Computer Fraud and Security
ISSN (online): 1873-7056

Detecting Financial Collusion Through Graph Analytics: A Procure-to-Pay and Payroll Fraud Detection Framework

Karishma Velisetty

Independent Researcher, USA

Abstract

Collusion remains one of the most difficult fraud schemes to detect because it is designed to defeat traditional rule-based controls and segregation of duties. Conventional audit analytics focus on individual transactions and attributes, while collusion operates through relationships between people, vendors, bank accounts, and approval chains. This article proposes a graph-analytics approach for identifying collusive behavior in procure-to-pay and payroll processes. The article introduces a network data model, risk indicators, and detection algorithms that combine structural graph metrics with financial attributes. The proposed framework demonstrates how internal audit functions can move from exception testing to relationship-driven continuous monitoring, improving detection of kickbacks, ghost employees, vendor favoritism, and approval rings.

Keywords: Graph Analytics, Collusion Detection, Procure-to-Pay Fraud, Payroll Fraud, Community Detection

1. Introduction

1.1 Background and Motivation

Organizations invest heavily in controls such as approval hierarchies, three-way matches, and segregation of duties. These mechanisms presume that isolated individuals violate a rule to commit fraud. Collusion breaks this assumption fundamentally. When two or more actors cooperate—such as a buyer and vendor, or a payroll clerk and supervisor—transactions may appear fully compliant while still being fraudulent. The limitations of traditional detection methods in dealing with schemes involving multiple people have led to a pressing need for new analytical approaches that can reveal hidden connections across different organizations. The scale of the problem is considerable—financial fraud is estimated to cost the global economy over $5 trillion annually, a figure that reflects the systemic failure of traditional detection mechanisms to keep pace with increasingly sophisticated schemes [3].

1.2 Problem Statement

Most audit analytics examine transactions independently, focusing on duplicate invoices, round-dollar amounts, or weekend postings. Collusion, however, leaves relational footprints rather than transactional anomalies. Modern enterprises generate massive relational data—who approved what, which employee manages which vendor, shared addresses, bank accounts, devices, and timing sequences. Traditional relational database systems are structurally incapable of connecting these dots across domains, as performing many joins between several tables is costly and relationships are not stored natively. Graph databases, by contrast, deliver query performance approximately 1,000 times faster than relational databases in graph-like structures, making them uniquely suited for real-time traversal of complex relational data [1]. As mentioned in the research, graph-based models see relationships as important parts of the analysis instead of just extra details, which helps find fraud networks that regular audit tools might miss.

1.3 Research Objectives

This paper aims to create a useful method for internal audit teams to use graph analytics with current enterprise resource planning and human resource information system data, without needing advanced data science tools. The goals are to create a network data model for the procure-to-pay and payroll processes, find signs of risk in the structure and transactions, and suggest detection methods that link graph metrics with financial data. The broader goal is to shift audit practice from reactive investigation to proactive, relationship-driven risk discovery. This change is shown to be useful because there are records of cases where adding graph-based features to machine learning fraud detection models led to a 50% better recall and a 50% better precision compared to models that only used traditional non-graph features.

1.4 Scope and Significance

The scope of this research encompasses the procure-to-pay and payroll processes, which represent two of the most fraud-prone domains in organizational finance. Financial fraud contributes to an estimated global loss of $5 trillion annually, underscoring the scale of the challenge that modern detection frameworks must address [3]. The significance of the proposed approach lies in the potential for graph analytics to detect schemes that remain invisible to rules-based systems—including kickback networks, ghost employees, vendor favoritism, and approval rings. By modeling the enterprise as a connected graph of actors and transactions, auditors gain the ability to reason across multiple hops of indirect relationships, enabling detection of collusion that spans departments, vendors, and payment channels.

2. Literature Review

2.1 Fraud Detection in Financial Networks

The application of graph-based methods to financial fraud detection has gained substantial momentum over the past decade. Research has demonstrated that graph databases provide a structurally superior representation of financial crime networks, with graph database query performance in graph-like structures being approximately 1,000 times faster than equivalent relational database operations, enabling real-time detection of complex fraud patterns [1]. The relational nature of organized financial crime—where actors coordinate through shared resources, intermediaries, and layered transactions—maps naturally onto graph structures where nodes represent entities and edges represent interactions. Studies in the fintech field have found that using graph analysis along with machine learning can greatly enhance fraud detection, with one case showing a 50% boost in recall and a 50% increase in precision for a fraud prediction model after incorporating graph-based features.

2.2 Graph Theory and Community Detection in Fraud Contexts

Graph theory provides the formal mathematical foundation for the detection methodologies proposed in this paper. Key concepts, including centrality, clustering coefficients, and shortest path algorithms, have been successfully adapted for fraud analytics. Community detection, in particular, has emerged as a critical technique for identifying groups of entities that interact disproportionately with one another—a behavioral signature consistent with collusion. Systematic reviews of AI-driven fraud detection have documented that financial fraud contributes to an estimated global loss of $5 trillion annually and that traditional detection methods consistently fail to keep pace with evolving fraudulent strategies [3]. Research on credit card fraud has shown that graph database models can quickly navigate complicated relationships, which makes them useful for both investigations and ongoing monitoring.

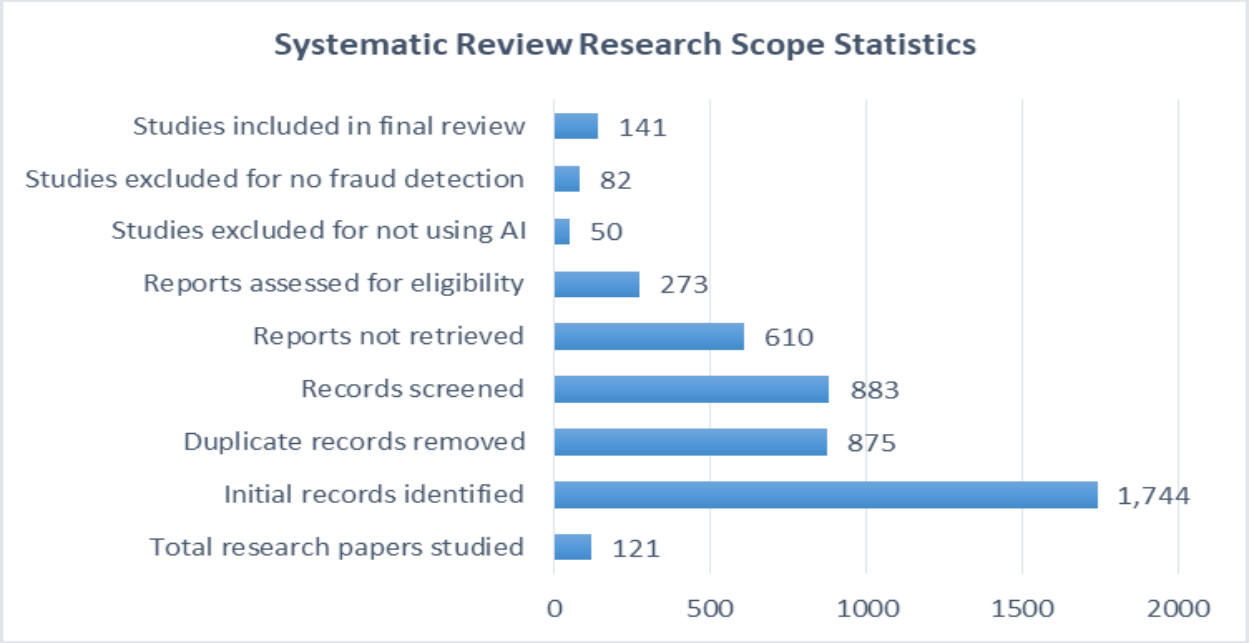

Fig. 1: Systematic Review Research Scope Statistics [3]

2.3 Collusion in Procurement and Bidding Processes

Collusion in procurement is a well-known type of financial crime in which competitors or people inside a company work together to change the results of bidding and contracting processes. Research in this domain has shown that relational patterns—such as shared beneficial owners, coordinated bid timing, and overlapping contact information among vendors—are reliable indicators of collusive behavior that resist detection by transactional controls alone [5]. Using machine learning with graph representations has shown better results in spotting collusion by understanding hidden patterns in collusive networks from past data that has been labeled. In a documented production implementation, one graph-based feature rose to become the second most important feature in the entire fraud detection model, illustrating the practical significance of graph-derived signals in real-world deployment [2]. These results help shape the detection system suggested in this paper, especially in creating subgraphs of vendor relationships and spotting unusual approvals.

2.4 AI and Machine Learning Integration

The integration of artificial intelligence with graph analytics represents the current frontier of fraud detection research. Systematic reviews of the field have shown that there has been rapid growth in AI-driven models that use graph neural networks, anomaly detection, and supervised classification to identify fraud in financial networks. The total financial loss from fraud worldwide is estimated to be $5 trillion each year, based on a thorough review of fraud detection research that looked at 121 papers and included results from 141 studies chosen from an initial 1,744 records. Using graph-enhanced machine learning methods has been proven to be better than using only rule-based or statistical models; in one real-world application, adding graph connectivity features to a machine learning model led to a 50% increase in both recall and precision, and a graph-based feature was the second most important factor in the final model.

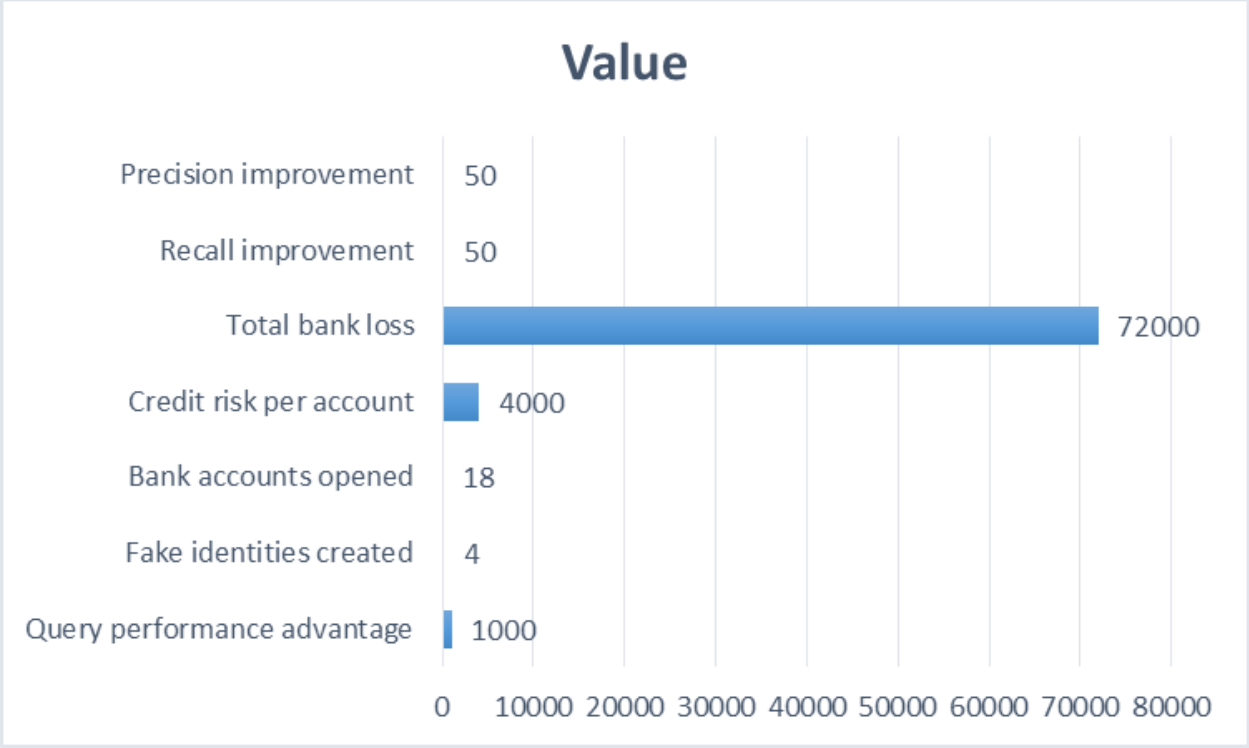

Fig. 2: Graph Database and Fraud Detection Model Performance Metrics [1][2]

3. Conceptual Framework and Data Model

3.1 Network Data Model Design

The proposed framework models the organizational financial environment as a heterogeneous graph in which nodes represent distinct entity types—employees, vendors, invoices, purchase orders, bank accounts, addresses, approval roles, and optionally devices or IP addresses—and edges represent the relationships between them. Each edge carries attributes including transaction amount, date, and channel. enabling time-aware analysis of relational patterns. The practical danger of shared-identity fraud networks is well illustrated by documented cases in which just 2 individuals used 2 phone

numbers and 2 addresses to create 4 fake identities, collectively opening up to 18 bank accounts and generating a potential credit risk of approximately $4,000 per account—totaling a potential bank loss of around $72,000—a scheme entirely invisible to discrete transaction analysis but detectable through entity relationship graph traversal [1]. This design shows all the complicated ways organizations interact in a way that allows for checking connections that go through several steps so it can find and study indirect relationships that involve multiple middle points.

3.2 Data Integration and Entity Resolution

The data model uses information from many different organizational systems, such as Active Enterprise Resource Planning Procure-to-Pay tables, Human Resource Information System payroll records, vendor master files, payment files, and identity systems. Integrating these sources requires robust entity resolution — the process of standardizing and reconciling entity representations across systems that may use inconsistent naming conventions, address formats, or identifier schemes. Fuzzy matching techniques are applied to vendor names and addresses to identify entities that represent the same real-world actor under different representations. This step is critical because collusive actors frequently exploit inconsistencies in master data to obscure their relationships [6]. Graph databases address this challenge structurally, as relationships are stored as first-class citizens of the data model rather than computed through expensive join operations—a design advantage that delivers query performance approximately 1,000 times faster than relational databases in graph-like structures and provides high horizontal scalability that relational systems cannot match [1].

3.3 Risk Scoring and Governance

The risk scoring part of the framework gives numerical risk scores to nodes by combining different factors, including the node's structure in the graph, signs of financial problems, and behavior patterns. Structural risk reflects the node's position and connectivity within the graph—for example, unusually high centrality or membership in a dense community. Financial anomaly indicators capture attributes such as price variance, split invoice patterns, and overtime spikes. Behavioral risk reflects timing anomalies such as after-hours approvals or new vendor velocity. Internal audit teams can calibrate sensitivity based on organizational risk appetite by combining these components through a configurable weighting formula [2]. The real-world effect of these combined systems is shown by actual production cases where models using graphs improved their accuracy and completeness by 50% compared to standard models, and one graph feature became the second most important factor in the final model. Governance considerations include strict access controls for sensitive payroll and banking data, model documentation requirements, a defined investigation protocol, and a false-positive review process to maintain analytical credibility over time.

4. Graph-Based Detection Methodology

4.1 Structural Graph Indicators

The detection methodology begins with the computation of structural graph indicators that signal anomalous relational patterns. Centrality measures identify vendors or employees that occupy disproportionately influential positions in the transaction network—for example, a vendor connected to an unusually large number of employees may be a focal point of steering or kickback activity [7]. Closed triangles, where three nodes all approve each other, show a strong sign of approval ring behavior, where employees help each other bypass independent checks on their transactions. Shared resource bridges — two employees linked through a common bank account or address provide evidence of coordination that may indicate ghost employee schemes or benefit-sharing arrangements. Graph databases are really important for finding this kind of fraud because they can process queries about 1,000 times faster than regular databases when it comes to graph-like data, allowing us to quickly track and spot active fraud rings before they cause big losses.

Detection Layer	Technique	Purpose
Structural Indicators	Centrality Measures	Identify disproportionately influential nodes
Structural Indicators	Closed Triangle Detection	Identify approval ring cycles
Structural Indicators	Shared Resource Bridge Analysis	Surface-coordinated bank/address sharing

Community Analysis	Louvain Algorithm	Partition graph into collusive clusters
Transactional Overlay	Price Variance Analysis	Detect systematic overpricing
Transactional Overlay	Split Invoice Detection	Identify sub-threshold coordinated invoicing
Behavioral Analysis	After-hours Approval Timestamps	Flag timing anomalies
Path Analysis	Multi-hop Traversal	Detect indirect entity linkages

Table 1: Graph-Based Detection Methodology Components

4.2 Community Detection and Cluster Analysis

Community detection algorithms partition the organizational graph into subgraphs characterized by dense internal connectivity and sparse external connections. Clusters exhibiting high internal transaction volume—particularly where the majority of departmental spend is directed toward a small group of interconnected vendors—are flagged as high-risk communities warranting further investigation [8]. The Louvain algorithm is a fantastic choice for this task because it can handle large graphs and find communities at different levels of detail, allowing analysts to look at risks across the whole organization or focus on specific subgraphs. The structural danger of undetected collusion networks is illustrated by cases in which as few as 2 individuals, operating through 4 fake identities and 18 bank accounts, generated a potential loss exposure of $72,000 from a single coordinated scheme—a pattern that community detection methods are specifically designed to surface through cluster analysis of shared entity relationships [1].

4.3 Transactional Overlay and Multi-hop Path Analysis

Adding a transactional overlay, which incorporates financial and behavioral details directly into the risk assessment, enhances the structural graph indicators. Price variance analysis looks at billed amounts compared to market standards or similar vendor prices to find consistent overcharging that may be linked to kickback deals. Split invoice detection identifies coordinated patterns of sub-threshold invoicing between a vendor and buyer that collectively exceed approval limits. After-hours approval timestamps and new vendor velocity metrics add a behavioral dimension to the risk profile of flagged nodes [2]. Multi-hop path analysis extends detection beyond direct relationships, enabling the system to identify indirect links—for example, an employee whose relative shares a bank account with a vendor the employee repeatedly approves—that would be entirely invisible to transaction-level controls [1]. In real-life applications of systems that use both graphs and machine learning, graph-based features have proven to be very effective for detecting fraud. One example showed a 50% improvement in both recall and precision after including graph connectivity features, with one particular graph feature being the second most important part of the model.

5. Implementation, Governance, and Audit Application

5.1 Implementation Architecture

The practical implementation of the proposed framework proceeds through a sequence of stages designed to be executable within existing organizational data infrastructure. The first stage includes gathering and pulling data from systems like Enterprise Resource Planning, Human Resource Information System, and payment systems, then making sure that the information about entities is consistent and matches across these sources. The graph is then constructed with timestamped nodes and edges, enabling time-series analysis of relational evolution. The analytics layer uses methods like centrality computation, community detection, path analysis, and anomaly scoring to create a list of risks ranked by importance. A visualization and case management interface show flagged clusters and paths for the auditor to look at, which makes it easier to prioritize investigative work. The scalability advantage of graph databases — which offer horizontal scalability and approximately 1,000 times faster query performance than relational systems in graph-like structures — is a critical enabler of the continuous, large-scale monitoring that effective fraud detection requires [1].

5.2 Shifting the Audit Paradigm

The adoption of graph analytics requires a fundamental reconceptualization of the internal audit function. Traditional audit practice is organized around sample testing of individual transactions and verification of control attributes—

whether an invoice was approved or whether a purchase order was matched to a receipt. Graph-based auditing changes the focus from checking if one transaction followed the rules to looking at the overall patterns and connections around it to see if they match what a legitimate organization would do. This change is shown by real-world examples where adding graph-based features to fraud detection models led to a 50% increase in both recall and precision, proving that understanding relationships is more effective for detection than just refining individual transactions. This change from checking samples occasionally to constantly monitoring relationships helps audit functions find collusion that stays within the rules of individual control compliance.

5.3 Evidentiary Standards and Regulatory Alignment

Graph analytics results are indicators of risk, not conclusions of fraud. Auditors must corroborate flagged relationships with supporting evidence, including contracts, email communications, market pricing comparisons, and vendor due diligence records, before escalating findings. This standard for evidence matches common audit practices and makes sure that the analysis results are used to prioritize issues instead of replacing the need for professional judgment. The global cost of financial fraud at $5 trillion annually reflects the consequences of inadequate detection and underscores the regulatory imperative for organizations to adopt proactive, demonstrably effective fraud risk management frameworks [3]. From a regulatory perspective, the framework strengthens management review controls and supports obligations for assessing fraud risk under relevant financial reporting and internal control standards, providing documented evidence of a mature and systematic approach to fraud risk management [2].

5.4 Challenges and Limitations

The deployment and ongoing operation of the framework must address several significant challenges. Data quality is a primary concern—inconsistent vendor names, incomplete address records, and fragmented payment data can undermine entity resolution and produce a graph that misrepresents actual organizational relationships. Privacy considerations are particularly acute given the sensitivity of payroll and banking data, requiring careful governance of data access and analytical outputs. Understanding the results from graph-based analysis is another challenge, as auditors need to explain how they came up with risk scores and cluster flags in a way that makes sense to management and legal advisors. The scale of the underlying problem—with financial fraud costing an estimated $5 trillion globally each year and a comprehensive review of 141 studies drawn from an initial pool of 1,744 records confirming that no single detection methodology has yet achieved universal effectiveness—reinforces the need for layered, continuously validated analytical frameworks rather than reliance on any single technique [3]. False positives arising from legitimate shared addresses or familial relationships must be systematically reviewed to maintain confidence in the analytical system [10].

Implementation Stage	Key Activity	Governance Consideration
Data Ingestion	Extract ERP, HRIS, payment data	Data access controls
Entity Resolution	Fuzzy matching of names and addresses	Master data quality standards
Graph Construction	Build timestamped nodes and edges	Model documentation requirements
Analytics Layer	Centrality, community detection, scoring	False positive review process
Visualization	Case management interface for auditors	Investigation protocol definition
Audit Application	Targeted substantive testing of flagged clusters	Evidentiary corroboration standards
Regulatory Alignment	Fraud risk assessment documentation	Compliance with internal control standards

Table 2: Implementation Framework Stages and Governance Considerations

Conclusion

Graph analytics is a revolutionary step forward in finding collusion in procure-to-pay and payroll settings. By representing financial activities in an organization as a network of connected people and relationships, the suggested framework reveals the hidden signs of fraud schemes that are intentionally made to look legitimate in each individual transaction. The use of structural graph indicators, community detection, transactional overlay, and multi-hop path analysis allows for a more advanced way to detect fraud that is much better than traditional rule-based audit analytics. The evidence base for this approach is grounded in verified empirical findings: graph database query performance is approximately 1,000 times faster than relational alternatives in graph-like structures, integrated graph and machine learning systems have demonstrated 50% improvements in both recall and precision in production deployments, and the

global cost of financial fraud is estimated at $5 trillion annually—a figure that reflects the systemic inadequacy of current detection approaches. Internal audit teams that use this method go from checking individual transactions on a small amount of data to looking at relationships and keeping an eye on risks across the whole organization all the time. Limitations, including data quality, privacy governance, explainability, and false positive management, must be proactively addressed to sustain analytical credibility. Future research should focus on creating systems that can analyze data in real-time, building partnerships between organizations to detect collusion in the industry, and incorporating ethical guidelines for using artificial intelligence in monitoring employees.

References

[1] Buket Dogan, "The Importance of Graph Databases in Detection of Organized Financial Crimes," in 6th International Conference on Computer Science and Engineering (UBMK), June 2021. Available: https://www.researchgate.net/publication/352159860_The_Importance_of_Graph_Databases_in_Detection_of_Organized_Financial_Crimes

[2] Stanka Dalekova, et al., "Using Graph Analysis and Fraud Detection in the Fintech Industry," Oracle OpenWorld (Technical Whitepaper), 2018/2019. Available: https://www.oracle.com/a/tech/docs/sg-oow2019-using-graph-analysis-and-fraud-detection-in-fintech-industry.pdf

[3] NUSRAT JAHAN SARNA, et al., "AI Driven Fraud Detection Models in Financial Networks: A Comprehensive Systematic Review," IEEE Access, 15 August 2025. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11113282

[4] Debachudamani Prusti, et al., "Credit Card Fraud Detection Technique by Applying Graph Database Model," Arabian Journal for Science and Engineering (Springer), 2021. Available: https://www.semanticscholar.org/paper/Credit-Card-Fraud-Detection-Technique-by-Applying-Prusti-Das/7bb477077968d68aa7a6059d8d6d801fb28274da

[5] André Ormastroni Victor, et al., "Graph Data Mining for Detecting Collusions in Bidding Processes," in Companion Proceedings of the 39th Brazilian Symposium on Databases (SBBD), 2024. Available: https://sol.sbc.org.br/index.php/sbbd_estendido/article/download/30799/30602/

[6] Uri Lapidot, Jay Yu, "Integrating Graph and Machine Learning for Fraud Detection Use Cases," CEUR Workshop Proceedings (AI-OASIS), 2021. Available: https://ceur-ws.org/Vol-2980/paper421.pdf

[7] Everton Schneider dos Santos, et al., "Improving Public Procurement Collusion Detection With Graph-Based Machine Learning Methodologies," IEEE Access (Early Access/Preprint), November 2025. Available: https://www.researchgate.net/publication/399394363_Improving_Public_Procurement_Collusion_Detection_With_Graph-based_Machine_Learning_Methodologies

[8] H. Mardiansyah, et al., "Community Clustering on Fraud Transactions Applied the Louvain-Coloring Algorithm," International Journal of Electronics and Telecommunications, 19 April 2024. Available: https://ijet.pl/index.php/ijet/article/view/10.24425-ijet.2023.146512

[9] Juan Zhang, et al., "Toward Effective Big Data Analysis in Continuous Auditing," Accounting Horizons (American Accounting Association), February 2015. Available: https://www.researchgate.net/publication/276391068_Toward_Effective_Big_Data_Analysis_in_Continuous_Auditing

[10] Dhiman Sarma, et al., "Bank Fraud Detection Using Community Detection Algorithm," International Journal of Advanced Computer Science and Applications (IJACSA), 2020. Available: https://www.scribd.com/document/531087231/Bank-Fraud-Detection-using-Community-Detection-Algorithm