Asia Pacific Fraud Journal
E-ISSN: 2502-695X, ISSN: 2502-8731
Volume 10, Issue 1 (January-June) 2025
*Asia Pacific Fraud Journal now has been accredited "**SINTA 3**" by Ministry of Research and Technology of The Republic of Indonesia (No SK. 225/E/KPT/2022).*
Available online at: http://apfjournal.or.id/index.php/apf

Fraud Syndicates Within Digital Ecosystem: Graph Network and Transaction Analysis Approach

[1]Ferdi Hidayat Irawandi, [2]Kristy Natasha Yohanes, ✉[3]Lalu Garin Alham
[1]Digital Cybersecurity and Fraud Management - PT Bank Multiarta Sentosa Tbk, Indonesia
[2]Data & AI - PT Dans Multi Pro, Indonesia
[3]Fraud Management & Authorization - PT Espay Debit Indonesia Koe (DANA), Jakarta, Indonesia

ARTICLE INFORMATION
Article History:
Received November 4, 2024
Revised May 21, 2025
Accepted June 20, 2025

DOI:
10.21532/apfjournal.v10i1.381

ABSTRACT
This paper aims to develop and test methods to detect organized crime, fraud syndicates, and Money Laundering schemes within an e-wallet ecosystem. Analytical An analyticalprocess framework combining Graph Analytics and Supervised Learning is developed and trained with sample spaces of fraud and non-fraud. The pipeline utilized Heterogeneous Graph Transformation (HGT), Graph Statistics (Centrality Measures and Community Detection), and a Gradient Boosting Model to produce models for the detection of fraud syndicate syndicatesand organized crime. Welch's t-test is employed to infer variance differences between samples. Findings confirm the hypothesis that fraud networks are markedly different, exhibiting a more centralized and isolated network compared to the organic, interconnected behaviors of non-fraudulent users. Fraud networks are further characterized by multiple isolated clusters, indicating distinctive groups or behaviors. The proposed method can provide validated methods of fraud and money laundering detection, especially for financial decision-makers and policymakers, to enhance fraud detection systems by improving the protection, integrity, and security of customers and digital transactions.

Keywords: Financial Crime, Fraud, Syndicate, Graph, Neural Network, Machine Learning.

How to Cite:
Irawandi, F. H., Yohanes, K. N., Alham, L. G. (2025). Fraud Syndicates Within Digital Ecosystem: Graph Network and Transaction Analysis Approach. *Asia Pacific Fraud Journal, 10*(1), 153-169. http://doi.org/10.21532/apfjournal.v10i1.381.

✉Corresponding author :
Email: alhamgarin@gmail.com

1. INTRODUCTION

The 2020-2023 COVID-19 pandemic transformed the global economic landscape with the emphasized ubiquity of digital remote financial activities. World Bank (2022) reported a surge of digital financial transactions within the timeframe of 2017 and 2021, with the growth rate reaching up to 67% globally. In addition to the report, around two-thirds of adult citizens worldwide have had experience receiving or paying digitally through electronic devices.

Indonesian regulatory body and central bank, Bank Indonesia, since 2019, to a greater extent, created foundational provisions to regulate and administer "*Sistem Pembayaran*" [Payment System] architectures and infrastructures through its Indonesian Payment System Blueprint 2025 (Bank Indonesia, 2019), signaling support for digitalization of banking & financial landscape. These digital financial frameworks further support financial inclusion for unbanked societies through alternative platforms such as "*Industri Keuangan Non-Bank* (IKNB)" [English: Non-Bank Financial Industries], payment services, payment infrastructures (switches, Payment Gateways), and others. The presence of *e-fraud* (electronic fraud) targeting both customers and the financial institution itself reported by Mastercard (2024) to increase by 17% in transactions from 2022 to 2023. Europol (2023), through Internet Organised Crime Threat Assessment (IOCTA) initiatives, defined several emerging threats of M/O (*Modus Operandi*) of fraudsters. Namely, identity fraud using *deep fakes/Generative* AI, phishing, shimming, Account Take-Over (ATO), and BEC (*Business Email Compromise*). The fraud acts mentioned above emphasize some forms of collusion between customer accounts, merchant accounts, or other fraud actors. FATF (2023) recognized that fraudsters are now more likely to work together by forming cells of subject-matter specialists. In other words, fraudsters are organizing resources, segregating their duties based on compartmented expertise, and coordinating their acts. *Cyber-enabled fraud* (CEF) became more apparent by attacking areas such as banks, payment, remittance services, card providers, virtual asset service providers (VASPs), and e-wallets.

Accordingly, this paper attempts to explore, examine, and test the effectiveness of analytical methods to detect the prevalence of organized crime & fraud syndicates and their inherent behaviors within the payment service's digital ecosystem.

2. LITERATURE REVIEW AND HYPOTHESIS

On Organized Crime & Fraud Syndicate

Black's Law Dictionary (retrieved 2017) and ACFE (n.d.) defined fraud as a deliberate misrepresentation & concealment of truth and material fact to incite others to act to his or her detriment. Those acts may include any deliberate deed that aims to deprive another of property or money of using trickery or cons. The act of fraud itself may or may not directly violate a Criminal Code, depending on the nature of the act itself, the involved parties, and the jurisdictions. Shulzenko (2020) and Polkowski (2013) mentioned that the issue of fighting fraud is becoming more challenging as the internet continues to be a vital part of global financial. Fraudsters alike are already on board and are digitally facilitated as well to find their way around using networked systems.

Levi (2008), Tremblay (1993), and Felson (2003) postulated that the likelihood of a fraud/crime being committed adheres to the accessibility, availability, and traits/characteristics suitability of co-offenders. As for the organized crime and fraud syndicates, the adage of "*It takes two to tango*" (Alekhin & Shmatenko, 2018) does perfectly encapsulate the discourses. This means that an act of fraud, especially in complex target profiles such as electronic transactions, mostly require more than one perpetrator to be successfully carried out. Hinting at a higher probability of success, regeneration, and 'sustainability' of criminal acts where corroborating fraud

actors are within proximity to each other. Whether it is in terms of geographical distance, access & communications, victimology/market profile, language barrier, or even socio-cultural settings. Additionally, scholars and practitioners within legal & law enforcement alike (Campana, 2024) have long argued that acts of criminals/frauds themselves are inherently complex. When an illicit profit is acquired, the corresponding flows of the funds are inherently relational in nature as they are made of interconnecting points of interest. For example, from one another; individual accounts, registered identities, devices and/or access points, financial institutions, or even across borders and jurisdictions.

Graph Theory & Social Network Analysis

Borgatti (2013) noted that inherently complex relations and behaviors of connection mentioned in the passages above can be conceptualized, analyzed, and mapped systematically through the study of Social Network Analysis (SNA). A theorem in which Euler (1736) previously concocted to optimally solve the *path traversal* problem as a branch of mathematics called *Graph Theory* (Sylvester, 1878) into multiple areas of study, including sociology, criminology, and Information Technology.

As an even larger volume of data is being produced nowadays from customers' activities, global movement alike have scrambled to utilize and develop Graph-related technologies as an alternative to conventional structured Relational Database Management System (RDBMS), with Graph & network analytics usage grown by 700% from 2021 to 2025 (Gartner Inc., 2022). Graph data enables inference of relationship patterns in a network of information not available in a conventional tabular form dataset, enabling tasks such as pathfinding & optimization problems, recommendation generation, and even fraud detection (Rodrigues, 2023).

Supervised Machine Learning

Supervised machine learning is a group of algorithms and statistical functions that work by generating a function of the response variable based on the input of the control variable (Nasteski, 2017). These algorithms work by approximation of the desired results and are corrected or reinforced to have its learner output the optimum predictive performance function, cue supervised. AI/ML (Sarker, 2021) allows users to simply start from pre-labelled data in building the models. This paper encompass Graph Theory and Machine Learning analysis of network behavior from both perspectives of fraud and non-fraud actors.

Hypothesis

Based on the laid out theoretical passages above, the possible conjectures for this writing can be devised as follows.

a. Null hypothesis (H0), where $s_1 = s_2$. There is not enough evidence on the existence of differentiating effects from the analysis between the fraud and non-fraud sample spaces, or
b. Alternative hypothesis (H1), where $s_1 \neq s_2$. There is enough evidence on the differentiating effect of the analytical model between fraud and non-fraud observed sample space.

3. METHODS

Data

The author extracted and utilized the data of transactions with a duration of 31 months, ranging from October 2021 until May 2024. The entities included in the data are the *debit party*/sender (source node), *credit party*/recipient (target node), as a ground to congregate a graph map. Datasets are then separated into '**fraudulent data'** (S_1) (data whose actors were identified as fraudulent actors) and '**non-fraudulent data'** (S_2)(data of transactions from legitimate customers), which then fed into the analytical pipeline. Dataset headers in this paper consist of nodes (source and target) and edges (connection & directionality [directed, undirected]) is composed of 2,021,750 edges (1,882,839 transactions to fraudulent merchants and 138,911 transactions from fraudulent customers) and 17,868 nodes (199 fraudulent merchants; 17,669

fraudulent customers). is composed of 2,021,750 transactions (randomized) and 113,818 nodes (12,008 merchants; 101,810 customers).

The description of the variables is as follows (Table 1).

a. source (v_1): the originating (debit party) account that initiated the transaction.
b. target (v_2): the destination (credit party) account that received the transaction.
c. trans_amount (v_3): amount of funds in the transaction (in Indonesian Rupiah/ IDR).
d. description (v_4): types of the transaction [e.g., "Buy Goods", "Peer-to-Peer Transfer"].

Conceptual Frameworks

Graph and Social Network Analysis

Network Analysis (Chiesi, 2001) involves representing actors and translating the data into a graph as nodes (e.g., customer, merchant) and their transactional relations as vertices/edges to analyze their connection structures. A graph function (Berge, 1958) can be described as.

G=(V,E)

Where

a. V is a set of nodes/vertices. A node represents a customer account or merchant that acts as either a source (sender) or a target (recipient).
b. E is a subset of nodes and y that formed the set of edges/connections. Mathematically annotated as $E \subseteq \{\{x,y\} | x,y | \in V \text{ and } x \neq y\}$. Where $\{x \neq y\}$ is a [sic] *loop*; in a payment service a transaction cannot be directly displaced into itself.

The set of analytical models that will be discussed in this paper are described in the following section.

Graph Statistics and Community Detection

Girvan-Newman (2002) Community Detection algorithm is a method to detect communities between *nodes* by progressively omitting *edges* from a clus-ter of connections until it finds the most probable connecting *edges* to form a "community". The list of graph statistics to be acquired.

a. Centrality measurements (Newman, 2010: Betweenness, Eigenvector, Closeness, Degree of connectedness (*degree, weighted degree*),
b. Community detection (Fortunato, 2010); Modularity.

Graph Neural Network

Graph Neural Networks (GNNs) represent a family of neural network architectures specifically tailored to address challenges posed by graph-structured data (Scarselli et al., 2009). Notable models include Relational Graph Convolutional Network (RGCN), Graph Transformer, and Heterogeneous Graph Transformer (HGT). The explanation is as follows.

a. RGCN (Schlichtkrull, 2017) is adept at capturing relational dependencies within graphs, crucial for tasks such as relational data analysis and knowledge graph reasoning.
b. Graph Transformer (Yun, 2019) extends the transformer architecture to process homogeneous graphs, excelling in tasks such as node classification and link prediction by leveraging self-attention mechanisms to capture both global and local contexts within the graph.
c. Heterogeneous Graph Transformer/ HGT (Hu, 2020) specifically targets the complexities of heterogeneous graphs, where nodes and edges belong to different types.

Table 1. **Preview of Sample Space**

No.	Source v1	Target v2	Trans Amount v3	Description v4
1	*****403-customer	******195-merchant	1000000	Buy Goods
n	…	…	…	…

Source: Processed Data

In this paper, we selected the Heterogeneous Graph Transformer (HGT) method for our GNN model due to its comparatively superior performance in assessing relationships between all nodes, regardless of transaction presence, and is tailored to handle heterogeneous data structure as demonstrated in Figure 1. Moreover, HGT incorporates hierarchical attention mechanisms to capture nuanced patterns across diverse entities within the graph, which enhances its interpretability and discriminative power compared to conventional models. The effectiveness of HGT in fraud detection is highlighted by Ghosh, (2023), underscoring its superiority in identifying collusion-driven fraud schemes. The distinctive advantages and operational differences between the models have been highlighted by Hu, (2020), Sun, (2024), summarized in Table 2.

The HGT operates through a mechanism that leverages node-specific and edge-specific parameters to effectively learn and represent heterogeneous data using a multi-headed attention mechanism, which then adapted to handle the heterogeneity in the graph. The formula for the attention mechanism in HGT can be expressed as follows:

$$\alpha_{uv} = softmax\left(\frac{(h_u W^Q)(h_V W^K + W^K_{\phi(uv)})}{\sqrt{d}}\right)$$

Where:

a. α_{uv} is the attention coefficient between nodes u and v.
b. h_u and h_V are the feature vectors of nodes u and v.
c. W^Q and W^K are the query and key transformation matrices for the attention mechanism.

Table 2. **Feature Comparison between Graph Neural Network (GNN) Analytic Methods**

No.	Feature	RGCN	Graph Transformer	Heterogeneous Graph Transformer (HGT)
1	Feature Incorpo-ration	Limited to relational features between direc-tly connected nodes (en-tities that have direct transactions).	Flexible, can incorporate various features (both direct and indirect relationships), but limited to a similar type/ homogeneous	Flexible, accommodates di-verse features (both direct and indirect relationships) across different types, i.e. cus-tomers, merchants.
2	Interpre-tability	Layers may lack interpre-tability (black box).	Provides insights via attention mechanisms (weight used to model the relationships)	Offers interpretability thro-ugh hierarchical attention mechanisms and node embeddings (attributes).
3	Scala-bility	Performs well on small to medium-sized graphs.	Can handle large-scale graphs, where networks can be densely connected.	Scalable to large and complex graphs (i.e., millions of tran-sactions and diverse entity types).
4	Perfor-mance	Effective for relational learning tasks (direct tran-saction).	Effective for capturing global and local patterns (subtle anomalous transactions).	Effective for handling diverse data representations and complex patterns (layers of transactions).
5	Suitable Applica-tions	Relational data analysis, knowledge graphs	Natural language processing, node classification.	Fraud detection, recommen-dation systems, and know-ledge graphs.

Source: Processed Data

d. $W^{k}_{\phi(uv)}$ is the edge-type-specific transformation matrix for the edge connecting u and v.
e. $\phi(uv)$ represents the type of edge between nodes u and v.
f. d is the dimensionality of the key vectors, used for normalization.

The attention scores are calculated using the softmax function to normalize the scores across all choices of u for each v, ensuring that they sum to one and highlight the most relevant connections for each node based on the heterogeneous context.

Following the computation of attention scores, the node features are updated by aggregating the features of neighboring nodes weighted by these scores, often incorporating transformation matrices that are specific to the type of node and edge.

$$h'_{u} = \sigma(\Sigma_{v \in N(u)}) \, \alpha_{uv} \, (h_{v} \, W^{V} + W^{v}_{\phi(uv)}))$$

Where;
a. h'_{u} is the updated feature vector of node u,
b. W^{V} and $W^{v}_{\phi(uv)}$ are the value transformation matrix and the edge-type-specific value transformation matrix.
c. N(u) is the set of neighbors of node u.
d. σ is a non-linear activation function, such as ReLU.

This methodology allows the HGT to dynamically adapt to the varying types of nodes and connections in the graph, effectively capturing the unique characteristics and relationships within heterogeneous data, which is crucial for tasks such as fraud detection in complex transaction networks.

Result parameters

Syndicate Score Calculation

The Syndicate Score uses normalized centrality measures of nodes, specifically on in-degree and out-degree. These metrics are normalized on scale of 0 and 1. The score is calculated as a weighted average,

Syndicate Score = 0.4 × Normalized InDegree + 0.4 × Normalized OutDegree + OutDegree + 0.2 × Normalized Betweenness Centrality.

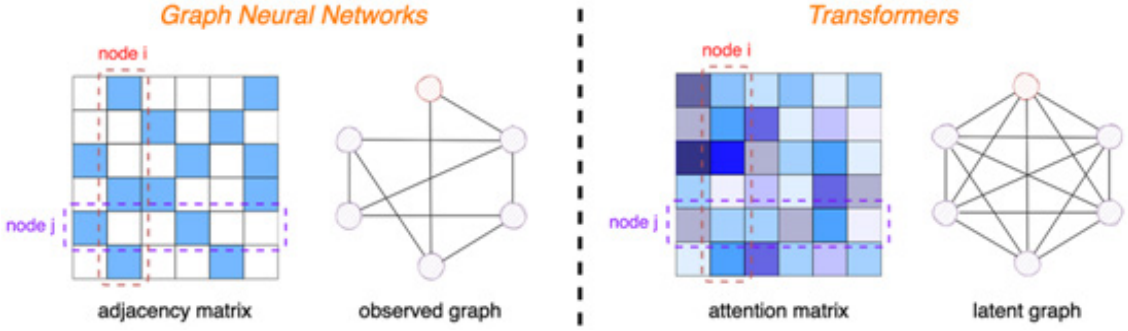
This score is then rescaled to ensure all values are between 0 and 1 by dividing by the maximum syndicate score found in the dataset and integrated back into the dataset.

Cluster Risk Score Calculation

The Cluster Risk Score is calculated using the Syndicate Score (which integrates normalized centrality measures such as in-degree, out-degree, and betweenness centralities) and is further refined by applying heuristic adjustments based on hit counts and predefined rules from past fraud cases, leveraging models like XGBoost for enhanced accuracy. Node scores are averaged within their respective clusters to calculate the cluster risk score. These cluster scores are categorized into risk levels—low, medium, and high—using the 33rd and 67th percentile thresholds:

Risk Level= {"high" if score ≥ 67th percentile;"medium" if score > 33rd percentile; "low" otherwise}

Figure 1. **Comparison between Graph Neural Network (GNN) and Graph Transformers**



Source: Clifford & Ferroukhi, 2021

This method segments nodes and clusters into manageable risk categories for targeted analysis and intervention in scenarios susceptible to fraud, such as financial networks.

Node Count

The Node Count for each cluster is a quantified number of nodes that belong to each distinct cluster identified within the network to provide insight into the size and potential influence of each cluster within the overall graph structure. This is computed by grouping nodes based on their cluster assignment (e.g., obtained from a clustering algorithm like HGT) and counting the number of nodes within each group. The result is a simple count that reflects how populated each cluster is to understand the distribution of nodes across different risk levels and identifying particularly dense areas that might require additional scrutiny or have a significant impact on network dynamics.

Supervised Learning

Supervised learning is a technique to predict or classify data based on a model of prior labeled examples. In the domain of fraud detection, where data sets are often complex and imbalanced, Gradient Boosting is particularly effective. This ensemble learning method, detailed by Alothman (2022), enhances predictive accuracy by sequentially combining multiple weak learners into a strong model. Each iteration focuses on correcting the errors of the previous one, gradually improving the model's ability to discern between fraudulent and non-fraudulent transa-ctions. The mathe-matical expression for Gradient Boosting can be represented as follows:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Where:

a. $F_m(x)$ is the model's prediction at iteration.
b. $F_{m-1}(x)$ is the model's prediction from the previous iteration.
c. γ_m is the learning rate at iteration.
d. h_m is the weak learner added at iteration to correct the residuals.

Integrating Gradient Boosting with Graph Neural Networks (GNNs), specifically in the context of HGT, further augment fraud detection capabilities by using supervised learning, initially by learning to represent and classify nodes based on their features and the graph structure. Post initial training, Gradient Boosting can be applied to refine the HGT model (where the model's predictive performance was not optimized). With each iteration, the combined model (HGT with Gradient Boosting) adjusts its parameters to optimally minimize errors.

Statistical Evaluation

The author compared two samples using Student's (1908) *t-test* as an inferential to test if two populations are *statistically* & *significantly* different. Furthermore, Welch's (1947) t-test will be used as the samples are non-overlapping (fraudulent sample to non-fraudulent sample) and the variances are not expected to be unequal. As a context, the author used this method to compare the extracted EDA (Exploratory Data Analysis) findings between the fraud and non-fraud sample pool datasets. The formula for Welch's t-test is as follows:

$$t = \frac{\underline{X}_1 - \underline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where;

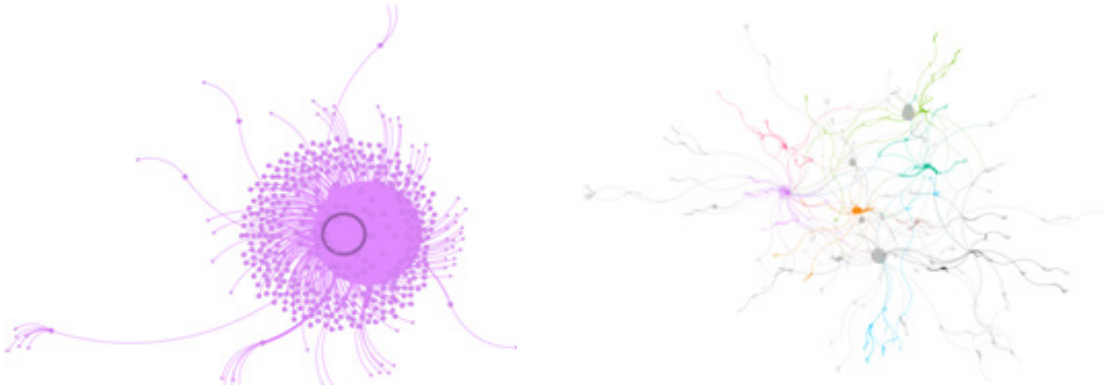
a. $\underline{X}_1$ and $\underline{X}_2$ are the means of the first and second groups, respectively.
b. s_1^2 and s_2^2 are the variances of the first and second groups, respectively.
c. n_1 and n_2 are the sample sizes of the first and second groups, respectively.

Analytical Pipeline Framework

The author outlined the following structured, staged method to identify and analyze fraud syndicates:

a. Data Acquisition: fraudulent and non-fraudulent instances of data.
b. Data Cleaning & Preprocessing: Remove noise, inconsistencies, and duplicates.
c. Feature Engineering: Extract and transform parameters to optimize inputs for modeling techniques.
d. HGT model and Network Analysis: Generate node embeddings and identify key features.
e. Combination of Metrics: Synthesize insights from both GNN and Network Analysis for a robust analysis.
f. Exploratory Data Analysis (EDA): Integrate insights to predict node behavior and potential link predictions.
g. Two Samples Paired t-test: Inferential test for statistical differences.
h. Heuristic Rule Development: Practical guidelines based on the findings to identify fraudulent patterns and inform decision-making

Figure 3. **Snapshot of Comparative Graph Network Maps, Fraud and Non-Fraud**



Observed Fraud Syndicate Observed Network of Non-Fraudulent Nodes

Source: Data Processing Using Gephi Network Visualization Software.

Table 3. **Observation of Profile Differences Between Fraud Versus Non-Fraud**

Profile	Fraud	Non-Fraud
Centrality	**High-Centrality**; high centrality, few nodes as central hubs and focal points through which many transactions are passed through. Apparent gatekeepers' presence.	**Dispersed-Connections**; characterized by more evenly distributed connections lacking prominent hubs, non-dependent use cases.
Clustering	**Dense Clustering**: higher clustering coefficient, tight-knit groups, obscuring detection. Indicators of repetitive, circular transaction patterns.	**Lower, Scattered Clustering**; lower clustering coefficient, less frequent community in normal business or social contexts.
Path lengths	**Short, Uniformed Path Lengths**; display shorter average path lengths between nodes compared to non-fraudulent networks, indicating closer and more direct/focused interaction among involved parties, which facilitates rapid coordination and execution of fraudulent activities.	**Varied Path Lengths**; vary more widely than in fraudulent networks, as legitimate interactions are not necessarily optimized for speed and coordination like fraudulent schemes networks where neighboring nodes are purpose-driven to move illegitimate funds.

Source: Processed Data

RESULT AND DISCUSSION

Graph Analytics

Figure 3 is created using the 'ForceAtlas2' layout algorithm (Jacomy, 2014) due to its intuitiveness and gravity-repulsion force approach in representing a social network. The layout of the graph mapping demonstrates certain different characteristics between fraudulent and non-fraudulent networks. Table 3 denotes the observed differences in profiles between fraud and non-fraud network mapping.

A more detailed and in-depth explanation for the result parameters of each fraud and non-fraud sample space's graph analytics can be viewed in table 4, table 5, and table 6.

Each group is classified into four clusters based on differences in observed patterns. This cluster's nodes are crucial for the flow of transactions, indicating a core role in the fraud syndicate. Within the discourse of Financial Crime and Money Laundering, these nodes act as the gatekeepers that control the flow of illegitimate funds (Utama, 2016).

Figure 4 highlights structural differences between fraudulent and non-fraudulent behaviors. The fraud-related network on the left is highly centralized, featuring a spoke-hub pattern with a few key nodes linked to many peripheral nodes, suggesting a design tailored for fraudulent activities. These peripheral

Table 4. **Fraud Syndicate and Cluster Score**

No.	HGT cluster	Syndicate Score	Cluster Risk Score	Cluster Risk Level	Node Count
1.	0	0.12	0.15	Low	2563
2.	1	0.54	0.67	Medium	11987
3.	2	0.76	0.82	Medium	553
4.	3	0.93	0.97	High	625

Source: Processed Data

Table 5. **Non-Fraud Syndicate and Cluster Score**

No.	HGT cluster	Syndicate Score	Cluster Risk Score	Cluster Risk Level	Node Count
1.	0	0.05	0.02	Low	10233
2.	1	0.15	0.10	Low	89795
3.	2	0.21	0.20	Low	11783
4.	3	0.32	0.35	Medium	3007

Source: Processed Data

Table 6. **Comparison between fraud and non-fraud clusters**

Cluster	Fraud Syndicate Analysis	Non-Fraud Syndicate Analysis
0	Lowest activity, Syndicate Score: 0.12, minimal degree and betweenness centralities, peripheral with insignificant risk.	Mostly inactive nodes, Syndicate Score: 0.05, minimal centrality, low node engagement, negligible risk.
1	Largest cluster, Syndicate Score: 0.54, low degree centrality, medium risk.	Holds bulk of nodes, Syndicate Score: 0.15, sparse interactions, typical of regular customers, low risk.
2	Higher risk, Syndicate Score: 0.76, higher transaction volume, limited external influence.	Slightly more frequent transactions, Syndicate Score: 0.21, somewhat higher centrality, low risk.
3	Highest risk, Syndicate Score: 0.93, significant network centrality, core role in fraud activities.	Most active in legitimate scenarios, Syndicate Score: 0.32, central role in transactions, medium risk level.

Source: Processed Data

nodes, with fewer connections, likely engage mainly with the central fraudulent node, underlining their specific roles in fraud operations.

In contrast, the right graph depicts a non-fraudulent network that is more organic and interconnected, with densely connected nodes facilitating legitimate transactions among a broad participant group. This network does not show the isolated clusters seen in the fraud network but instead exhibits integrated clusters indicative of genuine financial activities. Consistent with He et al. (2024), fraud networks are typically more isolated and concentrated in areas of low homogeneity, distinguishing them sharply from the more interconnected and diverse non-fraudulent networks.

Exploratory Data Analysis (EDA)

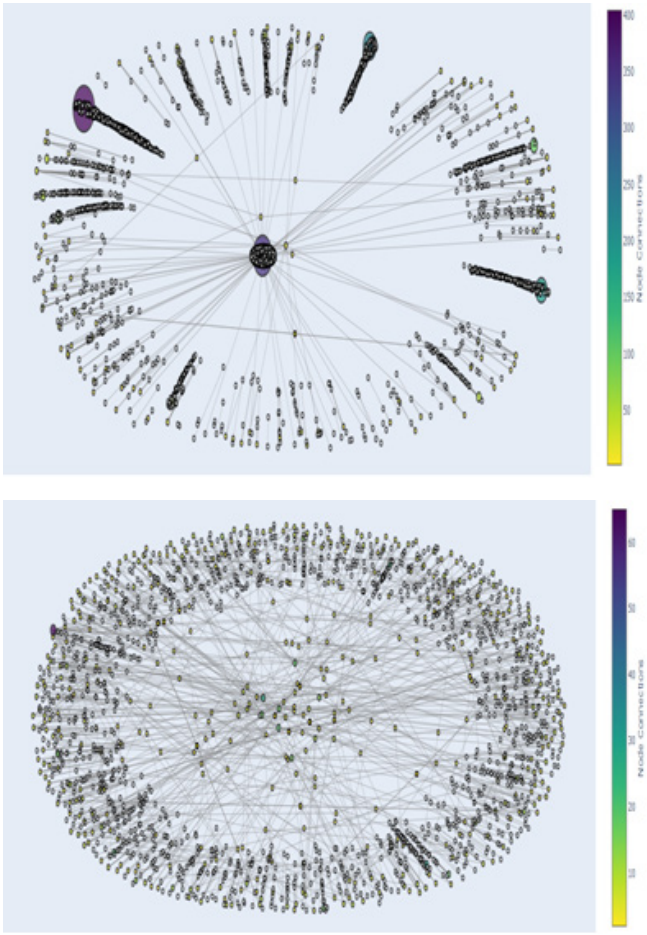
The acquired data of fraud and non-fraud Exploratory Data Analysis (EDA) results and their corresponding description are each shown in the following sections.

Fraud Exploratory Data Analysis:

Explanation for each fraud cluster (Table 7) is as follows.

a. Cluster 0: Includes high-frequency transactions like "Buy Goods", with "Business Adjustment" and "Transaction Reversal" as less frequent but significant activities,
b. Cluster 1: Dominated by "Business to Business Transfer" along with "Buy Goods", "Off-us Merchants", and "Business Adjustment", representing varied commercial activities,

Figure 4. **Fraud vs Non-Fraud Cluster Distribution**



Source: Data Processing Using Python Graph Package

c. Cluster 2: Primarily involves "Off-us" (outside of registered e-wallet ecosystem) transactions with a smaller amount of "Pay Bill" transactions, indicating specialized service transactions.
d. Cluster 3: Exclusively handles transactions labeled as "Commodity," which refer to a specific type of service or merchant central to this cluster's activities. Note that the targeted product of purchase is an easily convertible/cashable product and is commodifiable.

Most clusters include an "Off-us" transaction type which is a transaction going outbound from the observed e-wallet. This fits the widely recognized Money Laundering typology of *placement, layering,* and *integration* that illegitimate funds flowed outside of the ecosystem channels.

Non-Fraud Exploratory Data Analysis:
Explanation for each non-fraud cluster (Table 8) is as follows.

a. Cluster 0 focuses on TelcoMerchant B2B transactions and payment services like 'Transportation', 'Cash In', 'Pay Bill ', routine commercial activities,
b. Cluster 1 spans a range of B2B, customer depositing for a balance top-up, and peer-to-peer transactions. Diverse business interactions and financial services,
c. Cluster 2 emphasizes customer-oriented services, particularly in tele-communications, with a focus on reservations and payments, show-casing a service-specific transaction environment.
d. Cluster 3 centers on digital transaction services of Business-To-Business and related fees, demonstrating a specialized financial transaction use-case focus within the network.

Table 7. **Fraud Attributes and Transactions Statistical Distributions per Cluster**

HGT Cluster	Attributes	Transaction Amount (in IDR)	Transaction Type
0	Median = { 'degree_centrality': 0.70, 'in_degree_centrality': 0.69, 'out_degree_centrality': 0.01, 'betweenness_ centrality': 0.009 }	{'min': 500, 'median': 15000, 'max': 8000000, 'avg': 200000 }	description = { 'count of distinct transaction types': 3, 'common transaction type': 'Buy Goods', 'other types': 'Business Adjustment', 'Transaction Reversal'}
1	Median = { 'degree_centrality': 0.30, 'in_degree_centrality': 0.29, 'out_degree_centrality': 0.01, 'betweenness_ centrality': 0.005 }	{'min': 1000, 'median': 7000, 'max': 3000000, 'avg': 180000 }	description = { 'count of distinct transaction types': 4, 'common transaction type': 'Business to Business Transfer', 'other types': 'Buy Goods', 'Off-us Merchants', 'Business Adjustment'}
2	Median = { 'degree_centrality': 0.50, 'in_degree_centrality': 0.48, 'out_degree_centrality': 0.02, 'betweenness_ centrality': 0.010 }	{'min': 100, 'median': 30000, 'max': 5000000, 'avg': 250000 }	description = { 'count of distinct transaction types': 2, 'common transaction type': 'Off-us Merchants', 'other types': 'Pay Bill'}
3	Median = { 'degree_centrality': 0.85, 'in_degree_centrality': 0.83, 'out_degree_centrality': 0.02, 'betweenness_ centrality': 0.015 }	{'min': 10000, 'median': 500000, 'max': 8775000, 'avg': 337494 }	description = { 'count of distinct transaction types': 1, 'common transaction type': 'Commodity }

Source: Processed Data

Table 8. **Non-Fraud Attribute and Transaction Statistical Distributions per Cluster**

HGT cluster	Attributes	Transaction Amount (in IDR)	Transaction Type
0	Median = {'degree_ centrality': 0.15, 'in_ degree_centrality': 0.14, 'out_degree_centrality': 0.01, 'betweenness_ centrality': 0.001}	{'min': 10, 'median': 5000, 'max': 150000, 'avg': 20000}	{'count of distinct transaction types': 7, 'common transaction type': 'TelcoMerchant B2B Transfer In Cluster', 'other types': 'Cash In, TelcoMerchant B2B Transfer Fee, Customer Buy Goods, Pay Bill, Customer Withdrawal, Customer Pay Transportation'}
1	Median = {'degree_ centrality': 0.25, 'in_ degree_centrality': 0.24, 'out_degree_centrality': 0.01, 'betweenness_ centrality': 0.002}	{'min': 20, 'median': 16730, 'max': 1000000, 'avg': 45000}	{'count of distinct transaction types': 8, 'common transaction type': 'Business to Business Transfer', 'other types': 'P2P Transfer, B2B Transfer Telco, General B2B Transfer, Customer Buy Goods Reservation, Business Adjustment, APPLINK, General Customer Deposit'}
2	Median = {'degree_ centrality': 0.40, 'in_ degree_centrality': 0.38, 'out_degree_centrality': 0.02, 'betweenness_ centrality': 0.003}	{'min': 50, 'median': 25000, 'max': 2000000, 'avg': 51393}	{'count of distinct transaction types': 6, 'common transaction type': 'Customer Buy Goods', 'other types': 'Customer Reservation Telco For Other, Customer Reservation Telco For Self, Customer Online Payment, Distributor TelcoMerchant B2B Transfer In Cluster, Customer Buy Goods Reservation'}
3	Median = {'degree_ centrality': 0.55, 'in_ degree_centrality': 0.53, 'out_degree_centrality': 0.02, 'betweenness_ centrality': 0.005}	{'min': 100, 'median': 40000, 'max': 5000000, 'avg': 80000}	{'count of distinct transaction types': 5, 'common transaction type': 'TelcoMerchant B2B Transfer', 'other types': 'TelcoMerchant B2B Transfer Fee, Distributor TelcoMerchant B2B Transfer, Cash Out, Buy Telco Product'}

Source: Processed Data

Table 9. **Welch's Paired t-test Result**

No.	Data	t score	p value	
1.	Eccentricity	-37.431	< 0.00000000000000022**	|-4.767507|
2.	Closeness centrality	-37.43	0.00000000000004124**	|0.06555835|
3.	Betweenness centrality	-8.3836	< 0.00000000000000022**	|-0.0009419782|
4.	Harmonic closeness centrality	6.5413	0.00000000007945**	|0.05639332|
5.	Weighted Degree	-3.2809	0.001055*	|-1.354062|
6.	Eigenvector centrality	4.608	0.000004353**	|0.004862323|

Source: Data Processing Using R 'stats' Package.

Sample Comparison: Welch's Paired t-test ($\alpha = 0.05$).

There is enough evidence to reject the null hypothesis and accept that there are sufficient statistically significant differences of variance between fraud and non-fraud networks (rejected). The proposed framework can provide sufficient signals for the detection & prevention of fraud syndicates within the digital payment ecosystem. Further exhibited that all the appointed graph statistic parameters showed the capacity to prove inherent differences between fraud and non-fraud networks (Table 9).

5. CONCLUSION

This paper aims to test and deliver a composite approach in detecting the prevalence of fraud syndicates within a digital payment system. This paper also attempts to develop and validate a framework using graph analytics and machine learning to understand fraud behaviors. This significantly contributes to enhancing security measures within digital financial transactions by providing validated methods of fraud and money laundering detection. The findings confirmed the hypothesis that fraud networks are significantly distinctive from non-fraud in terms of a more centralized and isolated network compared to the organic, interconnected behaviors of non-fraudulent users. The structural differences identified align with existing literature by He, (2024), which emphasizes distinct network behaviors in fraudulent versus legitimate transactions in terms of their homogeneity. Some subclusters showed lower predictability (e.g., cluster 0 fraud), possibly due to the adaptive nature of fraud schemes not captured by static models. Implementing the study's findings can help financial risk decision makers and policymakers to enhance fraud detection systems, improving the protection, integrity, and security of customers and digital transactions. However, limitations exist in our study, such as the specificity of the data and the static nature of the models against evolving fraud tactics. Future researchers could explore adaptive models that respond in real-time to changing fraud strategies, potentially integrating technologies like blockchain for more dynamic fraud prevention mechanisms.

REFERENCES

ACFE. (2024). *Fraud 101: What is Fraud?*. ACFE. https://www.acfe.com/fraud-resources/fraud-101-what-is-fraud.

Alekhin, S., Shmatenko, L. (2018). *Corruption in Investment Arbitration - It Takes Two to Tango. New Horizons of International Arbitration.* Association of Private International and Comparative Law Studies.

Alothman, R., Talib, H. A., & Mohammed, M.S. (2022). Fraud detection under the unbalanced class based on gradient boosting. *Eastern-European Journal of Enterprise Technologies, 2*(2), 6–12. https://doi.org/10.15587/1729-4061.2022.254922.

Bank Indonesia. (2019). *Blueprint Sistem Pembayaran Indonesia 2025 Bank Indonesia: Menavigasi Sistem Pembayaran Nasional di Era Digital.* Bank Indonesia. https://www.bi.go.id/id/fungsi-utama/sistem-pembayaran/blueprint-2025/default.aspx#:~:text=BSPI%2025%20adalah%20arah%20kebijakan,era%20ekonomi%20dan%20keuangan%20digital.

Berge, C. (1958). *Théorie des graphes et ses applications* [English: The theory of graphs and its applications]. Wiley.

Black, H. C. (2018). *Legal Dictionary: Fraud. Black's Law Dictionary, 8th Edition.* Law.com.

Borgatti, S. P., Everett, M. G., Johnson, J. C. (2013). *Analyzing Social Networks.* SAGE Publications Ltd.

C. Xu and J. Zhang. (2015). Towards Collusive Fraud Detection in Online Reviews. *2015 IEEE International Conference on Data Mining, Atlantic City, NJ, USA*.

Campana, P., Antonopoulos, G. A. (2024). A Relational Approach to Organised Crime. *Trends in Organized Crime, 27*, 229–234. https://doi.org/10.1007/s12117-024-09534-4.

Chiesi, A. M. (2001). Network Analysis. *International Encyclopedia of the Social & Behavioral Sciences.* https://doi.org/10.1016/B0-08-043076-7/04211-X.

Cosmin, D., Polkowski, Z., & Gruber, Jacek. (2013). *E-Fraud*. DWSPiT Polkowice.

Dominik Olszewski. (2014). Fraud Detection Using A Self-Organizing Map to Visualize the User Profiles. *Knowledge-Based Systems, 70*, 324-334. https://doi.org/10.1016/j.knosys.2014.07.008.

Euler, L. (1736). Solutio Problematis Ad Geometriam Situs Pertinentis. *Commentarii academiae scientiarum Petropolitanae, 7*(1), 128-140.

Europol. (2023). *Online Fraud Schemes: A Web of Deceit*. Internet Organised Crime Threat Assessment (IOCTA).

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., & Monfardini, G. (2009). The Graph Neural Network Model. *IEEE Transactions on Neural Networks, 20*, 61-80.

FATF – Interpol - Egmont Group. (2023). Illicit Financial Flows from Cyber-Enabled Fraud, FATF, Paris, France. Link: fatf-gafi.org/content/fatf-gafi/en/publications/Methodsandtrends/illicit-financial-flows-cyberenabled-fraud.html

Felson, M. (2003). *The Process of Co-offending.* Martha Smith and DerekCornish (eds). Criminal Justice Press.

Gartner. (2022). *Market Guide for Graph Database Management Systems*.Gartner Research. https://www.gartner.com/en/documents/4018220.

Ghosh, S., Anand, R., Bhowmik, T., & Chandrashekhar, S. (2023). GoSage: Heterogeneous Graph Neural Network Using Hierarchical Attention for Collusion Fraud Detection. *ICAIF 23, Proceedings of the Fourth ACM International Conference on AI in Finance*. https://doi.org/10.1145/3604237.3626856

Girvan M., Newman M. E. J. (2002). Community Structure in Social and Biological Networks. *Proc. Natl. Acad. Sci. USA, 99*(12), 7821–7826

Grandjean, M. (2016). Applications of Graph Theory. *Proceedings of the IEEE, 106*(5), 784 - 786.

He, E., Hao, Y., Zhang, Y., Yin, G., & Yao, L. (2024). SCALA: Sparsification-based Contrastive Learning for Anomaly Detection on Attributed Networks. *Social and Information Networks.*

Hu, Z., Dong, Y., Wang, K., & Sun, Y. (2020). Heterogeneous Graph Transformer. *Proceedings of The Web Conference 2020*. https://doi.org/10.1145/3366423.3380027.

Islam, A. et al. (2011). Detecting Collusive Fraud in Enterprise Resource Planning Systems. In: Peterson, G., Shenoi, S. (eds) Advances in Digital Forensics VII. Digital Forensics 2011. *IFIP Advances in Information and Communication Technology, 361*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24212-0_11.

J. J. Sylvester. (1878). On an Application of The New Atomic Theory to The Graphical Representation of The Invariants and Covariants of Binary Quantics, with Three Appendices. *American Journal of Mathematics, 1*(1), 64-90.

Jacomy, M., Venturini, T., Heymann, S., Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization, Designed for the Gephi Software. *PLoS ONE, 9*(6): 1-12. https://doi.org/10.1371/journal.pone.0098679

Levi, M. (2008). Organized Fraud and Organizing Frauds: Unpacking Research on Networks and Organization. *Criminology and Criminal Justice, 8*(4), 389-419. http://dx.doi.org/10.1177/1748895808096470.

Li, Z., Fang, X., Sheng, O. R. (2017). A Survey of Link Recommendation for Social Networks: Methods, Theoretical Foundations, and Future Research Directions. *ACM Transactions on Management Information Systems, 9*(1), 1–26. https://doi.org/10.1145/3131782.

Mastercard. (2024). *E-commerce Fraud Trends and Statistics Merchants Need to Know in 2024*. https://b2b.mastercard.com/news-and-insights/blog/ecommerce-fraud-trends-and-statistics-merchants-need-to-know-in-2024/.

Nasteski, V. (2017). An Overview of the Supervised Machine Learning Methods. *HORIZONS. B. 4*, 51-62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

Rodrigues, C., Jain, M., R., Khanchandani, A. (2023). Performance Comparison of Graph Database and Relational Database. *Technical Report,* http://dx.doi.org/10.13140/RG.2.2.27380.32641.

S. Fortunato. (2010). Community Detection in Graphs. Physics Reports, 486, 1-103.

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI. 2*, 160. https://doi.org/10.1007/s42979-021-00592-x

Shulzhenko, Nadiia. (2020). Internet Fraud and Transnational Organized Crime. *Juridical Tribune - Review of Comparative and International Law, Bucharest Academy of Economic Studies, 10*(1), 162-172.

Snedecor, W. & Cochran, W. (1989). *Statistical Methods.* 8th Edition, Iowa State University Press.

Sun, Y., Zhu, D., Wang, Y., & Tian, Z. (2024). GTC: GNN-Transformer Co-contrastive Learning for Self-supervised Heterogeneous Graph Representation. arXiv, 2403.15520. https://doi.org/10.48550/arXiv.2403.15520

Tremblay, P. (1993). *Searching for Suitable Co-offenders*. Routledge.

Utama, Paku. (2016). Gatekeepers' Roles as a Fundamental Key in Money Laundering. *Indonesia Law Review: 6*(2), 180-206. https://doi.org/10.15742/ilrev.v6n2.215.

Welch, B. L. (1947). The Generalization of "Student's" Problem when Several Different Population Variances Are Involved. *Biometrika. 34* (1–2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28.

World Bank. (2022). *Press Release: "COVID-19 Drives Global Surge in Use of Digital Payments". The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19.* https://www.worldbank.org/en/news/press-release/2022/06/29/covid-19-drives-global-surge-in-use-of-digital-payments.

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., Li, J. (2021). Intelligent Financial Fraud Detection Practices in The Post-Pandemic Era. *The Innovation 2(4)*, 1-11 November 28, 2021. https://doi.org/10.1016/j.xinn.2021.100176.

Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph Transformer Networks. Cornell University, arXiv, 1911.06455. https://doi.org/10.48550/arXiv.1911.06455.

Figure 2. **Analytical Process Pipeline Diagram**