U.PORTO
FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

FRAUD: THE RIPPLING EFFECTS OF HIGH-RISK NODES REMOVAL IN FINANCIAL NETWORKS

Daniel Tavares Ribeiro

Dissertation

Master in Economics and Business Administration

Supervised by
Pedro Campos, PhD
Jacopo Bono, PhD

2025

Acknowledgements

First and foremost, I would like to thank those who guided me throughout this journey. To Professor Pedro Campos, for his unwavering support, insightful guidance, and for providing a strong academic and economic perspective that shaped this work from start to finish. To Dr. Jacopo Bono, for his availability, and for enriching this thesis with valuable real-world insights drawn from his experience at Feedzai. His practical perspective greatly complemented the academic approach and helped bring this work closer to real-world relevance.

I would also like to express my gratitude to my parents and my sister for their unconditional support, encouragement, and for being the foundation that allows me to pursue my goals.

Finally, to my friends, thank you for your motivation, patience, and for always being there throughout this process.

Abstract

This study investigates the dynamics of financial fraud propagation and detection through the lens of high-risk node behaviour within financial transaction networks. Utilizing a synthetic yet statistically grounded dataset derived from the PaySim simulator, this work explores the effectiveness of various machine learning and graph-based models, including Isolation Forests, Random Forests, Graph Neural Networks (GNNs), and Graph Recurrent Neural Networks (GRNNs), in detecting fraudulent transactions. The research implements extensive feature engineering and data augmentation techniques, notably SMOTE and Self-Attention Generative Adversarial Networks (SAGAN), to address the inherent class imbalance of fraud datasets. A graph-based representation of transactional data enables the analysis of relational structures and temporal dynamics, essential for understanding systemic risk introduced by high-risk nodes. Furthermore, the study integrates agent-based simulations to evaluate the ripple effects of systematically removing high-risk agents from the network. Complementing these analyses, the research incorporates centrality metrics and community detection through the Louvain algorithm to investigate the structural properties of the network and assess the systemic impact of targeted interventions. The results demonstrate that the removal of structurally influential agents induces measurable shifts in network topology, including increased modularity, fragmentation into smaller communities, and redistribution of structural influence among remaining nodes. These structural transformations effectively constrain pathways for coordinated fraudulent activity while preserving detection system stability. Empirical results further demonstrate that ensemble models incorporating Isolation Forests, Random Forest, and GRNN classifiers outperform standalone models in both detection accuracy and economic cost reduction. The findings highlight the critical role of network topology, structural influence, and dynamic agent behaviour in shaping the effectiveness of fraud detection strategies, offering insights into both technical performance and systemic impact. Collectively, these results contribute to a more robust, structurally informed, and economically grounded framework for combating financial fraud in digital ecosystems.

Resumo

O presente estudo investiga a dinâmica da propagação e deteção de fraude financeira através da análise comportamental de nós de alto risco em redes de transações financeiras. Utilizando um conjunto de dados sintéticos, mas estatisticamente fundamentado, gerado pelo simulador PaySim, o trabalho avalia a eficácia de diversos modelos de aprendizagem e de redes neuronais baseadas em gráficos, incluindo Isolation Forests, Random Forests, Graph Neural Networks (GNNs) e Graph Recurrent Neural Networks (GRNNs), na deteção de transações fraudulentas. Técnicas avançadas de engenharia de variáveis e geração de dados, como o SMOTE e Self-Attention Generative Adversarial Networks (SAGAN), são aplicadas para mitigar o desequilíbrio das classes. A representação gráfica dos dados permite analisar as estruturas relacionais e as dinâmicas temporais, aspetos essenciais para compreender o risco sistémico introduzido pelos nós de alto risco. Adicionalmente, o estudo integra simulações baseadas em agentes para avaliar os efeitos da remoção sistemática de agentes de alto risco da rede. Complementando estas análises, o trabalho incorpora métricas de centralidade e deteção de comunidades através do algoritmo de Louvain, com o objetivo de investigar as propriedades estruturais da rede e avaliar o impacto sistémico de intervenções direcionadas. Os resultados demonstram que a remoção de agentes estruturalmente influentes induz alterações mensuráveis na topologia da rede, incluindo o aumento da modularidade, a fragmentação em comunidades mais pequenas e a redistribuição da influência estrutural entre os nós remanescentes. Estas transformações estruturais contribuem para limitar os caminhos disponíveis para atividades fraudulentas coordenadas, enquanto preservam a estabilidade do sistema de deteção.Os resultados empíricos demonstram ainda que os modelos de ensemble que combinam Isolation Forests, Random Forest e GRNNs superam consistentemente os modelos individuais, tanto em termos de precisão na deteção como de redução do custo económico associado à fraude. As conclusões realçam o papel crítico da topologia da rede, da influência estrutural e do comportamento dinâmico dos agentes na eficácia das estratégias de deteção de fraude, oferecendo contributos relevantes tanto para o desempenho técnico como para a compreensão do impacto sistémico. No seu conjunto, estes resultados contribuem para um enquadramento mais robusto, informado pela estrutura da rede e economicamente sustentado, para o combate à fraude financeira em ecossistemas digitais.

Keywords

Financial fraud; Fraud detection; Machine learning; Graph neural networks; Agent-based modelling; Transaction networks; Anomaly detection; Economic cost analysis; Credit rationing; Market inefficiencies; Financial regulation; Information asymmetry; Ensemble methods

Palavras-chave

Fraude financeira; Deteção de fraude; Aprendizagem automática; Redes neuronais em grafos; Modelação baseada em agentes; Redes de transações; Deteção de anomalias; Risco sistémico; Análise de custo económico; Racionamento de crédito; Ineficiências de mercado; Regulação financeira; Assimetria de informação; Métodos em ensemble

Index

Index of Tables

Index of Figures

1. Introduction

Fraud in financial transactions is a deliberate act of deception or manipulation that results in financial gain for perpetrators or harm to victims. Fraud's pervasive nature is underscored by its economic and systemic impact. In 2019, fraudulent transactions across the Single Euro Payments Area reached 1870 million euros, with the eurozone accounting for 1030 million euros of fraudulent card transactions (European Central Bank, 2021). More recent figures reveal a sharp rise in financial fraud worldwide, with the European Banking Authority reporting 4300 million euros in fraud losses for 2022 and 2000 million euros in the first half of 2023 alone (European Banking Authority, 2024). The Federal Trade Commission reported 8800 million dollars lost due to fraud in 2022, 2600 million more than the previous year. In 2023 the value increased again to over 10000 million dollars (Federal Trade Commission, 2024).

This rise in financial fraud is intricately linked to the structure and dynamics of financial networks. A financial network can be understood as a graph-based representation of financial systems, where nodes signify entities such as accounts, banks, or institutions, and edges represent transactional relationships (Brito et al., 2018a). Within these networks, certain nodes exhibit characteristics that mark them as high-risk nodes, entities significantly more likely to be involved in fraudulent or suspicious activities. These high-risk nodes often play a critical role in the dissemination of fraud, as their connections and behaviours can propagate risk throughout the network. The increasing volume of online transactions has amplified the role of social network interactions within financial networks. These interactions encompass the behavioural and relational dynamics among nodes, such as trust-based relationships, influence, and the pathways through which fraud spreads. For example, fraudulent schemes frequently exploit interconnected nodes to propagate deceptive activities, leveraging existing trust relationships to minimize detection. The degree of technical knowledge required to commit fraud is now minimal, as card and personal details can be purchased on the dark web or even downloaded for free. Data breaches are driving the supply and providing the source material for the growing number of fraud cases. In January 2024, the "mother of all data leaks" was made public, and 2600 million records containing personally identifiable information and credentials were leaked (Scroxton, 2024). Such events highlight the vulnerabilities inherent in financial networks and underscore the importance of how social network interactions and high-risk nodes contribute to systemic risks. Addressing these elements is critical

to developing more robust fraud detection and prevention strategies, ensuring the resilience of financial systems in the face of evolving threats.

Beyond its direct financial costs, fraud introduces significant market inefficiencies. It exacerbates adverse selection, where lenders and financial institutions, unable to perfectly distinguish between honest and dishonest actors, may reject creditworthy borrowers to avoid fraud-related losses. This leads to credit rationing and limits access to finance for legitimate market participants (Claus & Grimes, 2003). Moreover, moral hazard emerges as some entities may engage in riskier behaviour or fraudulent actions, knowing that the costs may be externalized or mitigated by weak enforcement. Such distortions hinder optimal allocation of resources, reduce market confidence, and increase transaction costs. From a broader economic perspective, financial fraud represents a clear market failure, justifying regulatory intervention and enhanced oversight. As Stiglitz and Weiss demonstrated in their seminal work on credit rationing, information asymmetries in financial markets can result in suboptimal lending outcomes (Stiglitz & Weiss, 1981). The risk of fraud adds another layer of asymmetry, intensifying credit rationing and excluding many potential borrowers from financial access. This not only limits economic growth but also undermines the stability and inclusiveness of financial systems.

Addressing these challenges requires a nuanced understanding of fraud's interaction with financial network structures and behaviours. Developing robust fraud detection and prevention strategies is essential to mitigate systemic risk and promote the resilience of financial systems amid evolving threats.

Building on the preceding discussion of financial fraud's pervasive economic impact and the critical role of network topology in propagating fraudulent activity, the primary aim of this thesis is to investigate the economic consequences of removing high-risk nodes from financial transaction networks and the resultant effect on overall fraud risk. Specifically, we analyse how the targeted intervention of deactivating accounts identified as high-risk alters both the incidence of fraud and economic indicators. This entails quantifying the trade-offs between mitigating fraud (through node removal) and potentially disrupting legitimate financial flows or market efficiency. By integrating a network-centric perspective with economic analysis, the research seeks to determine whether strategic node removals can lower fraud exposure without disproportionately impairing overall transaction efficiency or connectivity. To this end, we employ network structural metrics, including degree, betweenness, closeness, and eigenvector centralities, to quantify the positional influence of individual accounts within the

transaction graph. In addition, we apply the Louvain method to detect community modularity and capture latent clusters of coordinated transactional behaviour. These analytical components are essential not only for identifying high-risk nodes, but also for enabling an impact analysis that measures how their removal reshapes the broader network structure, with specific attention to shifts in centrality and community cohesion.

To address these aims, this study employs a mixed methodological framework. We use the PaySim synthetic transaction dataset, a well-established mobile-money simulator that generates realistic, agent-based financial transaction records, (E. Lopez-Rojas et al., 2016). The raw dataset is subjected to extensive feature engineering: for each account (node) and transaction edge, we compute behavioural features. Given the rarity of fraudulent events in real-world data, the class imbalance is severe. We mitigate this by applying two data augmentation techniques: the Synthetic Minority Over-Sampling Technique (SMOTE), which synthetically generates new minority class (fraud) samples via interpolation in feature space; and a Self-Attention Generative Adversarial Network (SAGAN), which leverages attention mechanisms to produce highly realistic synthetic transactions with long-range dependencies. Together, SMOTE and SAGAN enrich the fraudulent sample pool, improving the robustness of subsequent learning models. The processed data is then represented as a financial transaction network, where nodes correspond to accounts or customers and directed edges represent transactional flows between them. This graph-based formulation permits the use of both classical and modern machine-learning methods for fraud classification. First, we train traditional classifiers, notably a Random Forest and an Isolation Forest, on the engineered feature set. These models produce a fraud-risk score for each node based on its transaction patterns and features. In parallel, we deploy graph-based deep learning models: a Graph Neural Network (GNN) and a Graph Recurrent Neural Network (GRNN). GNNs can aggregate information from a node's neighbours to capture structural patterns, while GRNNs further incorporate temporal or sequential memory via hidden states passed across connected nodes. These graph models can directly learn from the transaction network structure and have shown promise in capturing inter-connected fraud patterns. Finally, to evaluate systemic effects, we integrate these classifiers into an agent-based simulation of the financial network. In the simulation, agents conduct transactions over time following realistic behaviour rules, similar in spirit to the PaySim approach. We then iteratively remove those nodes flagged as high-risk by our models, observing the ripple effects. The simulation allows us to track how

node removals propagate through the network: for example, we measure changes in total fraud occurrences, transaction volumes and economic metrics. Crucially, we also monitor the evolution of centrality distributions and modularity scores post-intervention to assess changes in network influence and community structure. This procedure assesses whether and how the intervention of removing high-risk accounts reduces fraud and at what cost.

The central research question addressed in this thesis is: How does the targeted removal of high-risk nodes affect the overall level of fraudulent activity and the distribution of risk within the financial network? Furthermore, what are the economic and structural consequences of implementing such a removal strategy?

This thesis combines approaches from fraud detection and network analysis to explore and evaluate intervention strategies within financial transaction systems. The remainder of the document is organized to support this inquiry. 2. Literature Review presents a review of the relevant literature on financial fraud detection, data augmentation for imbalanced datasets, and graph-based approaches to modelling fraud. 3. Methodology and Data introduces the PaySim dataset and details the feature engineering process alongside the implementation of SMOTE and SAGAN for enhancing data representativeness, describes the computation of network centrality metrics and Louvain modularity. Outlines the design, training, and evaluation of the classification models, with a focus on their ability to identify fraudulent nodes within the transaction network. Lately, describes the agent-based simulation framework to be applied. 4. Results presents the results of the classification models applied to the transaction network, evaluating their effectiveness in identifying high-risk nodes associated with fraudulent activity. The performance of each model is assessed using appropriate evaluation metrics, and the results are contextualized through comparison with findings from existing literature. This comparison highlights both the contributions and limitations of the proposed approach relative to established methods in the domain of fraud detection. In addition, this section reports the outcomes of the centrality and modularity analyses, demonstrating how targeted removal of structurally influential nodes reshapes the network's topology. 5. Discussion provides a critical discussion of the results, exploring their implications for fraud prevention in financial networks. This section examines potential causal mechanisms underlying the observed model behaviours, including the role of network structure, feature representations, and data augmentation strategies. Additionally, it considers the broader consequences of model-driven interventions, such as the removal of high-risk nodes, on systemic risk, transactional dynamics, and economic stability, as inferred from observed shifts in

centrality and modularity. Through this discussion, the study reflects on both the practical utility and ethical considerations of deploying such models in real-world financial systems. 6. Conclusion concludes the thesis by summarizing the key findings and their relevance to the broader field of fraud detection and network analysis. It also outlines the limitations of the current work and proposes directions for future research. These include refining the simulation environment, exploring alternative intervention strategies, and extending the methodology to other types of financial networks. This final section emphasizes the need for continued interdisciplinary research to enhance the effectiveness, fairness, and interpretability of fraud detection systems in increasingly complex financial ecosystems.

2. Literature Review

2.1 Fraud Overview

The initial challenge we are faced in financial fraud detection lies in defining what constitutes fraud. A comprehensive understanding of its nature, significance, and implications is crucial for effectively addressing this issue. Fraud can be defined as the deliberate act of deceiving an individual or entity to gain an unfair or unlawful advantage, often resulting in financial gain or personal benefit. Typically involves misrepresentation, deceit, or the intentional withholding of critical information to influence the decisions or actions of others.

According to Wells, fraud is described as a criminal act involving deception for personal gain. For an act to meet the legal criteria to be constituted as fraudulent, four critical elements must be present: (1) the existence of a false statement made by the perpetrator that influences the victim's decisions, (2) the perpetrator's knowledge of the statement's falsehood, indicating a deliberate intent to deceive, (3) the victim's reliance on the false statement, which prompts an action, and (4) the resulting harm to the victim, typically manifesting as financial loss, due to their reliance on the deceptive statement (Wells, 2014).

Building on Wells' definition, we can draw financial transaction fraud as the deliberate act of deception or manipulation within financial transactions, with the intent of obtaining unlawful gain or inflicting harm. Akers and Gissel further characterize it as the misrepresentation, exploitation, or unauthorized use of access that violates trust, ethical standards, and legal principles within financial systems (Akers & Gissel, 2006). Furthermore, Karpoff argues that fraud not only erodes the trust within financial institutions but also impacts directly the economic stability of individuals, emphasizing the critical need for robust preventive measures (Karpoff, 2021).

Understanding the motivations behind financial transaction fraud involves considering a range of economic, psychological, and systemic factors. The scientific community has long debated the primary motivations for committing fraud, with one of the most influential concepts in social science and criminology being the fraud triangle. This framework, based on Cressey's research, posits that three key conditions drive fraudulent behaviour: perceived pressure, the perceived opportunity to commit fraud, and the ability to rationalize the fraudulent act in a way that makes it seem acceptable (Cressey, 1953). In their study, Hogan and colleagues found substantial evidence supporting the validity of the fraud triangle (Hogan et al., 2008).

The fraud diamond, introduced by Wolfe and Hermanson, is a widely recognized extension of the fraud triangle that incorporates a fourth condition: capability. While the original three conditions, perceived pressure, perceived opportunity, and rationalization, are crucial, the authors argue that an individual's personal traits and abilities play a significant role in determining the likelihood of fraud. These traits include the individual's role within the organization, intellectual capacity, confidence, ability to coerce others, effective lying, and immunity to stress, all of which are key factors influencing the occurrence of fraudulent behaviour (Wolfe & Hermanson, 2004).

In 2019, Vousinas introduced the S.C.O.R.E. model, which identifies key drivers of fraud: Situational pressures, Consequential rewards, Opportunities, Rationalization, and Lack of enforcement. According to this model, fraudsters often justify their actions through rationalization, perceiving them as necessary or deserved, especially during periods of financial strain or personal difficulty (Vousinas, 2019).

Financial fraud encompasses a variety of forms, each employing distinct methods. Some of the most well-known types include:

- Identity theft and account takeover, which involve the acquisition of sensitive personal information. Common tactics include phishing, physical theft, and exploitation of data breaches. Unauthorized access to financial accounts often leads to illicit transactions or unauthorized modifications to account details (Castell, 2013).
- Payment fraud, which targets financial transactions, particularly credit card and check fraud. In these cases, compromised payment information, such as stolen credit card details or falsified bank information, is exploited to authorize fraudulent transactions. Both businesses and consumers suffer financial consequences from payment fraud, including chargebacks and loss of goods or services (Mangala & Soni, 2023).
- Investment fraud involves fraudulent schemes that deceive individuals by promising high returns with minimal risk. These schemes often rely on misrepresentation or manipulation to lure investors into making unwise financial decisions (Hilal et al., 2022).
- Fraudulent charities, where fraudsters exploit individuals' goodwill by soliciting donations for a purported cause. These charities may be entirely fictitious or involve fake campaigns misusing the names of reputable organizations or well-known causes. Donors who provide personal or financial information on fraudulent websites risk

exposure to identity theft or credit card fraud, as scammers often misuse the collected data for illicit activities (Cordery & Baskerville, 2011).

- Cybercrime is a broad term encompassing various online criminal activities. These include:
    - Malware, which involves the use of viruses or malicious programs to exploit vulnerabilities in a victim's computer system or device.
    - Cryptojacking, where a cybercriminal covertly uses a victim's system to mine cryptocurrency without their knowledge or consent.
    - Ransomware, which involves deploying software that locks a victim's files or devices, rendering them inaccessible, while demanding a ransom for their release (Minnaar, 2019).

Effectively combating financial transaction fraud requires a multi-layered approach that combines advanced technology, robust organizational safeguards, and strict adherence to regulatory compliance.

Fraud prevention and detection increasingly rely on multifaceted strategies that integrate technological innovation, comprehensive organizational training, and strict adherence to ethical standards. According to Ameyaw and co-workers., these strategies constitute the foundation of an effective anti-fraud framework. Technological approaches, including advanced analytics, artificial intelligence (AI), and machine learning algorithms, are pivotal in enhancing the ability to identify and mitigate fraudulent activities (Maxwell Nana Ameyaw et al., 2024). These tools enable organizations to detect anomalies, anticipate potential fraud risks, and implement proactive responses, thereby mitigating vulnerabilities within financial systems.

Moreover, organizational training and ethical leadership play a pivotal role in fostering anti-fraud awareness across all levels of an organization. By nurturing a culture of integrity and providing employees with the necessary knowledge and skills to identify and report fraudulent activities, organizations can significantly mitigate the risks associated with both internal and external fraud.

Equally significant is the role of financial compliance, which entails strict adherence to regulatory standards and the establishment of robust internal controls. Financial compliance serves as a cornerstone of corporate integrity, ensuring that organizations operate within legal and ethical boundaries. Compliance officers are integral to this process, as they enforce anti-fraud policies, conduct regular audits, and promote a culture of accountability within the

organization. Their efforts not only ensure compliance with legal requirements but also bolster organizational resilience against fraudulent activities.

Thus, a synergistic integration of technological advancements, compliance frameworks, and ethical practices, as emphasized by Ameyaw and colleagues, underscores the importance of a comprehensive strategy to protect financial integrity and strengthen trust in organizational operations (Maxwell Nana Ameyaw et al., 2024).

Economic agents often rely on pre-existing social networks to mitigate risk and uncertainty, preferring to transact with individuals they know personally or who come recommended by trusted associates. These social networks provide both ex-ante and ex-post advantages. Prior a transaction, such networks assist in reducing information asymmetry between purchasers and vendors. Subsequent to the transaction, they mitigate opportunistic behaviour by cultivating social accountability and enabling the enforcement of sanctions against misconduct. Paradoxically, these same social connections can also heighten the potential for deception, deviance, and unethical practices (Granovetter, 1985).

Beyond the foundational frameworks discussed above, contemporary fraud research has evolved to incorporate sophisticated theoretical models that address the increasingly complex and interconnected nature of modern financial systems. These advanced approaches recognize that fraud operates within dynamic, multi-layered environments where traditional detection methods may prove insufficient. The first major advancement in fraud theory comes through Contagion Modelling which fundamentally reconceptualizes fraud from isolated criminal acts to systemic phenomena that spread through financial networks like infectious diseases, (Haldane & May, 2011). This paradigm shift emerged from the observation that modern financial systems are so interconnected that fraudulent behaviour in one area inevitably affects others, creating cascading effects that traditional individualistic models cannot explain. Drawing parallels from epidemiological studies, this approach conceptualizes fraudulent behaviour as a contagious process that spreads through financial networks via various transmission mechanisms. Unlike traditional fraud models that focus on individual perpetrators, contagion models examine how fraudulent activities propagate across interconnected institutions, creating cascading effects that can destabilize entire financial ecosystems. The contagion process operates through multiple channels: direct financial linkages between institutions, shared technological infrastructures, behavioural mimicry among fraudsters, and the erosion of trust that facilitates subsequent fraudulent activities. This theoretical framework has proven particularly valuable in understanding how single fraudulent events, such

as major data breaches or institutional failures, can trigger widespread fraudulent activities across seemingly unrelated entities. By mapping these contagion pathways, financial institutions can identify critical vulnerabilities and implement targeted interventions to prevent fraud from spreading throughout their networks (Battiston et al., 2012, 2016; Haldane & May, 2011). Building upon the systemic understanding provided by contagion theory, Game-theoretic model, add a crucial layer of analysis by examining fraud as a dynamic strategic interaction between rational actors. While contagion theory explains how fraud spreads, game theory illuminates why it spreads and how different actors respond to each other's strategies. This approach recognizes that fraud prevention is essentially a dynamic game where fraudsters, potential victims, financial institutions, and regulatory authorities continuously adapt their strategies in response to their opponents' actions. The game-theoretic framework illuminates several critical aspects of fraud dynamics: the arms race between fraud detection technologies and fraudster adaptation, the optimal allocation of security resources across different vulnerability points, and the role of deterrence in shaping fraudulent behaviour. These models reveal that effective fraud prevention requires understanding not only what fraudsters are currently doing but also predicting how they will respond to new security measures. Furthermore, game theory helps explain counterintuitive phenomena in fraud prevention, such as why increasing security in one area may lead to fraud migration to other vulnerable sectors, or why publicly announcing certain security measures may actually encourage rather than deter fraudulent activity (Alpcan & Başar, 2010; Grossklags et al., 2008). The strategic interactions identified by game theory naturally lead to questions about how fraudsters successfully deceive their targets, which is where Fraud Signalling Theory provides essential insights. This framework addresses the fundamental information asymmetries that characterize fraudulent interactions, examining how deceptive actors communicate false information while maintaining an appearance of legitimacy. Building upon classical signalling theory, this approach reveals how fraudsters craft convincing facades and how potential victims, and detection systems can distinguish between genuine and deceptive signals. The theory recognizes that successful fraud requires sophisticated signalling strategies that exploit cognitive biases and institutional weaknesses. Modern fraudsters increasingly employ multi-channel signalling approaches, combining technological manipulation with social engineering techniques to create convincing narratives. Understanding these signalling mechanisms is crucial for developing advanced detection systems that can identify subtle inconsistencies and patterns that betray fraudulent intent. The theory also explores how legitimate actors can

employ counter-signalling strategies to distinguish themselves from fraudsters, creating a complex signalling equilibrium that constantly evolves as both sides adapt their strategies (Akerlof, 1970; Spence, 2002). The final piece of this theoretical puzzle comes from Network Disruption Theory, which synthesizes insights from the previous frameworks to focus on the organizational structures through which modern fraud operates (Albert et al., 2000a; Borgatti, 2006). This theory emerges from the recognition that contemporary fraud increasingly functions through complex networks that span multiple jurisdictions, technologies, and social systems, networks that are revealed through contagion analysis, maintained through strategic interactions, and operated through sophisticated signalling mechanisms. Network disruption theory focuses on identifying structural vulnerabilities within fraud networks and developing targeted intervention strategies that can effectively dismantle these operations. Drawing from network science, this approach reveals how fraud networks form, evolve, and maintain operations despite law enforcement pressure. Key insights include the identification of critical nodes whose removal can fragment entire fraud networks, the role of network redundancy in maintaining operational resilience, and the importance of weak ties in facilitating network expansion. Crucially, this theory also examines how fraud networks adapt their structures in response to enforcement actions, often becoming more decentralized and resilient over time. This evolutionary dynamic requires law enforcement and fraud prevention specialists to adopt adaptive strategies that target networks' fundamental structural weaknesses rather than merely pursuing individual actors.

Together, these four theoretical frameworks provide a comprehensive, multi-layered understanding of contemporary fraud that addresses its systemic nature (contagion), strategic complexity (game theory), communicative deception (signalling), and organizational structure (network disruption). This integrated approach recognizes that effective fraud prevention in the modern era requires understanding not just individual criminal behaviour, but the complex systems, interactions, and structures that enable fraud to flourish in interconnected financial environments.

2.2 Fraud Detection Approaches

The increasing sophistication of fraudulent activities requires the adoption of advanced detection methodologies that extend beyond traditional rule-based systems. While rule-based systems have historically played a foundational role in fraud detection, they are inherently static and unable to adapt to the evolving, multifaceted nature of contemporary fraud. As a

result, the field has witnessed significant advancements in detection methodologies, embracing more dynamic and adaptive approaches, including agent-based models (ABMs), machine learning, graph-based analysis, quantum computing, and generative models. These diverse methodologies provide a more comprehensive and adaptable framework for tackling fraud across a variety of domains, such as financial transactions, e-commerce, social networks, and beyond.

The work of Koehler and colleagues illustrates the application of ABMs in generating large-scale, anonymized financial datasets to study fraud patterns (Koehler et al., 2005). Their use of NetLogo facilitated the simulation of heterogeneous financial networks, enabling the modelling of dynamic, high-volume transactions while preserving data privacy. Lopez-Rojas and Axelsson advanced the use of ABMs through tools like PaySim, RetSim, and BankSim, which generate realistic synthetic transaction data for fraud detection in banking, retail, and payment systems (E. A. Lopez-Rojas & Axelsson, 2014). These simulators, calibrated with social network analysis, offer a privacy-preserving solution to test fraud detection techniques, though the limited availability of real-world calibration data remains a challenge. Kleiner and co-workers developed an ABM to examine the interplay between economic agents and their institutional environment. The study emphasizes how agents' mentalities are influenced by societal institutions and, conversely, how these mental values can transform the institutional landscape (Kleiner et al., 2023). The BOND model introduced by Brito and colleagues, which utilizes ABMs to simulate the interactions of various agents and detect fraudulent patterns, incorporates behavioural tendencies and agent profiles (Brito et al., 2018). Behaviours can effectively be identified, such as fund dispersion through small transactions, a common tactic in money laundering. Similarly, Fontes and co-workers further expanded the utility of ABMs by integrating them with graph-based motif analysis, using entity graphs for entity interactions and transaction graphs for temporal transaction relationships, which allowed for the identification of overrepresented and underrepresented patterns in financial transactions. By focusing on network motifs, this combined approach provided structural insights that enhanced the identification of fraudulent activities (Fontes et al., 2021).

Leite and colleagues advanced the use of ABMs by incorporating reinforcement learning to train agents to adapt to evolving fraud tactics. Their multi-agent simulation framework demonstrated the versatility of ABMs in capturing dynamic fraud strategies, including changing transaction patterns in response to detection mechanisms. Complementing this, they employed a Visual Analytics (VA) approach combining automated machine learning with

human-in-the-loop reasoning to fine-tune detection algorithms, balance sensitivity and false alarms, and identify fraudulent relationships. By addressing fraud types such as unauthorized transactions, money laundering, and fake users, this integrated approach enhances fraud detection in scenarios where fraudsters continually adapt their tactics (Leite et al., 2018). Extending this framework, Sandeep et al. introduced a multilayer social network model that combined Soft Sign Activated Graph Autoencoders (SS-GAE) with Pareto Front-based K-Means Clustering. This hybrid model effectively integrated ABM principles with graph-based learning, significantly improving fraud detection across interconnected social networks by ensuring scalability and precision (Sandeep et al., 2024).

In more specialized domains, ABMs have also been applied to real-time payment systems. Mayo and colleagues developed an ABM that incorporated game-theoretic principles to model interactions between banks and fraudsters, emphasizing the strategic use of real-time payment threshold controls. Their model demonstrated how these controls could effectively mitigate fraud while minimizing customer disruptions (Mayo et al., 2024). Similarly, Yamin and co-workers applied ABMs to the detection of electoral fraud, showcasing the versatility of ABMs in simulating complex fraud scenarios that extend beyond financial contexts (Yamin et al., 2023).

Machine learning has become a cornerstone in modern fraud detection due to its ability to provide dynamic, data-driven solutions that evolve with the sophistication of fraudulent activities. Traditional techniques such as Support Vector Machines (SVMs) have shown considerable effectiveness, particularly when combined with other models. For instance, Udayakumar and colleagues proposed a hybrid model that integrates SVM with Feedforward Neural Networks (FFNN), achieving superior performance in fraud detection by utilizing data segmentation and nonlinear pattern classification (Udayakumar et al., 2023). Similarly, Rtayli and Enneya enhanced SVMs by incorporating Recursive Feature Elimination (RFE) and Synthetic Minority Over-Sampling Techniques (SMOTE), addressing the issue of imbalanced datasets that are common in financial fraud detection (Rtayli & Enneya, 2020).

Beyond classical machine learning models, ensemble methods such as Random Forests (RF) have demonstrated their robustness and adaptability to diverse datasets in fraud detection. Saddam Hussain and co-workers compared RF with other models and found it to provide superior accuracy in detecting fraud across various contexts (Saddam Hussain et al., 2021). Moreover, methods like SMOTE have been integrated to optimize the performance of RF, particularly in handling imbalanced data scenarios (Aburbeian & Ashqar, 2023). Other

ensemble techniques, such as the hyper-ensemble model proposed by Vashistha and Tiwari, combine multiple classifiers to achieve state-of-the-art precision, recall, and AUC-ROC scores, further advancing the effectiveness of fraud detection systems (Vashistha & Tiwari, 2024).

The role of deep learning in fraud detection has also expanded significantly. Rzayeva and Malekzadeh introduced a hybrid model that combines Deep Neural Networks (DNN) with K-Nearest Neighbours (KNN), a strategy that addresses class imbalances and concept drift in dynamic fraud detection environments ((Rzayeva & Malekzadeh, 2022). Louati and Ktata further advanced this by integrating autoencoders with multi-agent systems (MAS), which improved classification accuracy and data dimensionality reduction. Their system demonstrated particular efficiency and scalability in detecting intrusions in networked environments (Louati & Ktata, 2020). Wang and Zhu enhance behaviour-based fraud detection by fine-grained modelling of co-occurrence relationships in transactional data, utilizing network embedding techniques on heterogeneous relation networks. The approach, employs heterogeneous relation networks and a MAS behavioural model, addresses challenges in incomplete or sparse data (C. Wang & Zhu, 2022).

The study of fraud detection has increasingly leveraged graph-based techniques and social network analysis (SNA) to uncover complex patterns of fraudulent behaviour. By analysing the structural and dynamic properties of transactional and social networks, SNA uncovers patterns of interaction, influence, and information flow in systems ranging from social media to organizational networks. Building on this foundation Tian and colleagues developed the Adaptive Sampling and Aggregation-based Graph Neural Network (ASA-GNN), which improves the accuracy and robustness of fraud detection by focusing on meaningful patterns while filtering out noisy data. This method was further refined with the introduction of the Spatial-Temporal-Aware Graph Transformer (STA-GT), which incorporated temporal dynamics to better capture evolving fraudulent behaviours (Tian et al., 2023; Tian & Liu, 2024). In a similar vein, the Context Correlation Discrepancy Analysis framework by Wang and coworkers focused on identifying anomalies in financial transaction graphs by examining contextual relationships, achieving high precision and recall in detecting subtle fraudulent patterns (R. Wang et al., 2024).

Generative models, particularly in the form of Self-Attention Generative Adversarial Networks (SAGANs), have shown significant promise in enhancing fraud detection systems by simulating both normal and fraudulent behaviours. Zhao and colleagues employed SAGANs

to generate synthetic data that closely reflects real-world patterns, improving adaptability to evolving fraud tactics and reducing false positives (Zhao et al., 2024).

The application of quantum computing to fraud detection has introduced transformative capabilities, particularly in the processing of complex data structures. Innan and co-workers pioneered the use of Quantum Graph Neural Networks (QGNNs) in fraud detection, leveraging Variational Quantum Circuits to surpass the performance of classical neural networks. Their approach showcased significant improvements in both efficiency and accuracy, pointing to the potential of quantum computing to revolutionize fraud detection systems (Innan et al., 2024).

Dwivedi and Tajer introduced a Graph Recurrent Neural Network (GRNN) framework specifically designed for real-time prediction of cascading fault chains in power systems (Dwivedi & Tajer, 2024). This model effectively captured both spatial dependencies and temporal dynamics, by addressing the limitations of traditional GNNs, which, while adept at modelling spatial relationships within graph-structured data, often lack the capacity to account for temporal sequences inherent in dynamic systems. By integrating recurrent neural network components into the GNN architecture, the GRNN provided a unified framework for spatio-temporal data processing. This integration enabled the model to maintain a memory of past states, a feature critical for accurately modelling and forecasting the progression of faults over time. Further supporting the efficacy of GRNNs, Liu demonstrated that the model exhibits rapid learning and strong generalization capabilities, particularly in scenarios characterized by limited or imbalanced datasets, a common challenge in real-world (Y. Liu, 2024). Additionally, Rossi and colleagues developed an alternative GRNN based approach that yields significant performance gains over conventional GNN models, not only in predictive accuracy but also in computational efficiency (Rossi et al., 2020).

Recent advances in graph-based fraud detection have integrated structural centrality measures and community detection to enhance risk profiling and uncover coordinated illicit behaviour. For example, Bonato and colleagues utilize degree and eigenvector centralities to isolate influential or exposed banking nodes, followed by Louvain-based community partitioning to identify clusters associated with suspicious transaction patterns (Bonato et al., 2025). Similarly, Que and colleagues compare Graph Convolutional Network performance with unsupervised Louvain label propagation, demonstrating that community detection effectively reveals fraud-linked subgraphs and closely aligns with centrality-informed risk assessments (Que et al., 2015).

Complementing these findings, Chang and peers propose a centrality-guided deep random walk framework to identify key orchestrators within telecom fraud networks, showing that nodes with high betweenness or eigenvector centrality disproportionately facilitate illicit activity propagation (Chang et al., 2021). Likewise, Zanin and colleagues introduce a network approach to credit card fraud detection, leveraging structural network features, including centrality-based indicators, to significantly improve classification performance (Zanin et al., 2017). Collectively, these studies validate the dual utility of centrality metrics and Louvain modularity in both identifying high-risk entities and mapping fraudulent communities, providing robust justification for our methodological framework, which integrates these metrics into both node-level risk profiling and structural impact analysis.

Kottrachai and colleagues provide a comprehensive analysis of using explainable AI (XAI) techniques to enhance the interpretability of credit card fraud detection models (Kotrachai et al., 2023). The work focuses on leveraging SHapley Additive exPlanations (SHAP) to explain model predictions, enabling the understanding of the contribution of individual features to decisions. The authors underscore the critical role of explainability in fostering trust and transparency within financial decision-making systems.

In summary, the methodologies reviewed in this study each present unique strengths. ABMs offer a comprehensive macro-level perspective on systemic behaviours, while machine learning and graph-based techniques demonstrate exceptional capability in analysing transaction-specific data. Generative and quantum models provide advanced adaptability and computational efficiency. A holistic fraud detection ecosystem must integrate these methodologies into unified frameworks to represent a promising direction, by addressing the limitations of individual techniques and enhancing the overall robustness of fraud detection systems. By combining these advanced methodologies, future fraud detection systems can more effectively address the systemic, economic, and transactional dimensions of fraud, ensuring resilience against increasingly sophisticated tactics.

On *Table 1* we provide a concise summary of most pertinent studies discussed previously, presenting an overview of the key aspects analysed, including the methods employed for feature selection, classification techniques, datasets utilized, evaluation methodologies, and their respective metrics and final performance scores. This tabular representation offers a streamlined comparison, enabling a clearer understanding of the approaches and outcomes across the reviewed research.

Table 1 Overview of reviewed studies: Methodologies, Datasets, Research Objectives and Final Scores.

Authors	Methodology	Objective	Findings	Dataset
Brito et al., 2018	Employed Backward Feature Elimination and Sliding Window Aggregation alongside Machine Learning techniques (RF, Neural Networks, Bayesian Networks).	To develop an agent-based framework for detecting fraudulent transactions in economic networks.	Demonstrated the superior accuracy and stability of Random Forests. Proposed further refinements to address false negatives and real-time applicability.	Synthetic dataset generated via the BOND model.
Fontes et al., 2021	Utilized graph-based representations with a focus on heterogeneous network motif analysis.	To identify patterns in fraudulent banking transactions through motif-based graph analysis.	Provided critical insights into fraud patterns, identifying overrepresented and underrepresented motifs, which offer a nuanced understanding of fraudulent behaviour.	Real-world card transaction dataset from a banking network.
Rtayli & Enneya, 2020	Leveraged RFE and SVM with SMOTE for data balancing.	To enhance fraud detection performance via advanced feature selection and hyperparameter optimization.	Achieved high accuracy and reduced false positives, underscoring the robustness of the proposed methodology.	Real and synthetic card transaction datasets.
Saddam Hussain et al., 2021	Investigated RF in comparison with SVM and Decision Tree classifiers.	To evaluate the performance of machine learning algorithms in detecting financial fraud.	Highlighted the superior performance of Random Forests in terms of accuracy and reliability.	Financial transactions dataset (synthetic or real, not disclosed).
Vashistha & Tiwari, 2024	Introduced a hyper-ensemble model combining various classifiers, including Neural Networks and Decision Trees.	To enhance precision, recall, and adaptability in fraud detection systems.	Showcased superior performance metrics and computational efficiency, particularly in handling emerging fraud types.	Kaggle Bank Account Fraud Dataset Suite.

Tian et al., 2023	Developed ASA-GNN.	To enhance fraud detection through graph-based models incorporating sampling techniques.	Demonstrated improvements in accuracy and efficiency compared to traditional GNN approaches.	Financial datasets from Chinese financial institutions.
Tian & Liu, 2024	Introduced STA-GT.	To incorporate spatial and temporal dynamics into fraud detection using advanced GNNs.	Achieved high accuracy in detecting fraud patterns involving complex spatial-temporal relations.	Financial datasets from Chinese financial institutions
R. Wang et al., 2024	Proposed Context Correlation Discrepancy Analysis using graph neural networks.	To enhance anomaly detection by analysing contextual discrepancies in financial transaction graphs.	Achieved high accuracy and contextual relevance, outperforming traditional anomaly detection methods.	Financial transaction graphs.
Zhao et al., 2024	Leveraged SAGAN.	To reduce false positives and enhance adaptability to evolving fraud patterns.	Demonstrated superior performance compared to traditional models in real-time fraud detection scenarios.	Credit card transaction datasets.

3. Methodology and Data

This chapter outlines the methodological framework employed to investigate the economic and systemic effects of removing high-risk nodes from financial transaction networks. The central objective is to evaluate how such interventions influence the distribution of fraudulent activity and associated economic costs within the network. To achieve this, a multi-layered methodology is adopted, combining synthetic transaction data, advanced feature engineering, data augmentation techniques, machine learning classifiers, and agent-based simulation modelling. The overarching goal is to develop and evaluate a suite of predictive models capable of identifying high-risk entities and to simulate the systemic consequences of their removal. This approach enables the analysis of not only model accuracy in fraud detection but also the broader implications for financial network structure and transaction integrity. The methodological pipeline begins with the construction and processing of a realistic transaction dataset, which is described in the following section. A high-level visual representation of the methodology employed in this study is presented in Figure 1.

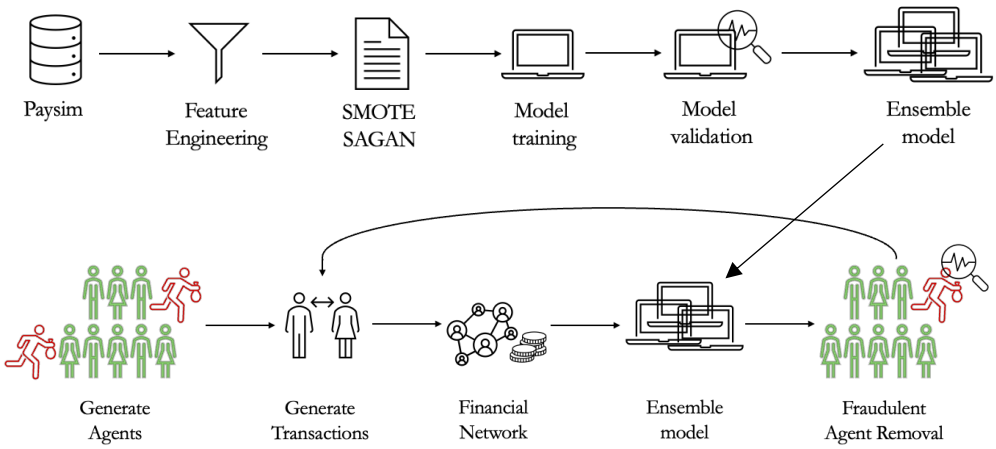

Figure 1 Diagram of development of the ensemble model and fraudulent agent removal workflow.

3.1 Dataset Overview

The PaySim dataset is a synthetic financial dataset generated by the PaySim mobile money simulator (E. Lopez-Rojas et al., 2016). The primary objective behind its creation is to provide a realistic yet publicly accessible resource for research on financial fraud detection. Given the highly sensitive nature of actual financial transaction data, the simulator leverages

aggregated real-world transactional logs to produce a dataset that mirrors genuine mobile money transactions, including the injection of fraudulent behaviour. This enables the development, test, and comparison of various fraud detection methodologies without compromising private financial information.

PaySim employs an agent-based simulation framework that is calibrated using sample aggregated financial logs from a mobile money service provider. The original data, which spans one month of real transactions from an operator in an African country, serves as the statistical basis for the simulation. In constructing the dataset, the simulator replicates normal transactional behaviour while deliberately integrating anomalous activity to mimic common fraudulent schemes such as unauthorized fund transfers and account takeovers. The simulation is organized in discrete time steps, with one step corresponding to one hour in real time. Over a simulated period of 30 days, the dataset therefore comprises 744 time-steps. This temporal granularity is beneficial for studying the evolution and dynamics of fraud, as it allows to explore how fraudulent patterns emerge and propagate over continuous intervals.

The dataset consists of 11 attributes that collectively capture key aspects of mobile money transactions. Each record represents an individual transaction and includes the following fields:

- **step**: An integer value indicating the time step at which the transaction occurred. One unit corresponds to one hour.
- **type**: A categorical variable denoting the transaction type. The dataset includes five distinct types: *CASH-IN*, *CASH-OUT*, *DEBIT*, *PAYMENT*, and *TRANSFER*. These categories reflect different modes of monetary transfer.
- **amount**: A numerical value that represents the amount of the transaction in the local currency.
- **nameOrig**: The unique identifier for the originating (sender) account. This identifier typically follows a specific format that may hint at the type of account (e.g., clients versus merchants).
- **oldbalanceOrg**: The balance of the originating account prior to the transaction.
- **newbalanceOrig**: The balance of the originating account after the transaction.
- **nameDest**: The unique identifier for the destination (recipient) account. For certain account types, such as merchant accounts (which often begin with a specific prefix, detailed balance information may not be available.
- **oldbalanceDest**: The initial balance of the destination account before receiving funds.

- **newbalanceDest**: The balance of the destination account following the transaction.
- **isFraud**: A binary indicator where a value of 1 represents a transaction that has been perpetrated by a fraudulent actor, while a value of 0 indicates a non-fraudulent transaction.

By embedding realistic transactional patterns and artificially injected fraudulent behaviour within a controlled environment, the PaySim dataset addresses a critical need in the domain of fraud detection research.

3.2 Feature Engineering

To enhance fraud detection performance on the PaySim dataset, a number of new features were engineered to capture transaction dynamics and anomalies not evident in the raw data. Building on these, the engineered features were designed to highlight unusual transaction patterns that may indicate fraudulent behaviour.

The first derived features quantify how each account's balance changes as a result of the transaction. Specifically, *balance_diff_org* was defined as the difference between the origin account's post-transaction and pre-transaction balances:

$$\Delta\text{balance}_{\text{orig}} = \text{newbalanceOrig} - \text{oldbalanceOrg}$$

Likewise, *balance_diff_dest* is the change in the destination account's balance. These differences directly measure the net effect of the transfer on each account. In legitimate transactions, the origin's balance decreases by the transaction amount while the destination's balance increases by the same amount. By contrast, fraudulent transactions can produce atypical imbalances. For instance, a fraudster withdrawing all funds would yield a large negative $\Delta\text{balance}_{orig}$, and an illicit deposit would yield a large positive $\Delta\text{balance}_{orig}$. Such extreme imbalances are uncommon in normal behaviour, making these difference features useful for detecting anomalies (Hilal et al., 2022).

To account for the scale of each account, relative balance change features were also computed. For each account, the relative change was defined as the balance difference divided by the prior balance. Formally, for the origin account:

$$\text{rel\_balance\_change\_org} = \frac{\text{newbalanceOrig} - \text{oldbalanceOrg}}{\text{oldbalanceOrg}}$$

with an analogous definition for *rel_balance_change_dest*. This normalization makes a fixed absolute change more significant when the initial balance is small. For example, a 1000€ transfer is far more impactful if the origin balance was only 100€. Large fractional changes (such as draining nearly all of an account's funds) are often much riskier than equivalent absolute

changes on larger accounts. These relative features therefore amplify cases where a transaction represents an unusually high proportion of the account's funds, helping to distinguish high-impact transactions common in fraud from routine transfers (Hilal et al., 2022).

Next, ratios of the transaction amount to account balances were derived.

$$\text{amount\_to\_oldbalance\_org} = \frac{\text{amount}}{\text{oldbalanceOrg}}$$

and analogously for the destination account, *amount_to_oldbalance_dest*. This ratio expresses the size of the transfer relative to the origin or destination's previous balance. A ratio near 1 (or higher) indicates the transaction size rivals or exceeds the account's funds, which is atypical under normal circumstances. For example, an amount-to-balance ratio close to 100% could indicate an attempt to empty an account in a single transaction. By providing this contextual ratio, the model can weigh the significance of each transaction relative to the account's capacity rather than relying on absolute amounts alone.

Because transaction amounts and balances can span several orders of magnitude, logarithmic transformations were applied to reduce skew and mitigate the influence of extreme values. Specifically, for each of amount*, oldbalanceOrg, newbalanceOrig, oldbalanceDest,* and *newbalanceDest*, $log(1 + x)$ was computed, yielding features like *log_amount, log_oldbalanceOrg*, etc. Formally:

$$log\_amount = log(1 + amount)$$

with analogous definitions for each balance. The $log(1 + x)$, transformation compresses large values, making their distribution more symmetric. This has two main advantages, it diminishes the effect of very large transactions, so they do not dominate the model, while still preserving the order of magnitudes among smaller values. In a fraud context, this means the model can learn patterns across both small and large transactions without being biased towards extreme cases.

Interaction features were introduced to capture multiplicative relationships between variables. For example, the feature *amount_x_balance_diff_org* is defined as:

$$\text{amount\_x\_balance\_diff\_org} = \text{amount} \times (\text{newbalanceOrig} - \text{oldbalanceOrg})$$

By combining the transaction amount with the origin balance change, this feature highlights instances where a large transfer causes a large imbalance. A high value may indicate a highly unusual transaction (for example, a large withdrawal from a nearly empty account). Similarly, *amount_x_balance_diff_dest* is the product of the amount with the destination's balance change. An interaction of log-transformed values was also included: for instance:

$$\text{log\_amount\_x\_log\_balanceOrg} = \log(\text{amount}+1) \times \log(\text{oldbalanceOrg}+1)$$

This captures multiplicative effects on a compressed scale. Such interaction terms allow the model to detect cases where high values of amount and balance jointly amplify risk more than either factor alone.

Polynomial features of degree two were included to capture non-linear effects.

$$\text{amount\_squared} = \text{amount}^2$$

$$\text{balance\_diff\_org\_squared} = \text{oldbalanceOrg}^2$$

Squaring the amount exaggerates the influence of very large transactions: while a modest transfer has a moderate squared value, an enormous transfer yields a very large squared term, signalling disproportionate risk. Similarly, squaring the balance change emphasizes dramatic shifts more heavily. Including these polynomial terms allows the model to capture situations where risk accelerates as values grow.

Binary flags were also designed to highlight special cases. *flag_zero_oldbalance_org* is set to 1 if the origin account's pre-transaction balance was zero (and 0 otherwise), and *flag_zero_oldbalance_dest* does the same for the destination account. These flags capture abnormal scenarios: for example, if *oldbalanceOrg* is equal to 0 but a positive amount is withdrawn, it implies money appearing without a source, a situation often characteristic of fraudulent behaviour. Conversely, an account with zero balance receiving funds may also be notable. These flags explicitly highlight such edge cases to the model.

In addition to these numerical features, temporal and categorical context features were added. The hour of day was extracted from the transaction time by computing:

$$\text{hour\_of\_day} = \text{step} \bmod 24$$

Yielding an integer from 0 to 23. Fraudsters may operate at unusual times (for example, late at night or early morning) compared to normal users, so including the time of day helps exploit such diurnal patterns. Transaction direction was also encoded, for example, by distinguishing transfers between customer accounts and payments to merchant accounts. The transaction direction feature thus captures different typologies of transfers and allows the model to differentiate behaviour across these classes.

After engineering all candidate features, a correlation analysis was conducted exclusively on the training dataset to identify the most predictive variables while minimizing redundancy. Restricting this analysis to the training data helps prevent information leakage from the test set, thereby preserving the integrity of the evaluation process. The primary objective of this

step was to isolate features that exhibit strong associations with fraudulent activity, ensuring that only the most informative variables were retained for model development.

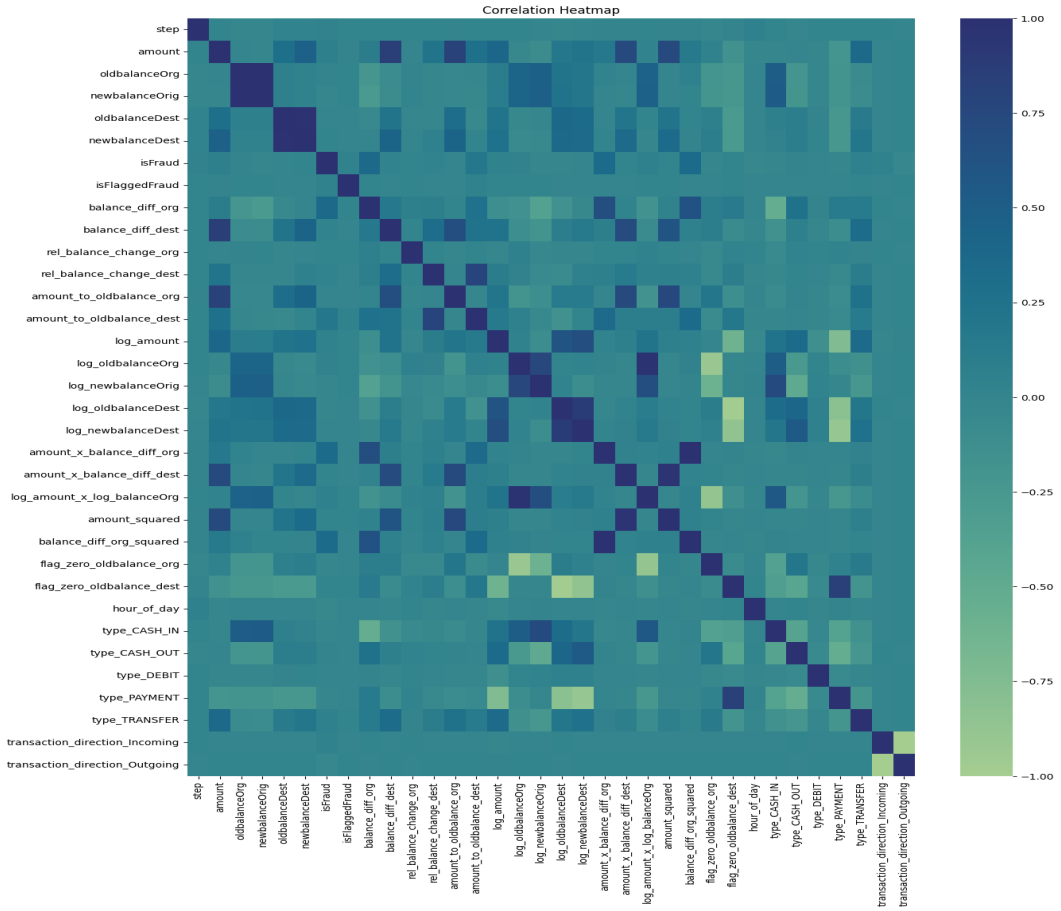

Figure 2 Correlation heatmap of all available features

From the analysis of Figure 2, the most relevant predictors included both original and engineered features: *amount, step, isFraud, amount_x_balance_diff_org, amount_to_oldbalance_dest,* and *type_TRANSFER*. Additionally, the identifiers *nameOrig* and *nameDest* were retained due to their relevance in distinguishing agents involved in transactions, which is critical for entity-level fraud detection. These selected features combine the raw data attributes (account IDs, transaction time and amount) with the engineered measures most indicative of fraud. Together, these features encompass the key aspects of transaction behaviour that distinguish fraudulent from legitimate activity.

3.3 Data Augmentation

Financial fraud detection datasets are notoriously imbalanced, since genuine transactions greatly outnumber fraud cases. The PaySim dataset is no exception as observed in Figure 3,

in a total of 6362620 transactions, only 8213 are labelled as fraudulent. Standard classifiers trained on such data tend to be biased toward the majority class, yielding poor detection of the minority fraud cases. To mitigate this, SMOTE was employed. It generates new synthetic minority-class examples by interpolating between existing fraud cases in feature space. In practice SMOTE was applied until the fraud/non-fraud ratio reached 0.5 (i.e. fraudulent examples became half as numerous as genuine ones). This intermediate target provides substantially more fraud samples for training while avoiding the excess redundancy of full class equalization. Note that oversampling to a full 1:1 ratio would simply replicate known patterns and risks overfitting the minority class.

The SMOTE process was implemented with care to preserve account identifiers. Since account IDs (*nameOrig*, *nameDest*) are nominal labels, they were excluded from the interpolation. SMOTE was applied only to the numerical transaction fields (such as amount and balance differences) and to the class label. As a result, each synthetic transaction inherits realistic numeric feature values without generating new arbitrary IDs. After oversampling, the original account ID columns were reattached to the augmented records. In this way no fictitious accounts are introduced: every synthetic transaction still links existing origin and destination IDs. Consequently, the augmented dataset maintains a consistent set of nodes, and each synthetic transaction remains anchored to existing account identities. However, while this approach preserves the identity of the nodes, it introduces a significant increase in the number of edges associated with fraudulent activity. This alters the structure of the transaction graph, particularly the distribution and density of fraudulent interactions, potentially impacting downstream graph-based models. Thus, although the node space remains stable, the edge-level augmentation introduced by SMOTE substantively modifies the topological and statistical properties of the graph.

Following the oversampling procedure, a graph-based generative model was employed to further enhance the minority class representations. Specifically, a SAGAN was utilized due to its capacity to generate high-fidelity, structurally coherent data by modelling complex dependencies across features (Zhao et al., 2024). SAGAN leverages attention mechanisms within both the generator and the discriminator, enabling the model to selectively focus on informative feature interactions during synthetic data generation. In this architecture, the generator is responsible for producing realistic synthetic samples by learning the underlying distribution of the minority class, while the discriminator evaluates the authenticity of the generated samples by distinguishing them from real data. Through adversarial training, these

two components iteratively refine their performance, resulting in high-fidelity synthetic instances that capture complex dependencies within the data. To integrate the synthetic outputs from SAGAN while preserving transaction integrity, the generation process was constrained to numeric transaction fields (e.g., transaction amount, balance updates), analogous to the SMOTE configuration. Nominal fields, particularly account identifiers, were reintroduced post-generation by associating each synthetic transaction with valid existing *nameOrig* and *nameDest* entries drawn from the fraud labelled dataset. This ensures that no artificial entities are introduced into the transaction graph. The combined use of SMOTE and SA-GAN thus produces a diversified yet coherent set of synthetic fraudulent transactions, contributing to a more balanced and expressive training dataset.

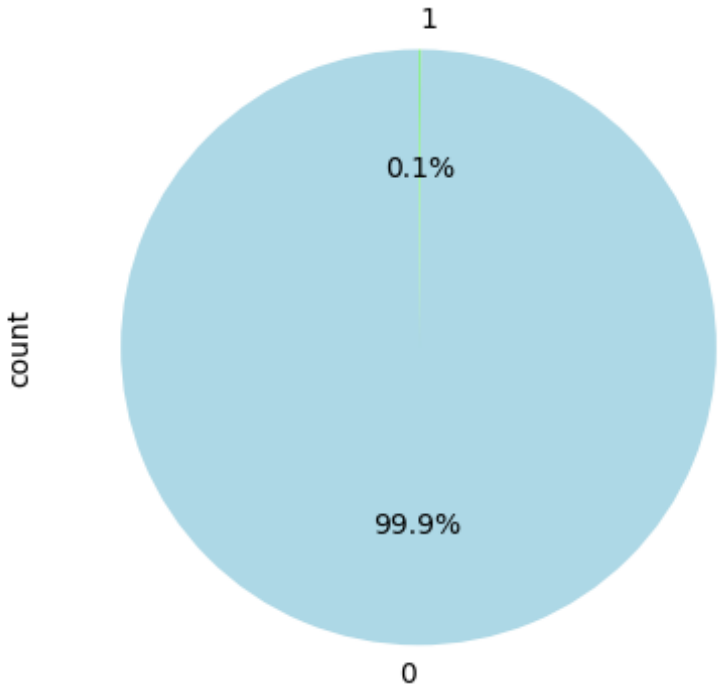

Figure 3 Distribution of fraudulent and non-fraudulent transactions

3.4 Creation of Financial Transaction Networks

With the balanced dataset prepared, a transaction graph was built to encode relational information for the GNN and GRNN. In this graph, each node represents a unique account, and each directed edge represents a transaction from the origin account to the destination account. Edge attributes were drawn from the transaction features: *amount; step; amount_x_balance_diff_org; amount_to_oldbalance_dest* and *type_TRANSFER*. The fraud label and account

IDs were defined as feature indicators on the edge.aiagentplatformaaa Node features were constructed by aggregating each account's transactional activity. For each node (account) statistics were computed from its incident edges to summarize its behaviour. Concretely, the total outgoing amount (sum of all amounts sent), the total incoming amount (sum received), and the counts of outgoing and incoming transactions. These aggregated measures encode an account's volume and activity frequency.

3.4.1 Network Structural Analysis

To deepen the structural understanding of the financial transaction network, we compute several node-level centrality metrics, degree, betweenness, closeness, and eigenvector (Peng et al., 2018), as well as a community-level measure: the Louvain modularity index (Blondel et al., 2008a). These measures highlight the roles and influence of accounts within transactional pathways and identify sub-structures of coordinated behaviour.

Degree Centrality: Quantifies the number of direct connections (edges) incident on each node v:

$$C_D(v) = k_v$$

where $k_v$is the degree of v, representing its immediate transactional activity. High-degree nodes often correspond to accounts with extensive direct interactions, which may serve as key transaction hubs (Peng et al., 2018).

Betweenness Centrality: Measures the extent to which a node lies on shortest paths between other node pairs:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from s to t, and $\sigma_{st}(v)$ counts those passing through v. Nodes with high betweenness serve as critical "bridges" in transaction flows and may control pathways susceptible to fraud propagation (Freeman, 1977; Holme et al., 2002).

Closeness Centrality: Indicates how efficiently a node can reach all other nodes via shortest paths:

$$C_C(v) = \frac{1}{\sum_{t \in V} d(v,t)}$$

where $d(v,t)$ denotes the shortest-path distance from v to t. Nodes with higher closeness are centrally positioned, enabling rapid dissemination or acquisition of information or funds (Pathan & Shrivastava, 2021).

Eigenvector Centrality: Assigns relative scores reflecting both node connectivity and the influence of its neighbours:

$$C_E(v) = \frac{1}{\lambda}\sum_{u \in V} A_{vu} C_E(u)$$

where A is the adjacency matrix and λ is the principal eigenvalue. Nodes connected to other prominent nodes achieve higher scores, indicating strategic importance in network influence (Peng et al., 2018).

Modularity & Louvain Method: Quantifies how well the network partitions into tight-knit communities. The algorithm, rapidly detects hierarchical communities by iteratively optimizing modularity and aggregating nodes:

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j)$$

Here, m denotes the total number of edges, k_i and $k_j$are the degrees of nodes i and j, and $\delta(c_i, c_j)$ equals 1 if nodes share a community and 0 otherwise. High modularity scores suggest the presence of tightly knit sub-groups, potentially indicating collusive or coordinated fraudulent clusters.

3.5 Training and Classification Models

In the context of fraud detection, the effectiveness of a detection system heavily depends on the choice and performance of the underlying classification models. This section outlines the machine learning and deep learning models employed for training and classification in the proposed system. The selected models represent a diverse spectrum of algorithmic paradigms, including ensemble-based learning, anomaly detection, and graph-based neural architectures. Specifically, this work explores an ensemble approach that integrates Random Forests, Isolation Forests, GNN, and GRNN. By evaluating various combinations of these models, the aim is to identify the most effective ensemble configuration for detecting fraudulent behaviour. Each model is trained on pre-processed transactional data to capture complex patterns indicative of fraudulent activity. The ensemble strategy is designed to leverage the strengths of each model type, enhancing overall robustness and accuracy. The following

subsections provide detailed descriptions of each classifier, including their theoretical foundations, implementation specifics, and relevance to fraud detection as well as the evaluation metrics to be used. To support this modelling approach, the transactional dataset was partitioned using a deterministic chronological splitting strategy, preserving the temporal dependencies inherent in real-world financial systems. The dataset was divided into training and validation subsets using an 80:20 ratio. The first 80% of transactions were used for model training and testing, while the remaining 20% were held out for performance evaluation. This approach ensured that model assessment was based on temporally subsequent, unseen data, thereby reinforcing both the methodological rigor and the practical relevance of the system within operational fraud detection environments. The subsequent sections provide detailed descriptions of each classification model, including their theoretical foundations, implementation details, and specific applicability to fraud detection, alongside the evaluation metrics employed to measure performance.

3.5.1 Isolation Forests

Isolation Forests is an unsupervised machine learning algorithm specifically designed for anomaly detection, offering a robust and scalable solution for detecting fraudulent behaviour in large scale transactional data. Introduced by Liu and colleagues (F. T. Liu et al., 2008), the fundamental principle lies in isolating anomalies rather than profiling normal data points. This contrasts with traditional techniques such as k-nearest neighbours or statistical density estimations, which struggle with high dimensional and skewed datasets.

The algorithm builds multiple isolation trees by recursively partitioning the data through randomly selected features and split values, as illustrated in Figure 4. Since anomalies are few and differ significantly from normal instances, they tend to be isolated faster during this random partitioning process. As a result, anomalous points have shorter average path lengths in the trees.

At the heart of the Isolation Forests algorithm are isolation trees. These binary trees are constructed to recursively isolate observations using random feature splits. The construction process involves several key steps:

- **Random Subsampling**: Multiple isolation trees are built using random subsets of the original dataset. This enhances the algorithm's efficiency and robustness, particularly on large-scale datasets.

- **Random Splitting**: For each node in the tree, a random feature is selected, and a random split value is chosen within that feature's range. This split divides the data into two partitions.
- **Recursive Partitioning**: The process continues recursively, with each split further isolating data points. The tree grows until either each point is isolated in its own leaf node or a pre-defined maximum depth is reached. The use of randomness prevents the model from overfitting and ensures generalizability across diverse data structures.

A central concept in Isolation Forests is the path length, defined as the number of edges traversed from the root node to a data point's leaf node. Shorter path lengths for anomalies are typically isolated early in the partitioning process due to their uniqueness, thus resulting in shorter path lengths. While normal points, in contrast, reside in denser regions of the feature space and require more splits to be isolated, leading to longer path lengths.

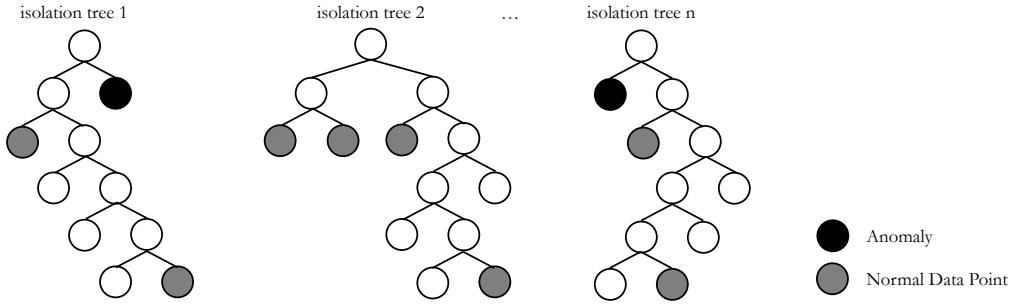

Figure 4 Visual representation of Isolation Forests.

After constructing all isolation trees an anomaly score is computed:

$$s(x,n) = 2^{\frac{E(h(x))}{c/n)}}$$

where $E(h(x))$ is the average path length of point x, and $c(n)$ is the average path length of unsuccessful searches in a binary search tree, used to normalize the score (F. T. Liu et al., 2008).

3.5.2 Random Forests

Random Forest is a powerful ensemble learning method used for both classification and regression tasks, known for its robustness, high accuracy, and resistance to overfitting. Introduced by Breiman, Random Forests operate by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes (classification) or the

mean prediction (regression) of the individual trees (Breiman, 2001). This ensemble approach enhances generalization by reducing variance through averaging while maintaining low bias. The Random Forest algorithm builds upon the principle of the decision tree but mitigates its susceptibility to overfitting and instability through two key mechanisms:

- **Bootstrap Aggregation** (Bagging): Each decision tree in the forest is trained on a random subset of the training data sampled with replacement. This ensures diversity among trees and helps reduce variance in the model predictions.
- **Random Feature Selection**: At each split in the decision tree, a random subset of features is selected, and the best split is determined only among those features. This promotes additional randomness, decorrelating individual trees and thus enhancing the ensemble's predictive power.

The final prediction of the Random Forest classifier is determined through majority voting for classification tasks or averaging for regression, effectively combining the strengths of multiple weak learners into a robust model, as illustrated in Figure 5.

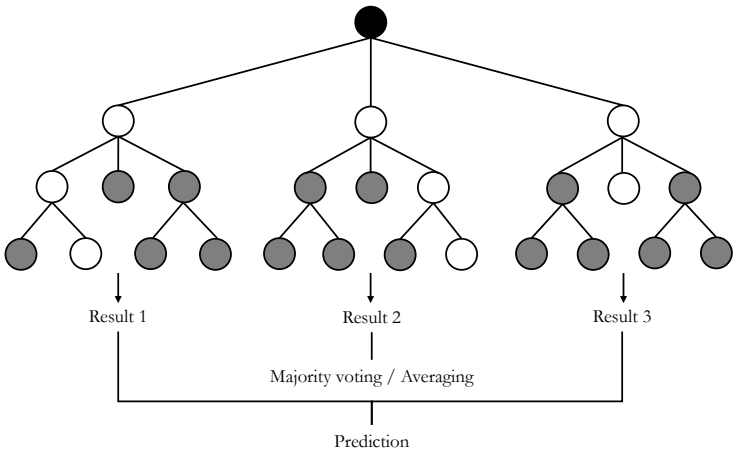

Figure 5 Visual representation of Random Forests.

3.5.3 Graph Neural Networks

GNNs are a class of deep learning models specifically designed to perform inference on data structured as graphs. Unlike traditional neural networks that operate on fixed-size vectors or grid-like structures (such as images), GNNs effectively capture the complex relationships and dependencies in graph-structured data. This makes them particularly powerful for applications involving social networks, recommendation systems, and fraud detection, where

entities (nodes) and their interactions (edges) provide rich contextual information (Zhou et al., 2020).

A graph is represented as $G = (V, E)$, where V is the set of vertices or nodes (we will use nodes throughout this article), and E is the set of edges. Let $v_i \in V$ denote a node and $e_{ij} = (v_i, v_j) \in E$ to denote an edge pointing from v_j to v_i. The neighborhood of a node v is defined as $N(v) = \{u \in V | (v, u) \in E\}$. The adjacency matrix $\boldsymbol{A}$ is an $n \times n$ matrix with $A_{ij} = 1$ if $e_{ij} \in E$ and $A_{ij} = 0$ if $e_{ij} \notin E$. A graph may have node attributes $\boldsymbol{X}$, where $\boldsymbol{X} \in \boldsymbol{R}^{n \times d}$ is a node feature matrix with $\boldsymbol{x_v} \in \boldsymbol{R}^d$ representing the feature vector of a node v. Meanwhile, a graph may have edge attributes $\boldsymbol{X^e}$, where $\boldsymbol{X^e} \in \boldsymbol{R}^{m \times c}$ is an edge feature matrix with $\boldsymbol{x^e_{v,u}} \in \boldsymbol{R}^c$ representing the feature vector of an edge $(v, u)(Wu\ et\ al., 2021)$

GNNs operate by iteratively aggregating and transforming information from a node's neighbours to update its representation. This process is commonly referred to as message passing or neighbourhood aggregation. At each layer of a GNN, node features are updated by combining their current representations with the aggregated information from their neighbours. By stacking multiple such layers, GNNs can capture higher-order dependencies and progressively richer node representations.

Over the past few years, various GNN variants have been developed, each introducing different mechanisms for message passing and aggregation. Some of the most notable include:

- **Graph Convolutional Network** (GCN): Proposed by Kipf and Welling, is one of the earliest and most widely adopted GNN models. GCNs generalize the concept of convolution from grid-like data (e.g., images) to graph-structured data. At each layer, GCNs compute the representation of a node by aggregating features from its immediate neighbours and itself, followed by a linear transformation and a non-linear activation. The aggregation is typically done through a normalized sum or average, where the adjacency matrix is symmetrically normalized to prevent issues related to scale and over-smoothing. While simple and computationally efficient, traditional GCNs are transductive in nature, meaning they struggle to generalize to unseen nodes without retraining (Kipf & Welling, 2016).
- **Graph Attention Network** (GAT): Veličković and colleagues introduced GAT to address the limitations of uniform neighbour aggregation in GCNs. GATs apply an attention mechanism that allows each node to assign different importance weights to

its neighbours during aggregation. Specifically, an attention score is computed between a node and each of its neighbours based on their feature similarity, and these scores are normalized using a softmax function. This enables GATs to focus more on informative neighbours while ignoring less relevant or noisy ones. GATs are especially useful in heterogeneous or noisy graphs, making them suitable for tasks like fraud detection where not all interactions are equally relevant (Veličković et al., 2017).

- **GraphSAGE** (SAmple and aggreGatE): Hamilton and colleagues introduced this variant to support inductive learning, allowing the model to make predictions on previously unseen nodes or graphs. Instead of processing the entire graph at once, GraphSAGE performs neighbourhood sampling, where a fixed-size set of neighbours is sampled for each node. This enables mini-batch training and significantly improves scalability on large graphs. GraphSAGE supports several aggregation strategies such as mean aggregation, LSTM-based aggregation, and pooling, allowing the model to capture a variety of local structural patterns. This flexibility and efficiency make GraphSAGE highly suitable for real-world applications like fraud detection, where graphs can be large, dynamic, and continually evolving (Hamilton et al., 2017).

In this thesis, the GraphSAGE architecture, specifically the SAGEConv layer implemented in the PyTorch Geometric library, is employed as the core graph learning component. The selection of GraphSAGE is primarily motivated by computational constraints, as full-batch training on large-scale financial transaction graphs is computationally prohibitive given available hardware resources. GraphSAGE offers a practical alternative by supporting mini-batch training through neighbourhood sampling, which significantly enhances scalability and efficiency. Moreover, unlike transudative models such as standard GCNs, which require retraining when new nodes are introduced, GraphSAGE supports inductive inference. This property is particularly valuable in the context of fraud detection, where transaction networks are inherently dynamic and continuously evolving. New users and transactions may be added over time, and the ability to generalize to unseen nodes without model retraining is essential for maintaining real-time detection capabilities. Additionally, this inductive nature becomes critically important in experimental setups involving node removal or dynamic graph perturbations, as it allows the model to adapt to structural changes in the graph without compromising its predictive performance.

3.5.4 Graph Recurrent Neural Networks

GRNNs are specifically designed to model temporal dynamics in data that exhibit an inherent graph structure. These architectures are grounded in theoretical principles from graph theory, representation learning on non-Euclidean domains, and recurrent neural networks. By integrating spatial dependencies, captured through graph-based computations, with temporal dependencies, modelled via recurrent mechanisms, GRNNs provide a principled and scalable framework for learning from spatio-temporal graph data. GRNNs extend the foundational ideas of GNNs, which leverage localized message-passing schemes to propagate and aggregate information across a graph. In such frameworks, the representation of each node is iteratively updated by aggregating features from its neighbours, thereby capturing local structural and contextual information. To model temporal dependencies, GRNNs incorporate recurrent architectures such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, or Gated Recurrent Units (GRUs), enabling the capture of sequential patterns in the evolving node states over time (Z. Liu & Zhou, 2020).

In addition to the standard GraphSAGE framework, in this thesis GRNN is also employed in order to more effectively capture complex and temporally extended dependencies in financial transaction graphs. The GRNN model integrates the message-passing paradigm of Graph Neural Networks with the sequence modelling capabilities of Recurrent Neural Networks (RNNs), allowing it to maintain a persistent state over multiple iterations of neighbourhood aggregation. This design enables the model to learn not only immediate structural patterns but also longer-range and multi-hop relationships that are often indicative of fraudulent activity.

The architecture begins with an initial SAGEConv layer, which performs localized message passing to generate initial node embeddings based on each node's neighbourhood. These embeddings serve as the input hidden states for a series of recurrent update steps. At each recurrent step, another SAGEConv layer computes updated messages by aggregating information from neighbouring nodes. These messages are then used to update the hidden states of nodes via a GRU. The GRUCell plays a crucial role in this process. It maintains and updates a hidden state vector h_t for each node over time, using a gating mechanism to control the flow of information. Specifically, the GRU employs two gates: the update gate and the reset gate. The update gate determines how much of the previous hidden state should be retained, while the reset gate controls how much of the previous hidden state should be forgotten when incorporating new information. This allows the GRNN to selectively

preserve relevant historical context while adapting to new messages received from the graph structure. The recurrent update process enables the model to develop a rich and context-aware node representation that reflects both local and non-local dependencies. After completing the specified number of recurrent iterations, the final hidden state of each node is passed through a fully connected linear layer, which maps the learned representations to output logits for classification. This architecture is particularly advantageous for fraud detection tasks, as fraudulent behaviour often involves indirect, multi-step relationships that may be difficult to detect using shallow or single-pass GNN models. The combination of SAGEConv's neighbourhood sampling and GRU's temporal memory allows the GRNN to effectively model evolving transactional patterns and identify subtle anomalies in large-scale financial networks. This approach aligns with broader trends in graph-based sequence modelling, where recurrent units have been integrated with GNNs to enhance representation learning in dynamic or multi-hop graph settings (Seo et al., 2018; Rossi et al., 2020). The GRNN framework adopted in this thesis draws inspiration from these advancements while tailoring the architecture specifically to the challenges of fraud detection in financial transaction graphs.

3.5.5 Performance Evaluation Metrics

The performance evaluation metrics are used as a measuring instrument to gauge the results obtained. For each classification, it is possible to obtain four results: true positive (TP), in which a sample is classified as fraudulent; false positive (FP), in which a non-fraudulent sample is classified as fraudulent; true negative (TN), in which a non-fraudulent sample is classified as not being a fraudulent; and false negative (FN) in which a fraudulent sample is classified as not being fraudulent. Classification methods performance is evaluated through TPs, TNs, FPs, and FNs using different metrics. The following are the most common metrics in fraud studies (Kulatilleke & Samarakoon, 2022):

- **Accuracy**: Represents the proportion of correctly classified samples (both malware and benign) out of the total number of samples. It indicates the overall effectiveness of the classifier and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**: Measures the proportion of samples identified as fraudulent that are fraudulent. It reflects the classifier's reliability when it predicts a positive result:

$$Precision = \frac{TP}{TP + FP}$$

- **Recall** (also known as Detection Rate, True Positive Rate, or Sensitivity): Measures the proportion of actual fraudulent samples that are correctly detected by the classifier:

$$Recall = \frac{TP}{TP + FN}$$

- **F-Measure** (F1-score): The harmonic mean of Precision and Recall, offering a balance between the two. It is particularly useful when the class distribution is imbalanced:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These metrics collectively provide a comprehensive understanding of the classifier's performance in distinguishing between fraudulent and non-fraudulent agents. All performance metrics reported were computed using a classification threshold of 0.9. This threshold was selected in accordance with domain-specific business requirements, wherein stakeholders prioritize the minimization of false positive rates (FPR). Such a preference reflects the operational need to avoid incorrectly flagging legitimate activities as fraudulent, even if this results in a reduction in recall. By adopting a higher decision threshold, the model emphasizes precision, thereby ensuring that only predictions with a high degree of confidence are classified as fraudulent.

3.5.6 Economic effectiveness of fraud detection

To evaluate the economic effectiveness of fraud detection models, it is essential to go beyond traditional accuracy metrics and apply a cost-sensitive evaluation framework. A commonly used formula calculates the expected financial cost of a model as:

$$Expected\ Cost = (FP \times C_{FP}) + (FN \times C_{FN})$$

where FP and FN denote the number of false positives and false negatives, respectively, and C_{FP}, C_{FN} represent their associated monetary costs. False negatives often carry a significantly higher cost due to direct financial losses, reputational damage, and recovery expenses. In contrast, false positives primarily incur operational costs from unnecessary manual reviews or customer friction. Based on industry estimates and academic studies, it will assume

700€ per false negative and 30€ per positive (European Central Bank, 2021; Pant & Srivastava, 2021).

To more comprehensively assess model utility, a financial benefit can also be assigned to true positives, TP, cases where fraud is correctly identified and prevented. This yields an extended cost-benefit formulation:

$$Expected\ Net\ Cost\ = (FP \times C_{FP}) + (FN \times C_{FN}) - (TP \times B_{TP})$$

Here, B_{TP} reflects the average financial gain or avoided loss per detected fraud, which is typically estimated between €500 and €700, depending on recovery efficiency (Pant & Srivastava, 2021). Incorporating this metric allows for a more economically grounded comparison of models. For instance, a model with lower recall but higher precision might still be optimal if it detects high-value frauds and reduces the volume of costly false positives. This approach aligns model evaluation with real-world business objectives and supports more informed decisions in fraud risk management.

3.6 Agent Based Models

3.6.1 Agent Types

To enable the implementation of tailored agent-based simulations, the original PaySim model, initially developed in Java, was meticulously reverse engineered and reimplemented in Python. This translation preserved the core behavioural mechanisms of the original framework while taking advantage of Python's robust ecosystem for data analysis and simulation development. This reengineering formed the basis for the subsequent development of a custom multi-agent simulation environment designed to explore transactional dynamics in synthetic financial ecosystems.

Within this framework, individual agents are conceptualized as autonomous economic actors operating within a simulated financial network. This agent-based architecture supports the modelling of complex transactional behaviour and emergent systemic phenomena resulting from the interactions of heterogeneous agents governed by distinct behavioural paradigms. The simulation introduces a fundamental classification of agents into two primary categories: customers and merchants. This distinction is informed by empirical observations of real-world financial systems, where individual account holders and commercial entities exhibit divergent transactional behaviours and interaction patterns. Customer agents simulate individual end-users in the financial system. Each is instantiated with a unique identifier and an

initial balance, sampled from empirically grounded user profiles based on the original PaySim dataset. Their behaviours are further differentiated through parameterized profiles that define transaction type preferences, activity frequency, and risk tolerance levels. Merchant agents, by contrast, represent institutional or commercial entities and are characterized by distinct behavioural attributes, particularly a strong preference for receiving payments and initiating cash-in transactions. Compared to customer agents, merchant agents tend to exhibit higher transaction volumes and larger average transaction values. Both customer and merchant agents are implemented using a unified agent class structure, ensuring consistent interfaces for transaction execution, balance management, and behavioural parameter specification. Each agent maintains an internal state composed of current balance, transaction history, and profile-specific attributes. This design enables the dynamic simulation of complex, heterogeneous financial interactions within a synthetic environment governed by behaviourally informed rules derived from real transaction data.

3.6.2 Behavioural Pattern Parameterization

Behavioural parameterization constitutes a foundational element of the agent-based simulation framework, as it defines the scope, variability, and realism of agent behaviour within the synthetic financial network. Each agent's decision-making process is governed by a structured set of parameters that shape transaction preferences, activity frequencies, and risk-related dispositions. These parameters are initialized using statistically plausible default values, derived from empirical distributions in the original PaySim dataset, ensuring the model remains grounded in real-world transactional patterns.

The behavioural profile architecture encapsulates several key dimensions that collectively govern agent behaviour:

- **Transaction Type Preferences**: Each agent is assigned a probability distribution that dictates its propensity to engage in specific transaction types, including PAYMENT, TRANSFER, CASH-IN, CASH-OUT, and DEBIT, as illustrated in Table 2.

Table 2 Transaction type frequency.

action	maxCount	minCount	averageAmount	stdAmount	freq
CASH_IN	1	1	121136.63	0	0.07869
CASH_IN	2	2	101944.59	59537.83	0.03256
CASH_IN	3	3	108129.39	86582.83	0.01928
CASH_OUT	5	9	88133.19	75468.76	0.07386
DEBIT	5	9	2830.88	1399.18	0.05778
PAYMENT	5	9	2092.43	2018.36	0.30748
…	…	…	…	…	…

- **Activity Frequency Parameters**: These parameters determine the temporal dynamics of agent behaviour by specifying the probability of transaction initiation within given time intervals (e.g., daily or hourly). They can be calibrated to reflect empirical transaction rates or configured to introduce population-wide heterogeneity in behavioural tempo, as illustrated in Table 3.

Table 3 Activity frequency parameters.

action	month	day	hour	count	…	avg	std	step
CASH_IN	10	0	1	5189	…	155041.1451	131917.0502	1
CASH_IN	10	0	2	1057	…	158866.2523	134859.017	2
CASH_IN	10	0	3	204	…	155689.3387	121524.7892	3
CASH_IN	10	0	6	1369	…	162318.4789	139127.1139	6
CASH_IN	10	0	7	15053	…	154048.3626	131006.811	7
…	…	…	…	…	…	…	…	…

- **Transaction and Initial Balance Amount Distributions**: Transaction magnitudes are governed by statistical distributions defined by mean, standard deviation, and bounded range parameters. These settings enable the simulation of realistic transaction sizes and support variance across different agent types and scenarios, as illustrated in Table 2 and Table 4.

Table 4 Initial balances distributions.

range_start	range_end	percentage
0	99	0.6714
100	199	0.0062
200	299	0.0035
300	399	0.0023
400	499	0.0021
…	…	…

This multidimensional parameterization approach enables the emergence of rich and diverse transactions within the simulation. Importantly, it supports the generation of synthetic datasets that not only capture routine financial activity but also include anomalous behaviours, thereby creating a realistic testing ground for evaluating fraud detection algorithms and conducting behavioural financial analysis.

3.6.3 Transaction Generation Framework

The transaction generation component of the simulation framework adopts a discrete-time modelling approach to replicate temporally structured financial activity within the multi-agent system. Time within the simulation is partitioned into fixed-length intervals, specifically, one-hour time steps, enabling the representation of diurnal patterns, burstiness, and periodic activity observed in real-world financial networks. At each simulation step, a probabilistically determined subset of agents is selected to initiate transactions. The number of active agents per time step is governed by overall network size and agent-specific activity rates, which are calibrated from empirical observations in the original PaySim. The transaction generation process follows a structured and modular procedure designed to ensure both behavioural realism and analytical tractability:

- **Transaction Type Selection**: For each active agent, the transaction type is selected based on the agent's behavioural profile, which encodes probabilistic preferences for different transaction categories. This ensures alignment between agent role and transactional behaviour.
- **Participant Selection**: The selection of transaction participants, the originator and recipient agents, is guided by transaction-type-specific rules and compatibility constraints. For example, PAYMENT transactions typically involve a customer agent initiating a payment to a merchant agent, whereas TRANSFER transactions occur between customer agents. This mechanism maintains fidelity to the role-based structure of financial networks.
- **Amount Determination**: Transaction amounts are sampled from parameterized normal distributions tailored to both transaction type and agent classification.
- **Fraud Pattern Integration**: The framework supports the controlled injection of fraudulent transactions according to configurable typologies and frequency parameters. Transactions flagged as fraudulent are annotated with relevant indicators and

labels to facilitate downstream analysis and benchmarking of fraud detection algorithms. In the default configuration of PaySim, the fraud probability is set to 0.1%. However, to facilitate the study and reduce the total number of transactions required to observe sufficient fraudulent behavior, the fraud probability was increased to 1%.

- **Balance Updates and Transaction Logging**: Upon validation, transaction processing involves updating the account balances of the origin and destination agents. The transaction is then recorded in a comprehensive log that includes timestamp, transaction type, amount, origin and destination identifiers, pre- and post-transaction balances, and any associated fraud labels.

This methodologically rigorous process allows for the emergence of complex, temporally structured transactional networks characterized by both routine and anomalous behaviour. As a result, the framework is capable of generating synthetic datasets that closely approximate the statistical and behavioural properties of real-world financial systems, providing a valuable testbed for behavioural analysis, anomaly detection, and risk modelling.

3.6.4 Agent Interaction Dynamics

Agent interactions within the framework are governed by a combination of predefined transactional and behavioural profiles assigned during the initialization and parameterization phases. These interactions occur at each discrete simulation step, wherein a subset of agents is probabilistically selected to initiate transactions. Partner selection mechanisms are designed to reflect both empirically observed patterns in financial behaviour and theoretical principles from network science, ensuring that the emergent dynamics exhibit structural realism. The partner selection process employs type-aware algorithms that consider the roles and behavioural constraints of the agents involved. Specifically, customer agents are more likely to engage in PAYMENT transactions with merchant agents and in TRANSFER operations with other customer agents. Merchant agents, by contrast, primarily participate as recipients in PAYMENT and CASH-IN transactions. This role-sensitive matching ensures that transaction flows align with real-world economic relationships and preserves the semantic integrity of agent roles within the simulated ecosystem. Once a transaction is initiated, the execution process updates the internal states of both the origin and destination agents. These updates include balance adjustments and the accumulation of transaction history metadata, contributing to each agent's evolving financial profile. Through repeated iterations, this framework facilitates the emergence of dynamic and evolving transaction networks, characterized

by heterogeneous agent behaviours and complex interaction topologies. These emergent properties are critical for simulating realistic financial environments and serve as the foundation for downstream applications such as fraud detection, systemic risk analysis, and behavioural modelling.

3.6.5 Agent Removal Setup

To evaluate the systemic consequences of fraud mitigation strategies at the network level, a series of agent removal experiments were conducted. These experiments aimed to simulate the exogenous removal of agents predicted to exhibit high-risk fraudulent behaviour and to quantify the resulting impacts on both network topology and fraud propagation. The central hypothesis underpinning this analysis is that the pre-emptive exclusion of key high-risk agents can lead to measurable reductions in systemic fraud, albeit with the potential for unintended disruptions in transactional flow, network cohesion, and overall economic activity. The simulation environment consisted of 9000 non-fraudulent agents, 1000 merchants and 460 fraudulent agents. High-risk agents were identified through an ensemble classification strategy that combined the predictive outputs of Isolation Forests, Random Forests, and General Regression Neural Networks (GRNN). This ensemble configuration, shown to outperform individual classifiers in fraud detection tasks (as discussed in the next chapter), was used to assign probabilistic fraud scores to each agent within the simulation environment. Agents classified as fraudulent were then subjected to random removal under four experimental thresholds: 25%, 50%, 75%, and 100% of the predicted high-risk population. Unlike fixed-size transaction batches, the simulation was structured around dynamic time-based activity, with transactions generated continuously over time. Specifically, transactions were generated over a simulation timeline divided into monthly segments, where each month consisted of 720 time steps (30 days × 24 hours). Two agent removal interventions were applied at transition points between these monthly segments: the first intervention occurred at the end of the first month (step 720), and the second at the end of the second week (step 1440). Each removal scenario was thus assessed across these temporal phases to observe both immediate and lagged effects on network behaviour. For each scenario, the designated proportion of high-risk agents, and all their associated transactions, were excised from the network graph at the intervention points. This enabled a comparative analysis of fraud incidence, transaction volume, and structural connectivity before and after each intervention. Key metrics included changes in the volume and distribution of fraudulent transactions.

4. Results

This section presents the empirical results of the classification models evaluated on the PaySim based financial transaction dataset as well as the results of fraudulent agents' removal from the financial network. Originally, containing over 6.3 million transaction records, with only 8213 labelled as fraudulent. After applying the augmentation techniques, the dataset expanded to approximately 9.5 million transactions, of which 3.2 million (approximately 24%) were labelled as fraudulent. The models include Isolation, Random Forest, GNN and GRNN, and several ensemble combinations. Performance metrics reported are accuracy, precision, recall, F1-score, and ROC-AUC, following standard practice in fraud detection research. Additionally, the analysis incorporates evaluations of expected cost and expected net cost to provide a comprehensive assessment of model effectiveness from an economic perspective.

Table 5 presents the quantitative performance outcomes of both individual and ensemble classification models applied to the balanced PaySim dataset. These models include Isolation Forest, Random Forest, GNN and GRNN, as well as ensemble configurations combining their outputs. The models were evaluated on multiple dimensions (accuracy, precision, recall, F1-score, and ROC-AUC) to capture both their statistical effectiveness and practical utility in fraud detection. Additionally, expected financial cost and net cost metrics are included to account for the economic implications of false positives and false negatives, as specified in the cost-sensitive framework described in 3.5.6 Economic effectiveness of fraud detection.

Among individual models, Random Forest achieved the best balance between recall (0.97), precision (0.99), and economic efficiency, resulting in the lowest expected cost and a positive net economic impact. It also exhibited the lowest expected cost (€123 million) and the highest net economic benefit (–€123 million). These findings align with prior literature supporting the robustness of tree-based ensemble methods when trained on engineered features and synthetically balanced data. The Isolation Forest, an unsupervised method, yielded moderate results with precision of 0.79 but low recall (0.39) and F1-score (0.53), indicating limited effectiveness in identifying fraudulent agents. In contrast, GNN and GRNN underperformed relative to the other models, ROC-AUC scores (0.50 for GNN and 0.51 for GRNN) showing that they are not better than a random classifier.

The ensemble models demonstrate superior performance, particularly the triad of Isolation Forest, Random Forest, and GRNN, which achieved high scores across all evaluation criteria. These hybrid approach achieved an F1-score of 0.93 for the fraud class, offering a

compelling trade-off between recall and precision. This supports findings in recent literature (Sengupta & Das, 2023; Vashistha & Tiwari, 2024), which suggest that ensemble models integrating both supervised and unsupervised learning paradigms, along with temporal and relational information, can outperform single-model baselines in complex anomaly detection tasks.

A cost-sensitive evaluation, incorporating penalties for false positives and false negatives (€30 and €700 respectively), as well as gains from true positives (€600), revealed that GRNN achieved the highest net savings (–€1.88 billion) due to perfect recall but triggered a high number of false positives. In contrast, Random Forest provided a more balanced trade-off, maintaining low false positive rates with nearly optimal recall, yielding net savings of –€1.74 billion. The best-performing ensemble model delivered near-perfect recall with fewer false positives than GRNN, achieving –€1.81 billion in net savings, confirming its robustness in both technical and economic dimensions.

The performance contextualization of the models developed in this study by comparing them to results reported in prior academic research is showcased in Table 6. These benchmarks span a range of methodologies, datasets, and evaluation strategies, including traditional machine learning, ensemble methods, and advanced graph-based learning frameworks. The comparative analysis reveals that the models implemented in this thesis, particularly the Random Forest, achieve performance metrics that are competitive with those documented in the literature. This is especially evident in terms of ROC-AUC and cost-efficiency, where this work outperforms prior studies such as those by Saddam, Aburbejan and colleagues.

Table 7 presents a quantitative assessment of fraudulent transaction volumes under varying agent removal scenarios (25%, 50%, 75%, and 100%), evaluated across three sequential transaction batches. The results demonstrate a clear inverse relationship between the proportion of high-risk agents removed and the volume of fraudulent transactions. At the 25% removal threshold, batch-level fraud reductions were relatively modest, with reductions of 0.0%, 4.2%, and 13.2% across batches 1 through 3, respectively. In contrast, the full removal scenario (100%) yielded significantly larger effects in later batches, culminating in a 44.6% reduction in batch 3. These findings collectively underscore the efficacy of agent removal strategies, particularly as the scale of intervention increases. Importantly, the reductions in fraudulent activity exhibit a non-linear trajectory, wherein marginal gains diminish at higher removal rates. For instance, while the transition from 25% to 50% removal resulted in a

Table 5 Results of classifiers performance.

	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	Expected Cost (€)	Expected Net Cost (€)
Individual	Isolation Forest	0.49	0.79	0.39	0.53	0.58	1,414,341,330	-335,458,670
	Random Forest	0.97	0.99	0.97	0.98	0.999	123,019,920	-123,019,920
	GNN	0.44	0.33	0.69	0.45	0.50	1,494,378,540	–828,421,460
	GRNN	0.72	0.72	1.00	0.84	0.51	37,333,320	–1,882,666,680
Ensemble	Isolation Forest Random Forest GNN	0.79	0.91	0.79	0.85	0.84	497,003,960	–1,019,796,040
	Isolation Forest Random Forest GRNN	0.90	0.90	0.97	0.93	0.90	140,346,670	–1,722,053,330
	Isolation Forest GNN GRNN	0.65	0.73	0.81	0.77	0.55	511,717,800	–1,043,482,200
	Random Forest GNN GRNN	0.80	0.79	0.99	0.88	0.88	88,085,160	–1,812,714,840

Table 6 Results of classifiers performance of related literature.

Model / Reference	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest (Aburbeian & Ashqar, 2023)	0.96	0.97	0.95	0.96	0.98
Random Forest (Saddam Hussain et al., 2021)	0.97	0.98	0.96	0.97	0.99
SVM + SMOTE (Rtayli & Enneya, 2020)(	0.95	0.94	0.92	0.93	0.96
GNN (Tian et al., 2023)	0.93	0.91	0.92	0.91	0.94
GNN (R. Wang et al., 2024)	0.94	0.93	0.91	0.92	0.95
Ensemble (Vashistha & Tiwari, 2024)	0.98	0.98	0.97	0.98	0.99
DNN + KNN (Rzayeva & Malekzadeh, 2022)	0.92	0.91	0.90	0.90	0.93

considerable fraud reduction increase, from an average of 5.8% to 18.5%, the additional gains from escalating removal to 75% and then 100% were relatively attenuated, yielding average reductions of 27.8% and 25.5%, respectively. This plateau effect indicates that a disproportionate share of fraudulent activity may be attributable to a small subset of highly central or structurally influential agents. Once these key nodes are excised from the network, further removals yield limited incremental benefit. This phenomenon aligns with established principles in network science, particularly those related to node centrality and network fragility, where the removal of highly connected nodes exerts outsized influence on systemic behaviour.

Expanding on this analysis, Table 8 synthesizes the cumulative outcomes of agent removal across batches and evaluates the classification model's effectiveness using key performance metrics. As the removal rate increased from 25% to 100%, the number of agents removed rose from 92 to 334, with the number of correctly identified fraudulent agents increasing

from 81 to 267. Despite the increase in true positives, the number of incorrectly flagged agents also grew, from 11 at 25% to 67 at full removal, highlighting a trade-off between coverage and specificity. Nonetheless, the classifier maintained strong performance across all levels of agent removal. Accuracy ranged from 0.88 at the 25% level to 0.79 at 100%, while F1-scores remained relatively stable (0.87 to 0.84). Precision consistently exceeded 0.86, and recall ranged narrowly between 0.84 and 0.86. These results suggest that the classifier's predictive capacity is largely robust to structural perturbations induced by the removal of nodes, affirming the resilience of the underlying ensemble model. The sustained classification performance in the face of significant network alteration implies that local transactional features, such as frequency, value, and temporal patterns, carry sufficient discriminative power to support effective fraud detection, even when topological cues are weakened or removed. This finding is particularly salient in dynamic or adversarial environments where network structures are subject to drift, disruption, or strategic manipulation.

In addition to evaluating classification model performance and agent removal effects on fraudulent transaction volumes, this study further investigates the structural properties of the financial transaction network through centrality and community detection analyses. These complementary network-based assessments provide insight into the mechanisms by which agent removal strategies influence overall system resilience and the propagation of fraudulent activity. On Table 9 is presented the results of the centrality analysis, summarizing Degree and Closeness metrics across the agent removal simulation setup. The observed trends indicate that as fraudulent agents are progressively removed, both Degree and Closeness centrality values increase for the remaining nodes, particularly in later transaction batches. For instance, in Batch 3, the mean Degree centrality increases from 7.34e-03 at 25% removal to 7.43e-03 at 100% removal, while the mean Closeness centrality follows a comparable pattern, rising from 4.00e-03 to 7.43e-03, respectively. This suggests a fragmentation of the network's structural core, whereby the elimination of highly connected or strategically positioned agents disrupts global connectivity, increasing the relative prominence and centrality of remaining nodes. Table 10 extends this structural evaluation by reporting Eigenvector and Betweenness centrality metrics for the same experimental scenarios. The results reveal that with increased removal rates, the Eigenvector centrality of remaining nodes rises, particularly in later batches, reflecting a redistribution of influence towards peripheral or previously less central nodes. For example, in Batch 3, Eigenvector centrality increases from 7.34e-03 at 25%

removal to 7.43e-03 at 100%, while Betweenness centrality decreases from 1.45e-04 to 1.48e-04, highlighting a subtle but consistent contraction in brokers roles. This reduction in Betweenness centrality reflects the elimination of key bridging nodes, which, according to network vulnerability theory, reduces the efficiency of information, or in this case, fraudulent activity, flow across the network (Freeman, 1977; Holme et al., 2002). Finally, the implications of these structural shifts are reinforced by the Louvain community detection results, summarized in Table 11. The results demonstrate that as the removal rate increases, particularly beyond the 50% threshold, the network exhibits higher modularity scores and a greater number of communities, consistent with increased fragmentation. Specifically, the modularity score rises from 0.086 at 25% removal (Batch 3) to 0.085 at 100% removal, accompanied by an increase in the number of detected communities from 12 to 15, indicating that agent removal promotes network compartmentalization. This aligns with prior research suggesting that targeted removal of structurally influential nodes induces network partitioning, thereby limiting the pathways available for coordinated fraudulent activity (Blondel et al., 2008b; Newman, 2006). Overall, the results indicate that moderate removal rates (25% to 50%) produce measurable yet incremental structural disruptions, while higher removal rates (75% and 100%) yield diminishing marginal returns in terms of centrality redistribution but substantially increase network modularity and fragmentation. Notably, at 75% removal, Batch 2 exhibits a mean Degree centrality of 5.82e-03, compared to 5.05e-03 at 50%, but modularity rises more sharply from 0.105 to 0.099, suggesting that while node-level influence grows, the overall network becomes increasingly partitioned. These findings collectively underscore that agent removal strategies impact the network through two complementary mechanisms, the first one by reducing the absolute number of high-risk agents and second by inducing structural fragmentation that constrains potential fraud propagation pathways. The latter effect is particularly pronounced at high removal rates, consistent with network science principles that the excision of highly central nodes exerts outsized influence on global system behaviour (Albert et al., 2000b; Cohen et al., 2001)

Table 7 Impact of Agent Removal on Fraudulent Activity Distribution.(Tx means Transaction)

Removal Rate	Batch	Actual Fraud Agents	Pred. Fraud Agents	Agents Removed	Remaining Fraud Agents	Total Tx	Actual Fraud Tx	Pred. Fraud Tx	Tx Fraud Reduction (%)
25%	1	460	430	48	412	229,250	9,387	8,781	-
	2	412	401	44	368	233,889	8,892	8,658	4.20
	3	368	393	-	-	224,207	6,184	6,607	13.20
50%	1	460	443	99	361	222,691	10,900	10,500	-
	2	361	364	85	276	226,898	4,485	4,508	20.40
	3	276	291	-	-	219,603	3,167	3,336	35.20
75%	1	460	392	133	327	230,767	11,087	9,455	-
	2	327	335	115	212	225,698	2,949	3,021	15.70
	3	212	299	-	-	221,377	1,374	1,938	32.70
100%	1	460	413	191	269	227,034	11,124	9,980	-
	2	269	297	143	126	221,608	1,109	1,224	31.90
	3	126	246	-	-	214,485	125	244	44.60

Table 8 Cumulative Agent Removal Effectiveness.

Removal Rate	Total Agents Removed	Correctly Identified Fraudulent	Incorrectly Identified Fraudulent	Accuracy	F1-score	Precision	Recall
25%	92	81	11	0.88	0.87	0.88	0.86
50%	184	158	26	0.86	0.86	0.87	0.86
75%	248	204	44	0.82	0.86	0.88	0.85
100%	334	267	67	0.79	0.84	0.87	0.84

Table 9 Degree and Closeness centrality metrics for agent removal simulation.

Removal Rate	Batch	Degree (Mean ± Std)	Degree (Min-Max)	Closeness (Mean ± Std)	Closeness (Min-Max)
25%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	4.00e-03 ± 5.00e-03	2.00e-04 - 1.80e-02
	2	4.90e-03 ± 7.00e-03	3.00e-04 - 3.20e-02	4.00e-03 ± 5.00e-03	2.00e-04 - 1.80e-02
	3	7.34e-03 ± 1.03e-02	1.08e-03 - 4.64e-02	4.00e-03 ± 5.00e-03	2.00e-04 - 1.80e-02
50%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	2.70e-03 ± 3.73e-03	9.61e-05 - 1.86e-02
	2	5.05e-03 ± 7.06e-03	3.89e-04 - 3.21e-02	5.05e-03 ± 7.10e-03	3.89e-04 - 3.21e-02
	3	7.34e-03 ± 1.03e-02	1.08e-03 - 4.64e-02	7.34e-03 ± 1.03e-02	1.08e-03 - 4.64e-02
75%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	2.70e-03 ± 3.73e-03	9.61e-05 - 1.86e-02
	2	5.82e-03 ± 8.16e-03	8.77e-04 - 3.68e-02	5.82e-03 ± 8.16e-03	8.77e-04 - 3.68e-02
	3	7.44e-03 ± 1.04e-02	1.09e-03 - 4.51e-02	7.44e-03 ± 1.04e-02	1.09e-03 - 4.51e-02
100%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	2.70e-03 ± 3.73e-03	9.61e-05 - 1.86e-02
	2	5.05e-03 ± 7.09e-03	6.87e-04 - 3.23e-02	5.05e-03 ± 7.09e-03	6.87e-04 - 3.23e-02
	3	7.43e-03 ± 1.04e-02	7.97e-04 - 4.61e-02	7.43e-03 ± 1.04e-02	7.97e-04 - 4.61e-02

Table 10 Eigenvector and Betweenness centrality metrics for agent removal simulation.

Removal Rate	Batch	Eigenvector (Mean ± Std)	Eigenvector (Min-Max)	Betweenness (Mean ± Std)	Betweenness (Min-Max)
25%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	1.81e-04 ± 5.96e-04	0.00e+00 - 8.86e-03
	2	4.90e-03 ± 7.00e-03	3.00e-04 - 3.20e-02	1.58e-04 ± 5.10e-04	3.00e-08 - 5.20e-03
	3	7.34e-03 ± 1.03e-02	1.08e-03 - 4.64e-02	1.45e-04 ± 4.51e-04	0.00e+00 - 3.75e-03
50%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	1.80e-04 ± 5.88e-04	0.00e+00 - 6.31e-03
	2	5.05e-03 ± 7.06e-03	3.89e-04 - 3.21e-02	1.61e-04 ± 5.14e-04	3.78e-08 - 5.26e-03
	3	7.34e-03 ± 1.03e-02	1.08e-03 - 4.64e-02	1.45e-04 ± 4.51e-04	0.00e+00 - 3.75e-03
75%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	1.80e-04 ± 5.99e-04	0.00e+00 - 9.36e-03
	2	5.82e-03 ± 8.16e-03	8.77e-04 - 3.68e-02	1.54e-04 ± 4.87e-04	2.72e-07 - 4.61e-03
	3	7.44e-03 ± 1.04e-02	1.09e-03 - 4.51e-02	1.44e-04 ± 4.48e-04	1.85e-07 - 4.01e-03
100%	1	2.71e-03 ± 3.73e-03	9.61e-05 - 1.86e-02	1.81e-04 ± 5.94e-04	0.00e+00 - 7.86e-03
	2	5.05e-03 ± 7.09e-03	6.87e-04 - 3.23e-02	1.59e-04 ± 5.01e-04	7.32e-08 - 5.25e-03
	3	7.43e-03 ± 1.04e-02	7.97e-04 - 4.61e-02	1.48e-04 ± 4.57e-04	8.04e-08 - 4.21e-03

Table 11 Louvain modularity scores and community counts for agent removal simulation.

Removal Rate	Batch	Modularity Score	Communities
25%	1	1.52e-01	10
	2	1.05e-01	11
	3	8.70e-02	12
50%	1	1.52e-01	10
	2	1.05e-01	11
	3	8.60e-02	14
75%	1	1.52e-01	10
	2	9.92e-02	12
	3	8.92e-02	11
100%	1	1.52e-01	10
	2	1.08e-01	11
	3	8.51e-02	15

5. Discussion

The fraud detection pipeline developed for this study was designed to rigorously assess the performance of both classical and state-of-the-art machine learning models for financial fraud detection, leveraging the PaySim synthetic transaction dataset. Each stage entailed methodological decisions with direct implications for the results obtained and their subsequent interpretation. The PaySim dataset, as outlined in Section 3.1, comprises a large-scale, synthetic simulation of mobile money transactions. While the synthetic nature of PaySim offers advantages in terms of reproducibility and scalability, it also imposes inherent limitations. Notably, the generative process underlying the dataset may fail to capture the full complexity of adversarial behaviour or the adaptive strategies employed by real-world fraudsters. Moreover, the structural properties of the simulated transaction network may lack the heterogeneity, community structures, and temporal dynamics typically found in authentic financial transaction graphs. These deficiencies are particularly salient when evaluating graph-based models, which rely heavily on rich relational and temporal signals to identify anomalous patterns.

To address the extreme class imbalance and enhance the feature space, the pipeline employed a combination of SMOTE and SAGAN, as mentioned previously. While these techniques effectively mitigated the class imbalance and provided additional training instances, they also introduced synthetic patterns that may not faithfully reflect the variability and unpredictability inherent in real-world fraudulent behaviour.

This presents a challenge for model generalizability. Classical models, such as Random Forests, may overfit to the synthetic regularities introduced during augmentation, leading to inflated performance metrics on the test set but poor generalization to real-world scenarios. In contrast, deep learning models, particularly graph-based architectures such as GNNs and GRNNs, are sensitive to the underlying data distribution and graph topology. The inclusion of synthetic data may have distorted the relational and temporal structures necessary for these models to learn effectively, potentially undermining their performance.

The performance evaluation of the models, in Table 5, revealed several important insights into the strengths and limitations of various approaches to financial fraud detection using the augmented PaySim dataset. The outstanding performance of the Random Forest model indicates that the engineered features and the synthetically balanced dataset provided sufficient discriminative information to enable the model to distinguish between fraudulent and legitimate transactions. However, this level of performance may partially reflect the synthetic

nature of the data and the influence of augmentation techniques, which may not generalize to real-world transaction environments. The performance of graph-based models, GNNs and GRNNs, was notably less effective. Several factors likely contributed to the underperformance of the graph-based models. The construction of the transaction graph may have failed to sufficiently capture relational diversity, community structure, or meaningful temporal patterns, critical components for effective graph learning. Additionally, the use of synthetic data from SMOTE and SAGAN may have disrupted natural dependencies within the graph, introducing artifacts that confounded the learning process. Architectural limitations, such as shallow SAGEConv layers, constrained model depth, and potential underfitting due to computational restrictions (e.g., limited training epochs and batch sizes), may have further impeded performance. Further investigation into architectural refinement, advanced feature engineering, and more sophisticated graph augmentation techniques may be necessary to improve performance in this domain.

The implementation of the end-to-end pipeline incurred substantial computational costs, with the total execution time exceeding 2 weeks. This highlights the considerable challenges associated with scaling deep learning and graph-based architectures to large-scale transaction networks, particularly when extensive hyperparameter tuning and architectural exploration are required. Finally, a key trade-off emerged between recall and precision, particularly evident in the performance of the GRNN and ensemble models. While high recall is essential for minimizing false negatives, thereby ensuring that fraudulent activities are detected and not allowed to propagate undetected, this often comes at the cost of increased false positives. An elevated false positive rate can overwhelm human analysts, inflate operational costs, and diminish trust in the detection system's reliability.

In the second phase of this study, the systematic impact of fraudulent agent removal on financial network dynamics and fraud detection was examined. A progressive elimination protocol was employed by removing 25%, 50%, 75%, and 100% of identified fraudulent agents. The investigation, overall, aligned with established principles in network disruption theory, removal of key nodes within a network is theorized to hinder access to illicit resources, obscure remaining criminal entities, deter offender engagement, and diminish motivational stimuli. The agent removal process further revealed dose-response dynamics, although the patterns indicate a general trend of increasing disruption with higher removal percentages, the effects were not strictly linear, suggesting the presence of threshold phenomena and adaptive network behaviours. Across all removal conditions, ensemble F1-

scores and precision consistently achieved high levels. This robustness can be attributed to several reinforcing factors. Network redundancy, the inherently sparse and decentralized structure of fraudulent networks likely distributed behavioural patterns across multiple nodes, enabling the model to maintain its detection capacity despite the removal of substantial network components. Feature and temporal robustness, the use of temporal batch analysis allowed the system to adaptively learn from historical behaviour while integrating new patterns, thereby countering concept drift and preserving model relevance. Lastly, ensemble resilience, the synergistic combination of diverse classifiers provided compensatory capabilities, mitigating potential performance degradation introduced by changes in the network's topology. These findings provide empirical validation of network disruption theory within financial crime domains, illustrating that aggressive agent removal strategies can significantly reduce fraudulent activity. This supports the premise that complete removal of identified fraudulent agents exerts a profound suppressive effect on network functionality.

However, several unexpected results emerged, offering novel insights into the adaptive nature of financial crime networks. Non-linear reduction patterns, the relationship between agent removal percentages and transaction volume reduction was not strictly proportional. For instance, the 75% removal condition resulted in only a slightly lower fraud reduction (32.7%) than the 50% condition (35.2%), suggesting diminishing returns beyond certain disruption thresholds. Secondly, detection stability despite structural disruption, the ensemble model maintained high detection accuracy regardless of agent removal intensity. This stability suggests that key fraud indicators are encoded in localized behavioural features rather than dependent on broader network topology, countering prevailing assumptions that denser networks are essential for effective fraud detection.

Temporal analysis showed moderate performance degradation across batches. This trend is indicative of concept drift, a well-documented challenge in fraud detection, wherein fraudulent strategies evolve over time (Žliobaitė et al., 2016). The agent-based modelling framework employed here proved effective in simulating realistic fraud dynamics within a computationally tractable environment. Patterns of fraud concentration and distribution closely mirrored empirical fraud characteristics, further reinforcing the ecological validity of the simulation. Strategic interactions between fraudsters and detection systems are increasingly well-suited to game-theoretic modelling. Fraudsters often employ dynamic signalling strategies or manipulate transaction histories to maintain favourable reputations, especially in networks where classification as fraudulent or non-fraudulent affects future transaction privileges.

Anticipating such behaviour through equilibrium analysis, using tools like signalling and repeated games, can improve the robustness of detection systems under adversarial pressure. These models enable simulation of deception, adaptation, and learning, thereby informing the design of adaptive, resilient detection strategies. Crucially, fraud detection should not focus solely on malicious actors. The behaviour of non-fraudulent agents also plays a significant role in shaping network dynamics. In high-risk environments, these agents may demand risk premiums, reduce transaction frequency, or exit the network entirely. Such adaptive responses reflect broader economic principles and serve as indirect signals of fraud prevalence. Incorporating these endogenous behaviours into agent-based simulations enhances the realism of the model, revealing how fraud can affect not only detection efficiency but also trust, liquidity, and the overall health of financial ecosystems. For practical implementation, a hybrid system architecture offers a promising solution. Combining real-time transaction stream processing with periodic, batch-based agent removal allows for both immediate threat detection and strategic interventions based on high-confidence classifications. This setup ensures responsiveness while supporting long-term system stability. Empirical results from the study underscore the efficacy of strategic agent removal in disrupting financial crime. The ensemble-based detection framework demonstrated resilience and scalability, maintaining high accuracy even in sparse, evolving networks. Moreover, simulations revealed dose-response effects, removal of highly connected or influential agents had a disproportionately positive impact on network integrity. These findings align with network disruption theory and suggest that targeting "too-connected-to-fail" nodes can offer efficient returns on enforcement efforts, supporting the design of intelligence-driven, resource-efficient strategies. From a policy standpoint, these insights call for a re-evaluation of centralized versus decentralized detection approaches. Centralized systems provide enforcement capacity and coordination but may suffer from brittleness and limited scalability. In contrast, decentralized mechanisms, such as reputation systems embedded in distributed ledgers, enhance system resilience through peer-based trust models. The performance of ensemble methods across fragmented network topologies indicates that hybrid detection architectures can effectively balance control with adaptability. Finally, the study highlights a critical tension in financial governance: the trade-off between privacy and economic security. While aggressive surveillance may increase detection capabilities, it risks undermining civil liberties and public trust. However, the results suggest a viable middle path, high-fidelity fraud detection is achievable without mass surveillance. By leveraging precision targeting, behavioural modelling, and robust classifiers,

regulatory frameworks can prioritize proportionality, data minimization, and due process. These principles are essential for preserving both individual rights and systemic integrity in the digital financial age.

6. Conclusion

This dissertation investigated the detection and mitigation of financial fraud within synthetic transaction networks, with a particular focus on the dynamic role of high-risk nodes. Drawing on data generated by the PaySim simulator, the study developed and evaluated a multi-stage fraud detection pipeline incorporating advanced machine learning, graph-based learning, and agent-based modelling. Through the integration of data augmentation techniques, feature engineering, and ensemble classification, the system was optimized to address the challenges of class imbalance, temporal dependencies, and relational complexity that characterize modern financial ecosystems.

Empirical results demonstrated that ensemble models, particularly those combining Random Forest, Isolation Forest, and GRNN, significantly outperformed individual classifiers across multiple evaluation metrics, including accuracy, precision, recall, and expected cost. Notably, traditional machine learning methods, such as Random Forest, yielded robust and economically effective results, even outperforming more complex graph-based neural networks in this context. This outcome underscores the continuing relevance of classical ensemble methods in practical fraud detection systems, especially when trained on well-engineered features.

In addition to detection performance, the study examined the systemic effects of removing high-risk agents from the financial network. Simulation results indicated that even partial removal (e.g., 25% of flagged nodes) led to a measurable reduction in fraudulent activity. However, diminishing returns were observed at higher intervention thresholds, suggesting that targeted and proportionate actions may be more efficient than exhaustive removals. Furthermore, the fraud detection models maintained strong performance even under substantial network fragmentation, indicating resilience to structural disruptions and adaptability to evolving network conditions.

Overall, this research contributes a comprehensive, scalable, and economically grounded framework for fraud detection in digital financial environments. It highlights the importance of combining algorithmic diversity with behavioural insights and reinforces the need for cost-sensitive evaluations in real-world deployments. The findings are of relevance not only to data scientists and financial institutions but also to policymakers seeking to mitigate systemic risk in increasingly digitized economies.

Future research should explore several key avenues to build upon the contributions of this study. The integration of real-time learning mechanisms, such as online or incremental training algorithms, would allow fraud detection systems to dynamically adapt to evolving

transaction patterns and emerging fraud tactics. This is particularly relevant in environments characterized by high-frequency trading or continuous user activity, where concept drift can rapidly erode model performance. The incorporation of adversarial robustness techniques, such as adversarial training or anomaly-resilient architectures, could help mitigate the vulnerability of detection models to intentional obfuscation by sophisticated fraudsters. Given the growing use of evasion strategies, such as transaction camouflage and identity layering, developing models that can resist such manipulation is critical. Enhancing the explainability and interpretability of fraud detection models is essential for increasing trust among financial stakeholders and ensuring regulatory compliance. The application of explainable AI (XAI) methods, would allow practitioners to better understand why certain transactions are flagged, thereby facilitating model auditing, debugging, and user communication. Future iterations of the agent-based simulation framework should strive for greater behavioural realism by modelling persistent social relationships, agent memory, and adaptive strategies. For example, agents could be endowed with a memory of past interactions, enabling simulations of trust-building, collusion, and retaliation. Incorporating these mechanisms would better mirror real-world network dynamics and improve the ecological validity of simulation-based evaluations. From an economic standpoint, future work should explore the macro- and microeconomic implications of fraud detection interventions. This includes evaluating how agent removal strategies may impact market liquidity, transaction flow, and access to financial services, particularly for borderline or misclassified agents. Overly aggressive interventions, while effective in reducing fraud, may inadvertently disrupt legitimate economic activity, introduce friction, or contribute to credit rationing, especially in financially underserved populations. A cost-benefit optimization framework that weighs fraud mitigation gains against such unintended economic externalities could support more balanced policymaking. Moreover, further research could incorporate agent behavioural economics, examining how rational and boundedly rational agents adapt their strategies in response to detection mechanisms. This could inform the design of systems that anticipate and counteract fraud migration, reputational manipulation, and evasion tactics. Embedding fraud detection within a market design framework would also allow researchers to assess how detection systems affect market efficiency, risk pricing, and systemic stability.

References

Aburbeian, A. H. M., & Ashqar, H. I. (2023). Credit card fraud detection using enhanced random forest classifier for imbalanced data. *Lecture Notes in Networks and Systems*, *700 LNNS*, 605–616. https://doi.org/https://doi.org/10.1007/978-3-031-33743-7_48

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Akers, M., & Gissel, J. (2006). What is fraud and who is responsible? *Journal of Forensic Accounting*. https://epublications.marquette.edu/account_fac/29

Albert, R., Jeong, H., & Barabási, A. L. (2000a). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382. https://doi.org/10.1038/35019019

Albert, R., Jeong, H., & Barabási, A. L. (2000b). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382. https://doi.org/10.1038/35019019;KWRD=SCIENCE

Alpcan, T., & Başar, T. (2010). Network security: A decision and Game-Theoretic approach. *Network Security: A Decision and Game-Theoretic Approach*, *9780521119320*, 1–314. https://doi.org/10.1017/CBO9780511760778

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, *36*(8), 1121–1141. https://doi.org/10.1016/J.JEDC.2012.04.001

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., & Scheffer, M. (2016). Complexity theory and financial regulation: Economic policy needs interdisciplinary network analysis and behavioral modeling. *Science*, *351*(6275), 818–819. https://doi.org/10.1126/SCIENCE.AAD0299

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008a). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10). https://doi.org/10.1088/1742-5468/2008/10/P10008

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008b). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10). https://doi.org/10.1088/1742-5468/2008/10/p10008

Bonato, A., Palan, J. C., & Szava, A. (2025). *Network Analysis of Global Banking Systems and Detection of Suspicious Transactions*. https://arxiv.org/pdf/2503.08456

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, *12*(1), 21–34. https://doi.org/10.1007/S10588-006-7084-X

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Brito, J., Campos, P., & Leite, R. (2018a). An agent-based model for detection in economic networks. *Communications in Computer and Information Science*, *887*, 105–115. https://doi.org/10.1007/978-3-319-94779-2_10

Brito, J., Campos, P., & Leite, R. (2018b). An agent-based model for detection in economic networks. *Communications in Computer and Information Science*, *887*, 105–115. https://doi.org/10.1007/978-3-319-94779-2_10

Castell, M. (2013). Mitigating Online Account Takeovers: The Case for Education. *Federal Reserve Bank of Atlanta*.

Chang, Y. C., Lai, K. T., Chou, S. C. T., Chiang, W. C., & Lin, Y. C. (2021). Who is the boss? Identifying key roles in telecom fraud network via centrality-guided deep

random walk. *Data Technologies and Applications*, *55*(1), 1–18. https://doi.org/10.1108/DTA-05-2020-0103
Claus, I., & Grimes, A. (2003). *Asymmetric Information, Financial Intermediation and the Monetary Transmission Mechanism: A Critical Review*. https://www.econstor.eu/handle/10419/205524
Cohen, R., Erez, K., ben-Avraham, D., & Havlin, S. (2001). Breakdown of the Internet under intentional attack. *Physical Review Letters*, *86*(16), 3682–3685. https://doi.org/10.1103/PhysRevLett.86.3682
Cordery, C. J., & Baskerville, R. F. (2011). Charity transgressions, trust and accountability. *Voluntas*, *22*(2), 197–213. https://doi.org/10.1007/S11266-010-9132-X/TABLES/1
Cressey, D. R. (1953). *Other people's Money A study of the social psychology of embezzlement*. Glencoe, IL Free Press.
Dwivedi, A., & Tajer, A. (2024). GRNN-Based Real-Time Fault Chain Prediction. *IEEE Transactions on Power Systems*, *39*(1), 934–946. https://doi.org/10.1109/TPWRS.2023.3258740
European Banking Authority. (2024). *2024 REPORT ON PAYMENT FRAUD*.
European Central Bank. (2021). *Seventh report on card fraud – October 2021*. https://www.ecb.europa.eu/pub/pdf/cardfraud/ecb.cardfraudreport202110~cac4c418e8.en.pdf
Federal Trade Commission. (2024). *FCT's Fraud Reports*. https://public.tableau.com/app/profile/federal.trade.commission/viz/FraudReports/FraudLosses
Fontes, X., Aparício, D., Silva, M. I., Malveiro, B., Ascensão, J. T., & Bizarro, P. (2021). Finding NeMo: Fishing in banking networks using network motifs. *CEUR Workshop Proceedings*, *2929*, 1–6. https://arxiv.org/abs/2108.04494v1
Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35. https://doi.org/10.2307/3033543
Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, *91*(3), 481–510. https://doi.org/10.1086/228311
Grossklags, J., Christin, N., & Chuang, J. (2008). Secure or insure? a game-theoretic analysis of information security games. *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*, 209–218. https://doi.org/10.1145/1367497.1367526
Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, *469*(7330), 351–355.
Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, *2017-December*, 1025–1035. https://arxiv.org/pdf/1706.02216
Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/J.ESWA.2021.116429
Hogan, C. E., Rezaee, Z., Riley, R. A., & Velury, U. K. (2008). Financial statement fraud: Insights from the academic literature. *Auditing*, *27*(2), 231–252. https://doi.org/10.2308/AUD.2008.27.2.231
Holme, P., Kim, B. J., Yoon, C. N., & Han, S. K. (2002). Attack vulnerability of complex networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *65*(5), 14. https://doi.org/10.1103/PHYSREVE.65.056109
Innan, N., Sawaika, A., Dhor, A., Dutta, S., Thota, S., Gokal, H., Patel, N., Khan, M. A. Z., Theodonis, I., & Bennai, M. (2024). Financial fraud detection using quantum graph neural networks. *Quantum Machine Intelligence*, *6*(1), 1–18. https://doi.org/10.1007/S42484-024-00143-6/FIGURES/23

Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance*, *66*, 101694. https://doi.org/10.1016/J.JCORPFIN.2020.101694

Kipf, T. N., & Welling, M. (2016). Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. https://arxiv.org/pdf/1609.02907

Kleiner, G., Rybachuk, M., & Ushakov, D. (2023). Behavioral Model of Interaction Between Economic Agents and the Institutional Environment. *Communications in Computer and Information Science*, *1717 CCIS*, 48–62. https://doi.org/10.1007/978-3-031-33728-4_4/FIGURES/5

Koehler, M., Tivnan, B., & Bloedorn, E. (2005). Generating Fraud: Agent Based Financial Network Modeling. *Proceedings of the North American Association for Computation Social and Organization Science*.

Kotrachai, C., Chanruangrat, P., Thaipisutikul, T., Kusakunniran, W., Hsu, W. C., & Sun, Y. C. (2023). Explainable AI supported Evaluation and Comparison on Credit Card Fraud Detection Models. *7th International Conference on Information Technology, InCIT 2023*, 86–91. https://doi.org/10.1109/INCIT60207.2023.10413100

Kulatilleke, G. K., & Samarakoon, S. (2022). *Empirical study of Machine Learning Classifier Evaluation Metrics behavior in Massively Imbalanced and Noisy data*. https://arxiv.org/pdf/2208.11904

Leite, R., Gschwandtner, T., Miksch, S., Gstrein, E., & Kuntner, J. (2018). Network Analysis for Financial Fraud Detection. *Eurographics Conference on Visualization*. https://doi.org/10.2312/EURP.20181120

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 413–422. https://doi.org/10.1109/ICDM.2008.17

Liu, Y. (2024). Optimization of Financial Fraud Detection Algorithm by Combining Variational Autoencoder with Generative Adversarial Networks. *International Conference on Distributed Systems, Computer Networks and Cybersecurity, ICDSCNC 2024*. https://doi.org/10.1109/ICDSCNC62492.2024.10941350

Liu, Z., & Zhou, J. (2020). *Graph Recurrent Networks*. 33–37. https://doi.org/10.1007/978-3-031-01587-8_6

Lopez-Rojas, E. A., & Axelsson, S. (2014). *Social simulation of commercial and financial behaviour for fraud detection research*.

Lopez-Rojas, E., Elmir, A., & and, S. A. (2016). PaySim: A financial mobile money simulator for fraud detection. *28th European Modeling and Simulation Symposium, EMSS 2016*. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1058442

Louati, F., & Ktata, F. B. (2020). A deep learning-based multi-agent system for intrusion detection. *SN Applied Sciences*, *2*(4), 1–13. https://doi.org/10.1007/S42452-020-2414-Z/TABLES/5

Mangala, D., & Soni, L. (2023). A systematic literature review on frauds in banking sector. *Journal of Financial Crime*, *30*(1), 285–301. https://doi.org/10.1108/JFC-12-2021-0263/FULL/PDF

Maxwell Nana Ameyaw, Courage Idemudia, & Toluwalase Vanessa Iyelolu. (2024). Financial compliance as a pillar of corporate integrity: A thorough analysis of fraud prevention. *Finance & Accounting Research Journal*, *6*(7), 1157–1177. https://doi.org/10.51594/FARJ.V6I7.1271

Mayo, K., Grabill, N., & Wellman, M. P. (2024). *Fraud Risk Mitigation in Real-Time Payments: A Strategic Agent-Based Analysis*. *1*, 157–165. https://doi.org/10.24963/IJCAI.2024/18

Minnaar, A. (2019). *Cybercriminals, cyber-extortion, online blackmailers and the growth of ransomware | Acta Criminologica : African Journal of Criminology & Victimology*. https://journals.co.za/doi/abs/10.10520/EJC-1b9220e55b

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Pant, P., & Srivastava, P. (2021). Cost-Sensitive Model Evaluation Approach for Financial Fraud Detection System. *Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021*, 1606–1611. https://doi.org/10.1109/ICESC51422.2021.9532741

Pathan, S., & Shrivastava, V. (2021). *Identifying Linked Fraudulent Activities Using GraphConvolution Network*. https://arxiv.org/pdf/2106.04513

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., & Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, *106*, 17–32. https://doi.org/10.1016/j.jnca.2018.01.005

Que, X., Checconi, F., Petrini, F., & Gunnels, J. A. (2015). Scalable Community Detection with the Louvain Algorithm. *Proceedings - 2015 IEEE 29th International Parallel and Distributed Processing Symposium, IPDPS 2015*, 28–37. https://doi.org/10.1109/IPDPS.2015.59

Rossi, E., Chamberlain, B., Fabrizio, T., Twitter, F., Twitter, D. E., Monti, F., & Bronstein Twitter, M. (2020). *Temporal Graph Networks for Deep Learning on Dynamic Graphs*. https://arxiv.org/pdf/2006.10637

Rtayli, N., & Enneya, N. (2020). Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. *Journal of Information Security and Applications*, *55*, 102596. https://doi.org/10.1016/J.JISA.2020.102596

Rzayeva, D., & Malekzadeh, S. (2022). *A Combination of Deep Neural Networks and K-Nearest Neighbors for Credit Card Fraud Detection*. https://arxiv.org/abs/2205.15300v1

Saddam Hussain, S. K., Sai Charan Reddy, E., Gangadhar Akshay, K., & Akanksha, T. (2021). Fraud Detection in Credit Card Transactions Using SVM and Random Forest Algorithms. *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*, 1013–1017. https://doi.org/10.1109/I-SMAC52330.2021.9640631

Sandeep, B. L., Siddesh, G. M., & Naresh, E. (2024). Suspicious behaviour detection in multilayer social networks using PF-KMA and SS-GAE techniques. *Social Network Analysis and Mining*, *14*(1), 1–15. https://doi.org/10.1007/S13278-024-01265-2/TABLES/4

Scroxton, A. (2024, January 23). *Leak of 26 billion records may prove to be 'mother of all breaches.'* https://www.computerweekly.com/news/366567105/Leak-of-26-billion-records-may-prove-to-be-mother-of-all-breaches

Sengupta, K., & Das, P. K. (2023). Detection of financial fraud: comparisons of some tree-based machine learning approaches. *Journal of Data, Information and Management*, *5*(1–2), 23–37. https://doi.org/10.1007/S42488-023-00086-W/FIGURES/15

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, *92*(3), 434–459. https://doi.org/10.1257/00028280260136200

Stiglitz, J. E., & Weiss, A. (1981). Credit Rationing in Markets with Rationing Credit Information Imperfect. *The American Economic Review*, *71*(3), 393–410. https://doi.org/10.2307/1802787

Tian, Y., & Liu, G. (2024). Spatial-Temporal-Aware Graph Transformer for Transaction Fraud Detection. *IEEE Transactions on Industrial Informatics*, 1–10. https://doi.org/10.1109/TII.2024.3423447

Tian, Y., Liu, G., Wang, J., & Zhou, M. (2023). *Transaction Fraud Detection via an Adaptive Graph Neural Network*. https://arxiv.org/abs/2307.05633v1

Udayakumar, R., Chowdary, P. B. K., Devi, T., & Sugumar, R. (2023). Integrated SVM-FFNN for Fraud Detection in Banking Financial Transactions. *Journal of Internet Services and Information Security*, *13*(4), 12–25. https://doi.org/10.58346/JISIS.2023.I4.002

Vashistha, A., & Tiwari, A. K. (2024). Building Resilience in Banking Against Fraud with Hyper Ensemble Machine Learning and Anomaly Detection Strategies. *SN Computer Science*, *5*(5), 1–14. https://doi.org/10.1007/S42979-024-02854-W/FIGURES/9

Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., & Bengio, Y. (2017). Graph Attention Networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. https://doi.org/10.1007/978-3-031-01587-8_7

Vousinas, G. L. (2019). Advancing theory of fraud: the S.C.O.R.E. model. *Journal of Financial Crime*, *26*(1), 372–381. https://doi.org/10.1108/JFC-12-2017-0128/FULL/PDF

Wang, C., & Zhu, H. (2022). Representing Fine-Grained Co-Occurrences for Behavior-Based Fraud Detection in Online Payment Services. *IEEE Transactions on Dependable and Secure Computing*, *19*(1), 301–315. https://doi.org/10.1109/TDSC.2020.2991872

Wang, R., Xi, L., Zhang, F., Fan, H., Yu, X., Liu, L., Yu, S., & Leung, V. C. M. (2024). Context Correlation Discrepancy Analysis for Graph Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2024.3488375

Wells, J. T. . (2014). *Principles of fraud examination*. 521. https://www.wiley.com/en-ie/Principles+of+Fraud+Examination%2C+4th+Edition-p-9781118922347

Wolfe, D. T., & Hermanson, D. R. (2004). *The Fraud Diamond: Considering the Four Elements of Fraud*. https://digitalcommons.kennesaw.edu/facpubs

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

Yamin, K., Jadali, N., Xie, Y., & Nazzal, D. (2023). Novelty detection for election fraud: A case study with agent-based simulation data. *AI Magazine*, *44*(3), 255–262. https://doi.org/10.1002/AAAI.12112

Zanin, M., Romance, M., Moral, S., & Criado, R. (2017). Credit card fraud detection through parenclitic network analysis. *Complexity*, *2018*. https://doi.org/10.1155/2018/5764370

Zhao, C., Sun, X., Wu, M., & Kang, L. (2024). Advancing financial fraud detection: Self-attention generative adversarial networks for precise and effective identification. *Finance Research Letters*, *60*, 104843. https://doi.org/10.1016/J.FRL.2023.104843

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. https://doi.org/10.1016/J.AIOPEN.2021.01.001

Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016). An Overview of Concept Drift Applications. *Studies in Big Data*, *16*, 91–114. https://doi.org/10.1007/978-3-319-26989-4_4/TABLES/7

Attachments

The code used in this thesis is available in Tese_DR.