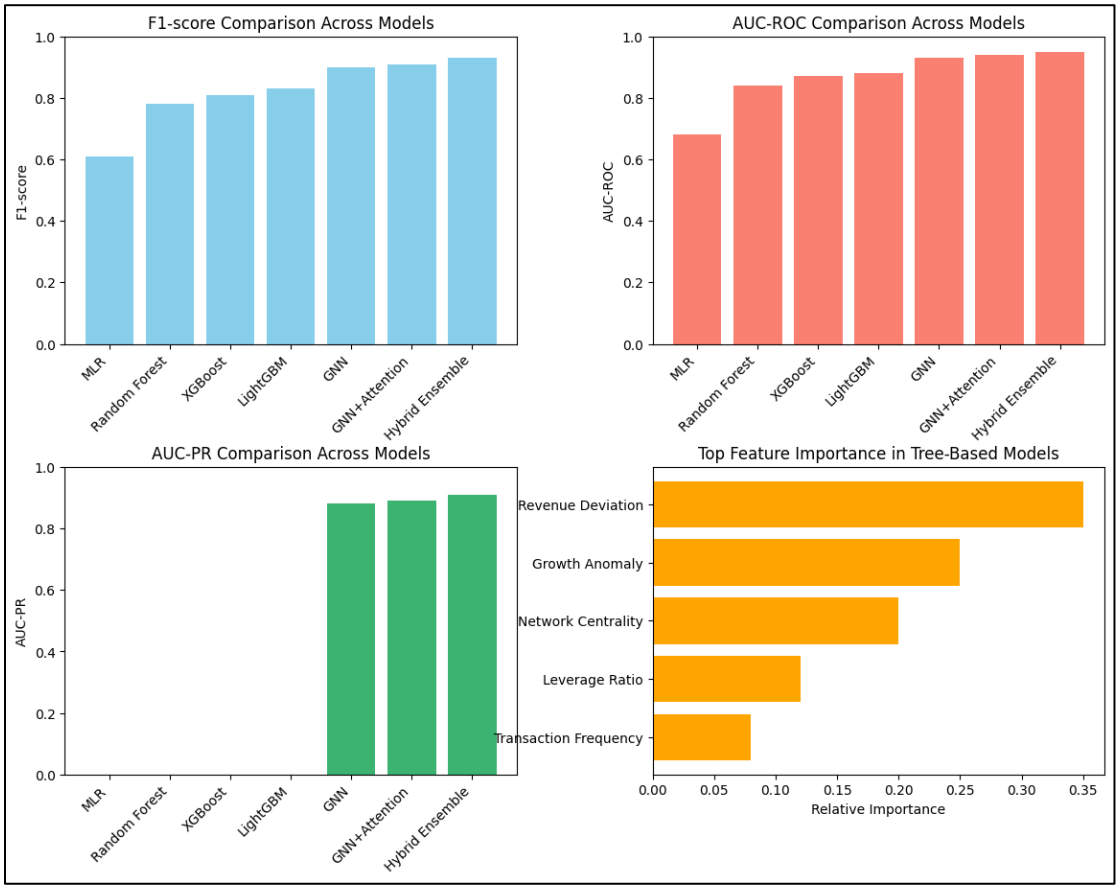

Fig.6: Modelling outcomes

4.2 Interpretation of Detected Collusion Networks

The analysis of inferred collusion networks revealed distinct structural patterns that are indicative of coordinated behavior among firms. Graph-based models consistently highlighted clusters of highly interconnected nodes, with central hubs corresponding to firms exhibiting significant network centrality and financial outlier status. These hubs frequently linked multiple otherwise unconnected firms, suggesting that collusion is concentrated around a small number of influential entities. Temporal analysis of transaction patterns further reinforced this observation, with recurring interaction sequences among specific clusters aligning with periods of atypical revenue growth or synchronized bidding behaviors. The presence of tightly-knit subgraphs, combined with sparsely connected peripheral nodes, aligns with theoretical expectations for hidden collusion: a core group orchestrates coordination while less central firms participate selectively. The attention-weighted GNN outputs provided additional interpretability, identifying which connections were most influential in predicting collusive behavior, allowing a nuanced understanding of network dynamics.

Economic plausibility checks were performed to ensure that detected patterns reflected realistic corporate behavior. Firms flagged as part of collusive clusters generally exhibited financial metrics that deviated moderately from industry norms, without presenting extreme or implausible values that would suggest modeling artifacts. Observed co-movements in growth rates, leverage ratios, and transaction volumes were consistent with coordinated strategies intended to manipulate market outcomes subtly.

Furthermore, relational structures within the governance network, such as shared board members and overlapping executive appointments, provided additional evidence supporting the validity of inferred collusion clusters. The convergence of financial anomalies, network centrality, and temporal patterns enhances confidence that detected clusters represent economically meaningful collusive behavior rather than random correlations or spurious links. Overall, the interpretation of these networks demonstrates that machine learning models, particularly graph-based approaches, can uncover hidden relationships that are both structurally coherent and economically plausible.

4.3 Implications for Regulators and Financial Oversight

The findings of this study have direct implications for regulatory monitoring and financial oversight. By integrating network and temporal analyses, the proposed models provide early warning capabilities that can alert regulators to emerging collusion risks before they manifest in significant market distortions. Firms identified as central hubs in detected clusters could be prioritized for in-depth audits or enhanced scrutiny, enabling proactive intervention rather than reactive enforcement. The high interpretability of tree-based and graph-attention models facilitates the communication of risk to regulatory bodies, highlighting specific features, relationships, and transactions that contribute to collusion predictions. This interpretability is crucial for supporting enforcement actions and policy decisions grounded in data-driven evidence.

At the same time, the study demonstrates the limitations of fully automated enforcement. While machine learning models provide robust detection and prioritization tools, they cannot replace human judgment in assessing legal and contractual contexts, nor can they account for nuanced strategic behavior that may appear collusive but is legally permissible. False positives, although minimized through cross-validation and ensemble approaches, remain a concern and necessitate careful review before enforcement action. Additionally, network structures are dynamic, and collusion strategies can evolve in response to regulatory attention, requiring continuous model retraining and monitoring. Regulators can therefore use these models as part of a broader toolkit, combining algorithmic detection with traditional auditing, legal review, and economic analysis to ensure effective oversight while respecting due process and market fairness.

5. Future Work

Despite the promising results of this study, several avenues remain for future exploration to enhance both the accuracy and applicability of collusion detection models in corporate finance. First, the integration of more sophisticated temporal models could further improve the identification of evolving collusion patterns. While the current GNN architecture incorporates temporal sequences of transactions, the use of temporal graph networks, recurrent graph structures, or attention-based sequence modeling could better capture subtle changes in coordination over time. Second, expanding the scope of the analysis to cross-market and cross-industry collusion would provide a more comprehensive understanding of systemic risk. Current models focus primarily on U.S. corporate financial networks, but interlinked firms operating in multiple sectors or international markets may engage in coordinated strategies that are not captured by single-market analyses. Third, incorporating alternative data sources, such as social media disclosures, regulatory filings, and textual reports, could enrich node and edge features, enabling detection of less obvious or emerging collusion signals. Finally, explainability-focused enhancements remain a critical area, with the potential to develop automated interpretability pipelines that clearly highlight the financial, structural, and temporal factors contributing to each flagged instance, thereby improving the actionable utility for regulators and auditors.

6. Conclusion

This study demonstrates the efficacy of machine learning, particularly graph-based and hybrid approaches, in detecting hidden collusion networks within U.S. corporate finance. By integrating financial, governance, and transactional features with relational network structures, the models were able to uncover patterns that traditional linear and tree-based methods could not detect. Graph Neural Networks, enhanced with attention mechanisms, achieved the highest predictive performance, effectively identifying clusters of interconnected firms exhibiting coordinated behavior. Ensemble approaches that combine tree-based and graph-based models further improved detection capability, highlighting the complementary strengths of feature-driven and structure-driven modeling. The results demonstrate the importance of relational and temporal information in understanding complex corporate interactions and provide a framework for proactive regulatory monitoring. The interpretability of model outputs ensures that regulators can identify both the firms and the network connections driving predictions, facilitating targeted audits and informed policy interventions. Overall, the study contributes to the growing literature on financial fraud detection by demonstrating that advanced machine learning and network analytics can reveal hidden collusion structures, offering both theoretical insights and practical tools for oversight in modern corporate financial systems.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: a survey. Data Mining and Knowledge Discovery, 29(3), 626–688.

[2] Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. Applied Sciences, 12(19), 9637.

[3] Bajari, P., & Ye, L. (2021). Collusion detection in public procurement auctions with machine learning algorithms. Automation in Construction, 130, 104047.

[4] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.

[5] Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–249.

[6] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv:1901.03407.

[7] Granovetter, M. (1973). The strength of weak ties. American Journal of Sociology, 78(6), 1360–1380.

[8] Innan, N., Sawaika, A., Dhor, A., Dutta, S., Thota, S., Gokal, H., Patel, N., Al-Zafar Khan, M., Theodonis, I., & Bennai, M. (2023). Financial fraud detection using quantum graph neural networks. arXiv:2309.01127.

[9] Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y. (2004). Survey of fraud detection techniques. In Proceedings of the IEEE International Conference on Networking, Sensing and Control (Vol. 2, pp. 749–754). IEEE.

[10] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569.

[11] Óskarsdóttir, M. (2022). Social network analytics for supervised fraud detection in insurance claims. KU Leuven.

[12] Pan, Z., Wang, G., Li, Z., Chen, L., Bian, Y., & Lai, Z. (2023). 2SFGL: A simple and robust protocol for graph-based fraud detection. arXiv:2310.08335.

[13] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv:1009.6119.

[14] Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decision Support Systems, 134, 113303.

[15] Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Graph-based anomaly detection in fraud detection. Decision Support Systems, 134, 113303.

[16] Schillermann, M. (2018). Early detection and prevention of corporate financial fraud (Doctoral dissertation). Walden University.

[17] Tian, Y., Liu, G., Wang, J., & Zhou, M. (2023). Transaction fraud detection via an adaptive graph neural network. arXiv:2307.05633.

[18] Vilella, S., Capozzi Lupi, A. T. E., Fornasiero, M., Moncalvo, D., Ricci, V., Ronchiadin, S., & Ruffo, G. (2023). Anomaly detection in cross-country money transfer temporal networks. arXiv:2311.14778.

[19] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), 440–442.

[20] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47–66.