Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

DOI: 10.37943/22ZSKI6025

Rustem Baigabyl
Master's student, MBA "Financial technology and digital transformation of business"
r.baigabyl@astanait.edu.kz, orcid.org/0009-0002-1741-8078
Astana IT University, Kazakhstan

Aliya Nugumanova
PhD, Head of the Scientific and Innovation Center «Big Data and Blockchain Technologies»
a.nugumanova@astanait.edu.kz, orcid.org/0000-0001-5522-4421
Astana IT University, Kazakhstan

Mariia Sodnomova
Bachelor's degree, BS "IT-Entrepreneurship"
222288@astanait.edu.kz, orcid.org/0009-0006-4035-9618
Astana IT University, Kazakhstan

USING GRAPH CENTRALITY METRICS FOR DETECTION OF SUSPICIOUS TRANSACTIONS

Abstract: Detecting suspicious transactions remains a persistent challenge due to increasingly sophisticated methods of money laundering and fraud within modern financial systems. This study introduces a graph-based analytical framework utilizing social network analysis for identifying potentially illicit transactions. Financial entities (individuals or institutions) are represented as nodes connected by directed edges symbolizing transactions. Applying centrality measures – including degree, betweenness, closeness, and eigenvector centrality - we quantify each node's influence and involvement in financial flows. Nodes exhibiting high betweenness and degree centrality values emerge as potential 'bridges,' controlling significant transaction pathways. Our analysis demonstrates these high-centrality entities often mediate substantial transactional volumes or integrate otherwise disconnected sub-networks, highlighting them as prime targets for investigation. Visualizing localized subgraphs around these pivotal nodes further uncovers densely interconnected structures suggestive of hidden clusters involved in complex money-laundering operations. Integrating our method with real-time machine learning analytics significantly enhances both the speed and precision of suspicious account detection. Empirical validation using anonymized banking data illustrates that centrality-based screening enables proactive identification of anomalous patterns, substantially improving traditional, reactive anti-money laundering measures. This approach not only addresses existing gaps - such as the static nature of traditional supervised learning models - but also overcomes computational and scalability barriers characteristic of prior advanced techniques. Additionally, the proposed approach provides enhanced interpretability, supporting compliance officers in making informed decisions. The findings emphasize the necessity of continuously adapting analytical techniques to emerging threats. Ultimately, our research provides financial institutions with robust, actionable tools for early-stage fraud detection, underscoring graph analytics as vital to financial security within our interconnected global economy.

The relevance of this research stems from the increasing complexity of financial fraud, which often evades traditional rule-based detection systems. The study aims to address this gap by employing graph-based techniques that allow for the structural analysis of transac-

Received: 22.02.2025 Accepted: 26.06.2025 Published: 30.06.2025

DOI: 10.37943/22ZSKI6025

tion networks. Centrality metrics are used to identify key actors whose positions may indicate hidden coordination or anomaly patterns. A directed transaction graph is constructed from synthetic financial data, and a set of experiments is conducted to compute centrality measures, extract subgraphs, and visualize network topology. The results are evaluated in comparison with machine learning benchmarks to assess the effectiveness and interpretability of the proposed approach.

Keywords: social network analysis, centrality measures, financial fraud detection, betweenness centrality, anti-money laundering, transaction networks, graph-based anomaly detection, explainable AI.

Introduction

Modern financial systems are increasingly reliant on high-speed, digital transaction processes, making them more efficient yet simultaneously vulnerable to sophisticated schemes of fraud and money laundering. A notable historical case that underscores these vulnerabilities is the collapse of the Bank of Credit and Commerce International (BCCI) [1]. Due to insufficient transparency and inadequate oversight mechanisms, the BCCI scandal illustrated how multinational financial institutions could conceal illicit activities for prolonged periods. Although technological advancements now allow for real-time monitoring and data analysis, criminal networks continue to evolve, capitalizing on decentralized structures and multiple layers of intermediaries.

In this context, methods based on graph analytics and social network analysis (SNA) have emerged as powerful tools for identifying suspicious nodes and transaction flows in complex financial networks. By representing senders, beneficiaries, and the corresponding monetary transactions as nodes and edges, respectively, researchers can leverage centrality metrics – such as degree, betweenness, closeness, and eigenvector – to detect anomalies, pinpoint pivotal actors, and assess the overall resilience of the network. This approach enables financial institutions and regulatory authorities not only to analyze historically known fraud patterns but also to proactively identify new, undiscovered threats. Consequently, the integration of graph-based methodologies into anti-money laundering (AML) systems holds significant promise for safeguarding financial integrity and preventing large-scale abuses similar to those exemplified by BCCI [1].

This perspective is aligned with the international guidelines of the Financial Action Task Force (FATF), which emphasize the importance of adopting proactive, data-driven monitoring systems in financial institutions to counter money laundering and terrorist financing [2].

Literature review and problem statement

The historic collapse of the Bank of Credit and Commerce International (BCCI) revealed astonishing levels of institutionalized fraud, as investigators discovered that the bank's sprawling transnational network facilitated unauthorized fund transfers on multiple continents [3]. Because regulatory bodies struggled to coordinate their oversight across diverse jurisdictions, BCCI's top executives successfully obscured illicit transactions behind intricate layers of corporate structures. These findings underscored just how critical international regulatory coordination is to prevent large-scale money laundering, while also highlighting the inherent vulnerability of financial institutions operating in rapidly globalizing markets. In an academic context, BCCI's case serves as an exemplar of systemic failure: the sheer volume of illicit activity that occurred – even under partial regulatory scrutiny – illuminated the pressing need for advanced detection methods capable of transcending traditional auditing mechanisms.

Building on the lessons from this significant financial scandal, subsequent scholarship on anti-money laundering (AML) and fraud detection pivoted toward sophisticated, data-driven

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

techniques. Ahmed et al. [4] systematically reviewed anomaly detection methodologies in the financial sector, concluding that single-method approaches – such as basic statistical thresholds or rule-based systems – struggle to adapt to constantly evolving criminal strategies. Their meta-analysis showed that while simpler models were sometimes easier to interpret, they frequently missed subtle anomalies or generated spurious alerts, thereby reducing operational effectiveness. This indicated that more adaptive solutions, integrating machine learning or hybrid detection strategies, were necessary to mitigate both false negatives (overlooking genuine fraud) and false positives (flagging benign transactions).

Huang et al. [5] further advanced the conversation by applying a machine-learning-based k-means clustering algorithm to financial datasets, revealing distinct transaction "clusters" suggestive of anomalous behavior. Empirically, they demonstrated that the algorithm could effectively adapt to newly emerging transaction patterns, a crucial advantage given the dynamic tactics employed by illicit actors. However, their study also reported substantial computational overhead in high-volume environments, where continuously streaming data demanded frequent model updates. From an academic perspective, these results underscore a fundamental tension in AML research: while increasingly complex algorithms can improve detection accuracy, they may also necessitate extensive computing resources and careful parameter tuning. This trade-off highlights the need for scalable system architectures, a balanced approach to model interpretability, and ongoing updates to account for new types of fraudulent activity – all of which remain core challenges in the modern AML landscape.

Subsequent studies increasingly adopted supervised learning paradigms, reflecting a desire to improve the precision and scalability of AML efforts under rapidly changing conditions. Savage et al. rigorously tested Random Forest and Support Vector Machine (SVM) models on transaction-network features, demonstrating robust detection rates for known fraudulent patterns [6]. Notably, their results suggested that these models could accurately classify many high-risk entities, indicating strong predictive power when sufficient historical training data was available. However, an important takeaway was the pronounced performance decline when criminals deployed entirely new schemes or significantly altered their methods—circumstances where the underlying training data no longer mirrored real-world conditions. From an academic standpoint, this underscores a key limitation of supervised models: their reliance on representative training sets makes them vulnerable to evolving criminal activity that does not match existing patterns.

Mirroring these concerns, Sun et al. [7] developed a specialized decision tree algorithm that incorporated a focused feature selection process designed specifically for AML applications. Their study showed that limiting the model to the most relevant transaction variables enhanced both accuracy and interpretability. The decision tree approach allowed compliance officers to trace exactly how the model flagged certain operations, satisfying regulatory demands for transparent, justifiable decisions. However, the authors also cautioned that the decision trees, by nature, can become too simplistic if crucial features are omitted – potentially missing complex behaviors that more sophisticated models might capture.

Building on these supervised foundations, Pambudi et al. [8] introduced principal component analysis (PCA) as a dimensionality-reduction tool, streamlining large input vectors into a smaller set of latent features. Their method improved classification speed and yielded higher overall accuracy, primarily because it eliminated redundant or weakly predictive features that could clutter traditional classifiers. Yet a core limitation emerged: in reducing dimensionality, PCA sometimes obscured outlier points that were critical for detecting rare but highly damaging fraudulent behaviors. This trade-off exemplifies an enduring tension in AML model design: balancing computational efficiency and the capacity to capture exceptional but meaningful data anomalies.

DOI: 10.37943/22ZSKI6025

Along a similar line, Zhu et al. [9] benchmarked artificial neural networks (ANN) against logistic regression in credit card fraud detection, finding that ANN excelled in modeling non-linear relationships within transaction data. Their experiments suggested that neural architectures capture subtle indicators of deception missed by linear or stepwise methods, improving recall in complex, real-world scenarios. However, logistic regression continued to offer better real-time applicability due to its lower computational overhead, an indispensable factor for financial institutions that process hundreds or thousands of transactions per second.

Yet, despite the efficacy of purely attribute-based (or tabular) machine learning systems, researchers increasingly recognize that modeling financial transactions as a network—linking senders, beneficiaries, and intermediaries – can reveal deeper patterns of illicit behavior. Building on this idea, Sousa Lima [10] conducted an extensive study utilizing real-world Brazilian banking data, where he applied social network analysis (SNA) techniques to trace multi-layered schemes orchestrated by shell companies. Notably, his results showed that nodes which appeared innocuous in conventional tabular analyses were, in fact, part of tightly woven clusters funneling suspicious funds across international jurisdictions. By detecting atypical "hub" nodes and suddenly formed network bridges, Sousa Lima's approach substantially reduced false positives, primarily because it captured relational structures that standard pointwise algorithms overlooked. This outcome highlights the critical advantage of examining how specific entities function within a collective web of interactions rather than focusing solely on isolated variables like transaction volume or frequency.

Expanding on the network perspective, Deprez et al. [11] offered a comprehensive overview of state-of-the-art graph-driven anti-money laundering methodologies, surveying 97 scholarly sources from Web of Science and Scopus databases. Their large-scale comparison, employing the Elliptic dataset of 203,769 Bitcoin transactions, revealed two pivotal insights. First, Graph Convolutional Networks (GCNs) achieved higher accuracy in detecting malicious nodes compared to simpler classification techniques, such as logistic regression or decision trees, confirming their superior capacity for modeling intricate relationships in transactional graphs. Second, while GCN-based methods outperformed others by a significant margin in capturing evolving money-laundering patterns, the authors noted steep computational demands, including increased training time and memory usage. This underscores a trade-off routinely encountered in real-world applications: although complex architectures often excel in uncovering hidden clusters or subtle transactional anomalies, they also pose challenges in scalability and operational feasibility – especially in environments requiring real-time or near-real-time screening. Deprez et al. [11] further emphasized that many existing models focus on historically known fraud typologies, raising questions about how effectively they adapt to novels, rapidly mutating criminal behaviors. Academically, these findings suggest that future AML research must balance the algorithmic sophistication afforded by network-based deep learning with the need for efficiency, interpretability, and continuous model updating to stay ahead of increasingly agile money-laundering networks.

Further emphasis on centrality metrics can be found in the study by Gerbrands et al. [12], who investigated multiple real-world money-laundering cases to illustrate how degree, betweenness, and closeness centrality can pinpoint nodes exerting disproportionate influence over financial flows. A major finding was that high-betweenness nodes, or "choke points," often orchestrated significant movement of funds between otherwise isolated network segments. Crucially, the authors noted that such nodes tended to correlate with accounts already flagged by compliance teams, underscoring betweenness centrality's strong empirical grounding in actual detection contexts. Yet, Gerbrands et al. [12] also criticized over-reliance on any single metric—particularly betweenness – because it can overlook cunning, low-degree intermediaries that link sub-networks in ways less obvious than classic "hub" structures. From an academ-

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

ic perspective, this affirms the complexity of money-laundering strategies: some bad actors deliberately minimize overt connectivity while secretly bridging essential transaction paths, a tactic easily missed by simplistic screening strategies.

In a bid to capture such multidimensional risk indicators, Smith and Allen [13] proposed a hybrid detection framework that integrates social network analysis (SNA) with graph-based outlier detection, enabling continuous recalibration of suspicious-entity profiles. Validated against extensive real-world transaction logs, their approach showed that static analyses – regardless of how detailed – often fail to detect subtle, newly emerging schemes once criminals shift tactics or forge unorthodox connections. Instead, periodically updated network metrics proved more adept at spotting emergent clusters, albeit at a higher computational cost. Lee and Park [14] extended these insights into the cryptocurrency domain, focusing on high-throughput environments where wallets can swiftly transfer funds across multiple addresses. Their real-time anomaly detection achieved consistently high recall rates for flagging malicious activities, yet the authors cautioned about critical scalability concerns in fast-growing blockchain ecosystems. Taken together, these studies suggest that while multi-metric, continuously updated analyses hold promise for detecting sophisticated laundering operations, they also introduce significant operational challenges in terms of computational overhead and data management—underscoring the need for well-optimized systems and robust institutional support to ensure timely, effective AML monitoring.

One of the main technical challenges in applying graph-based methods is the high computational cost of calculating centrality metrics, especially in real-time environments. Metrics such as betweenness and eigenvector centrality require building and analyzing all shortest paths, which becomes resource-intensive for large-scale graphs.

To optimize performance, it is possible to use approximate algorithms, stream-processing frameworks (e.g., Apache Flink, Apache Kafka), and graph engines that support parallel computation – such as cuGraph, GraphX, and other GPU-accelerated solutions.

In pilot implementations, simplified approaches are often used—for example, constructing subgraphs only around suspicious nodes or aggregating transactions by time, geography, or participant type. These strategies significantly reduce graph size without losing essential structural information.

Additionally, pre-filtering transactions before the graph construction phase is recommended. This reduces the data volume and improves system responsiveness.

In another critical contribution, Wang et al. [15] introduced a dynamic graph representation learning framework to capture the fluid, time-sensitive evolution of suspicious transactions in large-scale financial networks. Through extensive experiments on real-world banking data, they demonstrated that updating node embeddings to reflect newly formed or dissolved relationships substantially improved detection recall compared to static graph models. However, their results also revealed the considerable computational burden of recalculating graph embeddings, particularly for institutions processing millions of daily transactions. Wang et al. therefore argued that unless organizations invest in specialized data architectures and parallelized computing resources, the practical adoption of real-time dynamic graph analytics may remain limited.

Building on this notion of temporal complexity, Kim and Choi [16] focused on high-frequency trading (HFT) environments, analyzing minute-by-minute or even second-by-second changes in node centralities. Their findings indicated that abrupt surges in betweenness or degree centrality served as strong precursors to orchestrated market manipulation, including front-running and wash trading schemes. Critically, they underscored that while the real-time recalculation of these metrics significantly enhances early fraud detection, it introduces laten-

DOI: 10.37943/22ZSKI6025

cy and throughput challenges – demanding near-instantaneous processing of vast, continually updating trade logs.

A parallel research trajectory applies multi-layer Social Network Analysis (SNA) specifically to cross-border remittances. In Zhao and Liu's work [17], specialized centralities at various financial layers (such as peer-to-peer versus bank-to-bank) uncovered hidden corridors often exploited in large-scale laundering operations. The authors observed an appreciable jump in detection accuracy—especially for covert, smaller transactions – yet warned that integrating heterogeneous data sources from multiple layers could inflate system complexity and data-cleaning overhead. Similarly, Verma and Devi [18] combined embedded node features (like account type or historical risk rating) with SNA centralities to obtain highly granular alerts, thereby reducing false positives in complex, multi-layer settings. While their integrated approach demonstrated promising results in pilot studies, they also acknowledged the increased need for domain expertise to tune and interpret the expanded feature sets.

Despite these algorithmic and methodological gains, effective analytical tools must also be accessible to human investigators – a point emphasized by Moreno et al. [19] in their examination of visual analytics for large-scale AML detection. Through user-centered evaluations of interactive dashboards, they found that overlaying real-time SNA metrics, such as degree or betweenness, onto intuitive graphical interfaces accelerated the discovery of abnormal clusters by nearly 40%. However, the authors also noted that the learning curve for interpreting high-dimensional transaction networks could impede adoption, necessitating focused training programs. Finally, Jha et al. [20] proposed a "hybrid AI" paradigm that blends rule-based detection, neural network embeddings, and SNA attributes, achieving more proactive AML coverage across diverse financial products. While their approach notably improved the detection of emerging fraud typologies, they cautioned that fragmented institutional structures—or a lack of cohesive data-sharing protocols – could nullify these benefits. Taken together, these studies highlight both the immense promise of dynamic, multi-layer graph analytics for AML and the significant operational, technological, and institutional hurdles that practitioners must navigate to implement them on a scale.

Research Aim and Objectives

This study aims to develop and validate a graph-based methodology that leverages centrality metrics – particularly degree, betweenness, and closeness – to detect suspicious transactions within financial networks. By applying social network analysis techniques in a real-time data context, the research seeks to provide a cohesive set of tools for the early detection of money laundering schemes and related illicit activities. Accordingly, the primary goal is to model financial transactions as a directed graph, where nodes represent senders and beneficiaries and edges signify transaction flows. This representation may also integrate additional features, such as transaction volumes or timestamps, to enhance analytical precision.

In pursuit of this goal, the research entails the dynamic calculation of centrality metrics to identify nodes exhibiting atypical patterns that could indicate potential involvement in money laundering or fraud. It also involves the establishment of a real-time monitoring system capable of recalculating these measures as new transactions are recorded, thus allowing for rapid alerts when certain thresholds of suspicious activity are surpassed. A comparative analysis of degree, betweenness, and closeness centralities will help determine how each measure, individually and jointly, can illuminate behavioral anomalies in the network.

Validation of the proposed methodology is envisioned through a combination of case studies and simulation experiments that mirror the complexity of large-scale, real-world financial ecosystems. Performance evaluations will be carried out to assess metrics such as precision, recall, and F1-scores, alongside an examination of scalability for higher-volume networks. The

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

study will conclude with the formulation of practical guidelines for financial institutions seeking to adopt social network analysis in their anti-money laundering strategies, including recommendations for potential integration with machine learning approaches and consideration of additional metrics – such as eigenvector centrality – that may further optimize early-stage detection of illicit transactions.

The overall structure of the proposed approach, including the use of centrality metrics and their integration into transaction network analysis, is illustrated in the conceptual model below (see Figure 1).

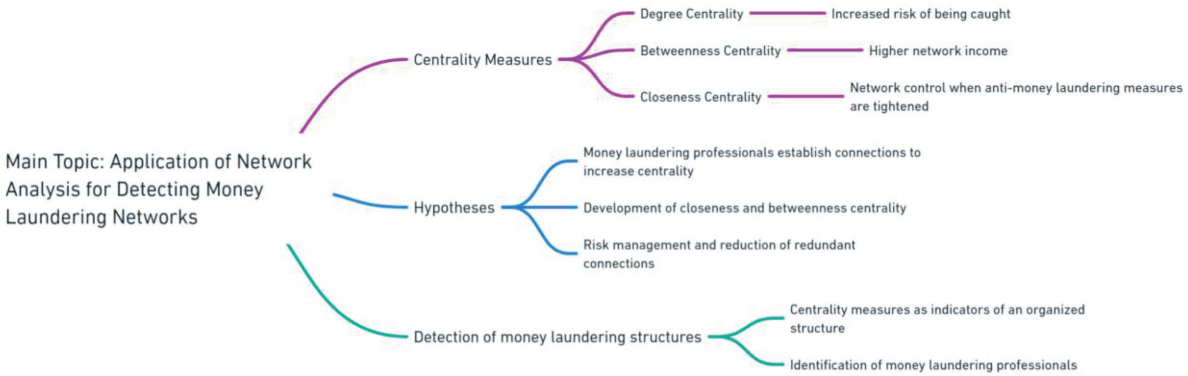

Figure 1. Conceptual Model of Research

Methods and Materials

Let us introduce the basic definitions necessary for applying centrality metrics for fraud detection purposes. In graph theory, a simple graph G (N, V) is a collection of two sets - the set of vertices of the graph N and the set of its edges E – unordered pairs of different elements of the set N. In this study, nodes are senders and beneficiaries, and edges are transactions that connect senders and beneficiaries and represent the transfer of money from the sender to the beneficiary. One of the most popular areas where graphs are used is Process Mining, an area that focuses on discovering, analyzing, and optimizing business processes based on event log data, and Social mining, an area that focuses on identifying social connections in social networks. However, in this study, the connections are not found in social networks, but in banking transactions.

The characteristic "centrality" allows us to determine the degree of importance of a graph node based on its location. Let's consider several ways to calculate it.

A. Degree centrality

Degree centrality measures how important a particular node is in terms of the number of connections it has to other nodes in the network, and for a weighted graph is calculated as follows:

$$C_d(i) = \sum_j^N w_{ij} \quad (1)$$

Where:
- i is the index of the vertex in question
- w_{ij} is the weight of the edge (i, j)
- N is the number of vertices in the graph.

B. Eigenvector centrality

The disadvantage of the previous measure is that it takes into account only the nearest neighbors of the node in question, while this measure takes into account the "influence" of the

DOI: 10.37943/22ZSKI6025

(central) nearest neighbors themselves. The principle of the measure can be described as follows: "if my friends are influential, then I will be more influential." The formula for calculating this measure is:

$$c_e(i) = \frac{1}{\lambda}\sum_j^N w_{ij} c_e(j) \quad (2)$$

Where:
- i is the index of the vertex in question
- w_{ij} is the weight of the edge (i, j)
- λ is some normalization coefficient

To calculate eigenvector centrality, we need to transform this formula by introducing the notations,

$$\vec{v} = (c_e(1), c_e(2), \dots, c_e(N))^T \quad (3)$$

$$W = w_{ij} \quad (4)$$

Where
- $\vec{v}$ is the vector consisting of the centrality values of each vertex
- W is the weight matrix of the graph in question

Using the entered notations, the original formula is transformed to $Wv = \lambda v$, and this is already a classic problem of finding the eigenvectors of a matrix; as a final answer, it is necessary to take the eigenvector corresponding to the maximum eigenvalue.

C. Closeness centrality

The previous measures are usually classified as structural, while the next two measures under consideration are usually classified as geometric, since they are based on the shortest paths in the graph. Closeness centrality $Cc(i)$ for the i-th vertex of the graph is calculated by the formula:

$$C_c(i) = \frac{1}{\frac{1}{N-1}\sum_{j=1, j\neq i}^N d_{ij}} \quad (5)$$

Where
- i and j are vertex indices of the graph under consideration
- d_{ij} is the shortest path from vertex i to vertex j, meaning the minimum number of edges one must traverse to get from i to j

This measure has a straightforward physical meaning: the smaller the distances from vertex i to the other vertices j in the graph (in the extreme case $d_{ij} - 1$, i.e., vertices i and j are connected by an edge), the smaller the denominator in the formula for $Cc(i)$ becomes, and thus the greater its centrality value. It is important to note that this measure only makes sense for connected graphs, as the presence of isolated vertices or entire components would make the shortest paths to these objects effectively infinite, with all the resulting consequences.

D. Betweenness centrality

This measure is very popular, and often in various literary sources on Network Science, when the term "centrality" is mentioned, it is precisely betweenness centrality that is meant. The formula for calculating the value of this measure for the i-th vertex of the graph will already look more complicated:

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

$$C_b(i) = \frac{2}{N(N-1)} \sum_{(j,k),j \neq k} \sigma_{jk}(i)\sigma_{jk} \quad (6)$$

Where
- σ_{jk} is the number of shortest paths from vertex j to vertex k
- $\sigma_{jk}(i)$ is the number of shortest paths from j to k that pass through vertex i
Summation in this formula is performed over all possible pairs of vertices (j,k)

Simply put, this measure shows how often the vertex i acts as a "transshipment point" when traveling from one vertex of the graph to any other. It is quite effective in identifying "bottlenecks" in a graph – vertices that are part of one or several edges connecting two clearly defined clusters.

After the most important vertices by centrality have been identified, additional analysis of these vertices is carried out in order to clarify their role in the resulting social graph.

For a deeper analysis of the node with the highest betweenness centrality in the graph, several directions can be considered:

Neighbor analysis: study the nodes that are directly connected by a vertex. This helps to understand which nodes interact with the node most frequently.

```
neighbors = list(G.neighbors(top_node))
# top_node - это узел с максимальным посредничеством
print("Соседи:", neighbors)
```

Subgraph: Create a subgraph that includes the maximum-betweenness vertex and its neighbors to better understand its role in the local structure.

```
subgraph_nodes = neighbors + [top_node]
subgraph = G.subgraph(subgraph_nodes)
```

Roles in different paths: how often a given node participates in the shortest paths between other nodes. To do this, we select random pairs of nodes and check how often a node participates in linking them.

```
all_paths = dict(nx.all_pairs_shortest_path(G))
top_node_in_paths = sum(top_node in paths for paths in all_paths.values())
print(f"Вершина {top_node} участвует в {top_node_in_paths} кратчайших путях")
```

Structural holes: A high-betweenness vertex often serves as a bridge between several otherwise unconnected parts of the graph.

Visualization: We visualize the subgraph by highlighting the vertex and its neighbors to better see its position in the network.

```
import matplotlib.pyplot as plt
pos = nx.spring_layout(subgraph)
#можно выбрать другой метод размещения
nx.draw(subgraph, pos, with_labels=True, node_color='lightblue',
edge_color='gray', node_size=500, font_size=10)
plt.show()
```

These methods make it possible to understand why exactly a given peak ended up in the top for mediation.

DOI: 10.37943/22ZSKI6025

Results and Discussion

Let's conduct an analysis on the application of the above methodology using masked transaction data from one bank using Python.

A. Loading and preparing data

We use the following libraries:

- *pandas (pd)*- for working with data in tables (Excel);
- *networkx (nx)*- for constructing graphs;
- *plotly.graph_objects (go)*- for creating interactive 3D graphics.

Loading data from an Excel file (*DATA.xlsx*), containing transactions between senders and receivers, into an object *DataFrame (df)* using pandas. The data structure is oriented to represent unique identifiers of the sender *Sender_BIN* and the recipient *Beneficiary_BIN*.

B. Construction of the graph

Using the NetworkX library, we create a directed graph based on sender-receiver pairs, where each transaction between two entities is a directed edge, which allows us to visualize and analyze the connections between entities as nodes and edges. The graph uses columns *Sender_BIN* and *Beneficiary_BIN*, indicating senders and recipients in transactions.

C. Visualization of the graph

The graph visualization is implemented using the Plotly library. The layout of the nodes is configured, the distance between them and their position are controlled for visual display. For this, *spring_layout*, where parameters *k*, *iterations* and *scale* regulate the visual presentation of the graph. The resulting transaction graph is shown in Figure 2.

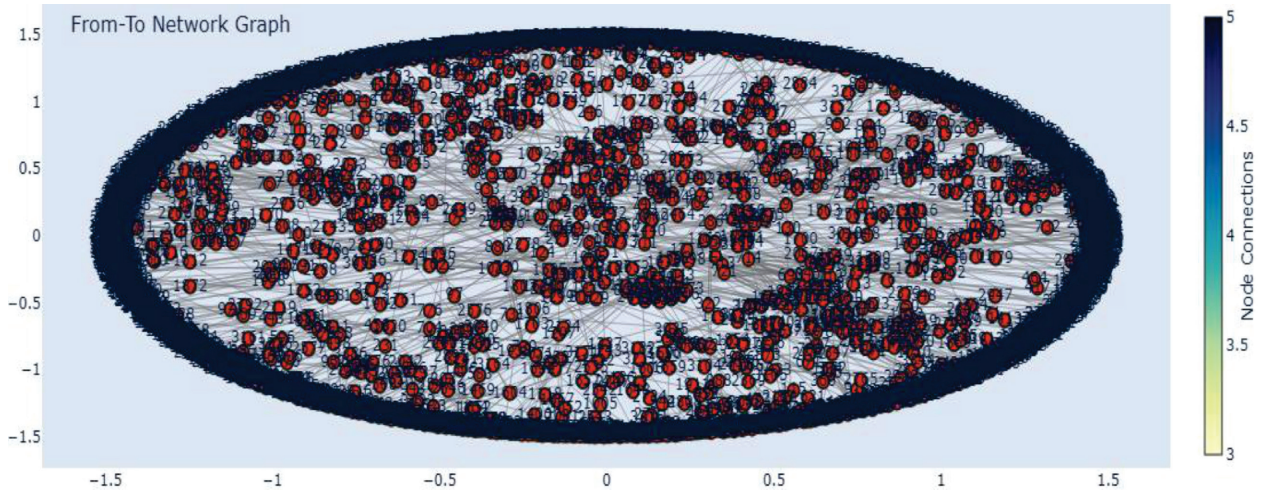

Figure 2. From-To Network Graph

In addition to the standard 2D layout, the same transaction network is also rendered using an interactive 3D visualization, which provides enhanced clarity for observing structural clusters and node positioning (see Figure 3).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

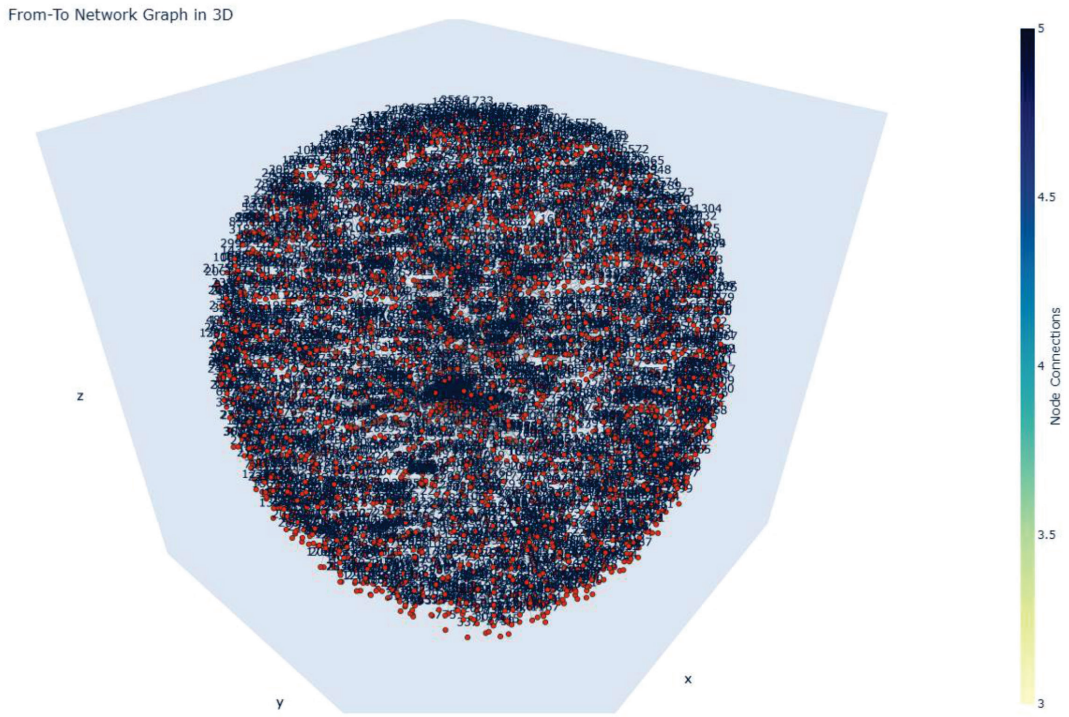

Figure 3. From-To Network Graph in 3D Visualization

D. Calculating centralities

To analyze the importance of nodes in a network, centrality measures are used:

- Degree centrality shows the number of direct connections of a node;
- Closeness centrality reflects how close a node is to all others;
- Eigenvector centrality indicates nodes with high importance through their connection with other important nodes;
- Betweenness centrality identifies nodes through which many shortest paths pass.

For example, snippet sorts nodes by their betweenness centrality values and outputs the top 10 nodes with the highest influence in the graph. Betweenness centrality values *betweenness centrality* indicate how important a node is for transmitting information by measuring its participation in the shortest paths between other nodes. The higher the value, the more significant the node is in the network structure.

Table 1. Top 10 most influential nodes based on betweenness centrality values

Node number	**Betweennes centrality value**
Node 729	0.00024152492420198153
Node 1500	0.00018672574645268257
Node 1873	0.0001770763911582743
Node 870	0.00017205292972288062
Node 829	0.00016549113941275785
Node 694	0.00014992953826333907
Node 1274	0.00013251781270987348
Node 9	0.00013110257393336028
Node 97	0.00012659954146263642
Node 3164	0.0001223967111566275

DOI: 10.37943/22ZSKI6025

Node 729 has the highest value (0.0002415249242019815 3), making it the most significant "middleman" or "center" in the graph. Other nodes, such as 1500 and 1873, also have relatively high betweenness centrality values, although lower than node 729. Computed nodes can play an important role in combating fraud, as they are critical points through which the greatest number of shortest paths link different parts of the graph pass.

E. Selecting key nodes

We determine the top 10 nodes for each centrality metric to identify the most significant network nodes in terms of their influence and role in the transaction structure. We calculate key centrality metrics for graph nodes and select the top 10 nodes with the highest values in four categories: by degree, by proximity, by eigenvector, and by intermediation.

```
Топ-10 узлов по мерам центральности:
Центральность Degree: [729, 701, 714, 866, 9, 1615, 286, 755, 1402, 724]
По степени: узлы, которые имеют наибольшее количество прямых связей (ребер)
Данные узлы наиболее активно связаны с другими, то есть имеют максимальное число
«соседей», и могут быть важными «центрами» в графе.

Центральность Closeness: [829, 870, 729, 830, 871, 820, 750, 773, 749, 866]
По близости: узлы с наибольшей близостью находятся ближе всего ко всем другим узлам в
графе. Они имеют низкие суммарные расстояния до всех остальных узлов.
Данные узлы могут быстро передавать информацию по графу или влиять на другие узлы через
минимальное количество шагов.

Центральность Eigenvector: [729, 819, 714, 1402, 820, 713, 866, 712, 700, 809]
По собственному вектору: оценивает узлы по их важности, учитывая также значимость
узлов, к которым они подключены.
Данные узлы не только хорошо связаны, но и подключены к другим значимым узлам. Они
могут влиять на важные узлы, а не только на тех, кто имеет много связей.

Центральность Betweenness: [729, 1500, 1873, 870, 829, 694, 1274, 9, 97, 3164]
по посредничеству: показывает через какие узлы проходят наибольшее количество
кратчайших путей. Узлы с высокой межузловой центральностью контролируют потоки
информации в графе.
Данные узлы выполняют роль «мостов» и могут быть критически важными для передачи
информации или для объединения различных частей графа.
```

Node 729 stands out as important across all centrality measures, indicating its key role in the graph. Other nodes such as 870, 866, 829 also appear in several metrics, confirming their importance and versatility in the graph.

F. Visualization of subgraphs

Function implemented *visualize_subgraph* to create subgraphs around the central node, visualizing direct and indirect connections. In the subgraph, nodes are highlighted in different colors (central - red, direct connections - blue, indirect - green), which allows you to observe the structure of the local network and evaluate the influence of key nodes on neighboring ones. The result is shown below.

G. Finding common connections

The final stage of analysis looks for common direct and indirect connections between top nodes, which can be useful for identifying clusters and interconnected groups in the network.

Analyzing common links between top nodes helps identify clusters and patterns of interactions. Nodes with many common links (both direct and indirect) are more tightly integrated into the network and may indicate highly interconnected groups of nodes that are important for the transfer of information or resources.

For top nodes, direct and indirect link analysis is performed to identify which entities are directly or indirectly connected to key nodes. Function *analyze_connections* returns a list of direct and indirect links for each node.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

Table 2. Analysis of direct and indirect connections for top 10 nodes

Node number	Direct connections	Indirect connections
Node 729	15	22
Node 1500	8	4
Node 1873	5	6
Node 870	3	11
Node 829	4	23
Node 694	8	11
Node 1274	7	7
Node 9	16	5
Node 97	4	7
Node 3164	2	2

The subgraphs of the most influential nodes by betweenness centrality are displayed in Figures 4 to 13.

Nodes with a large number of direct and indirect links (Fig.4), (Fig.11) are the central elements of the network. They can quickly reach other nodes and probably play a significant role in the transmission of information.

Nodes with a small number of direct but significant number of indirect links (Fig.7), (Fig.8) indicate nodes that, although they do not have many direct contacts, can influence a significant number of nodes through intermediate links.

Nodes with a small number of both direct and indirect links (Fig.13) are generally considered less important for disseminating information or resources throughout the graph.

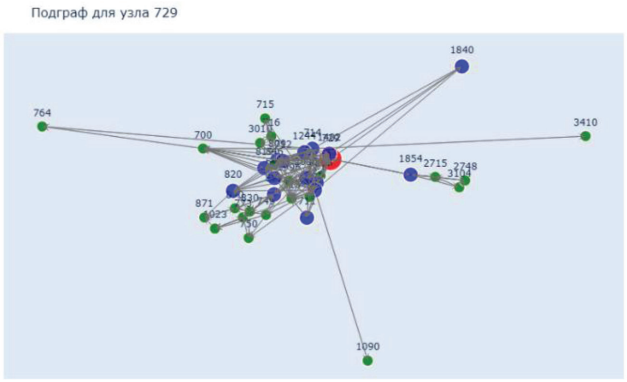

Figure 4. Subgraph for node 729

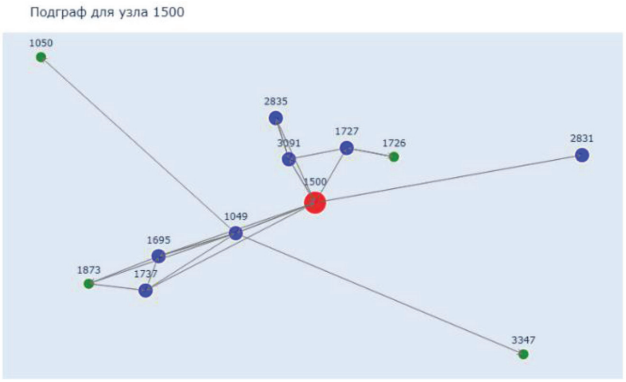

Figure 5. Subgraph for node 1500

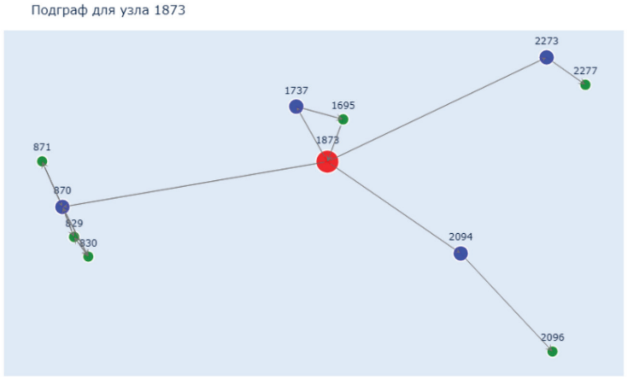

Figure 6. Subgraph for node 1873

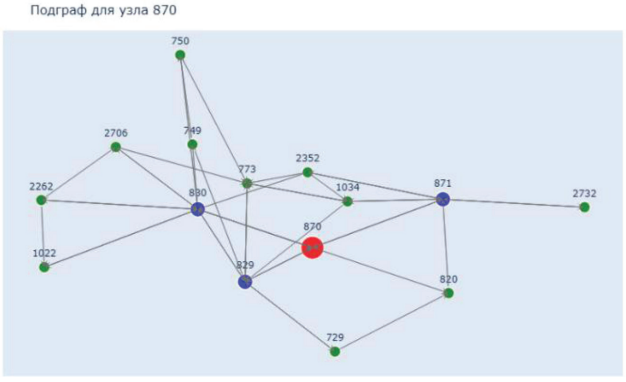

Figure 7. Subgraph for node 870

DOI: 10.37943/22ZSKI6025

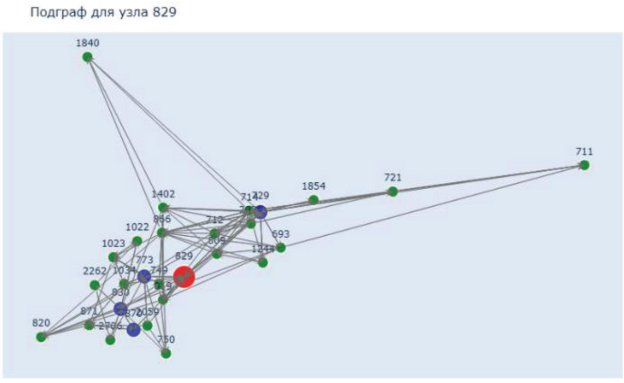

Figure 8. Subgraph for node 829

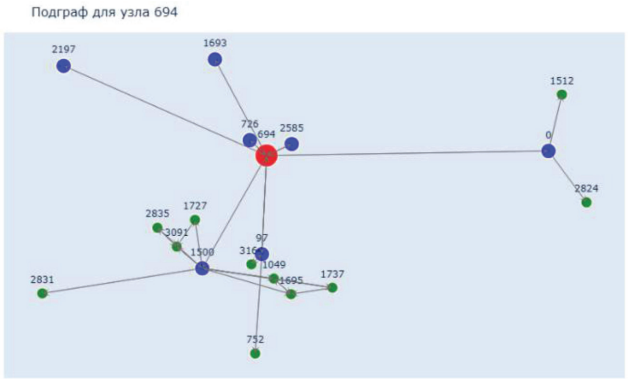

Figure 9. Subgraph for node 694

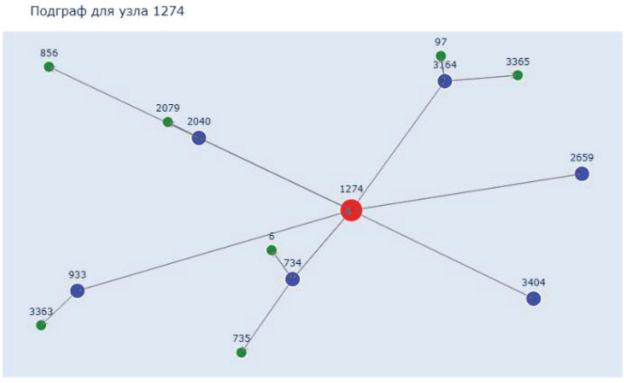

Figure 10. Subgraph for node 1274

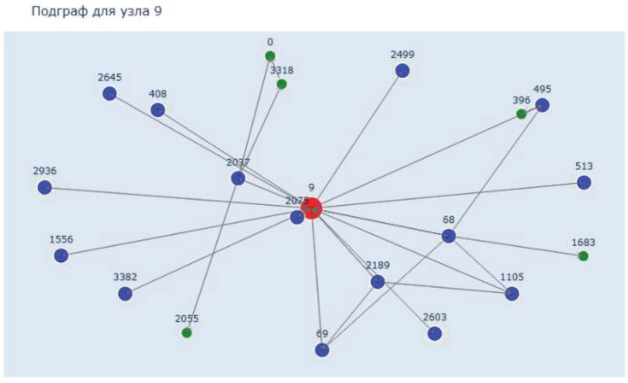

Figure 11. Subgraph for node 9

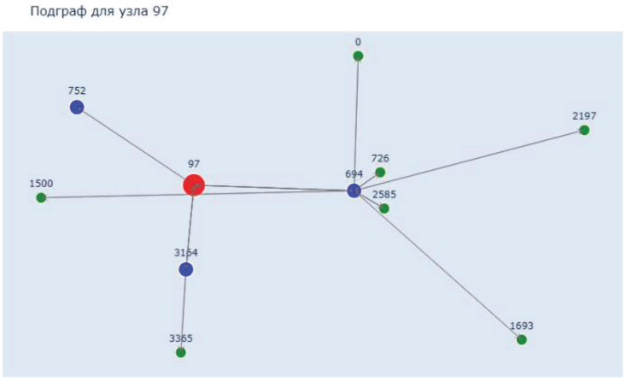

Figure 12. Subgraph for node 97

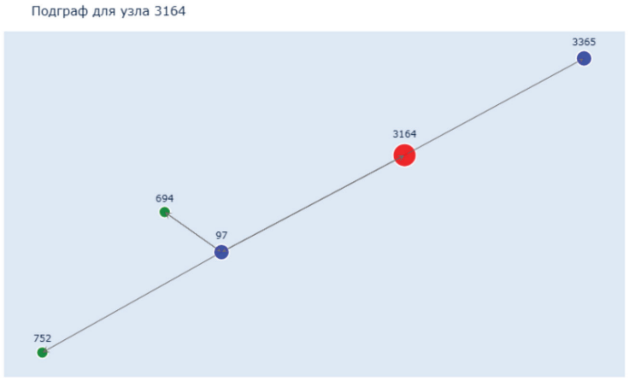

Figure 13. Subgraph for node 3164

The identified nodes represent companies that acted as intermediaries in financial transactions. Structural analysis determined that these companies actively interact with many different counterparties, which indicates its possible involvement in the withdrawal of money or participation in shadow schemes. The reason for its dubiousness appeared due to the uncharacteristic activity of the identified company – a high degree of connectivity with nodes.

Thus, the frequency of using this company as a transshipment can be determined by analyzing its centrality in the graph and identifying suspicious patterns of behavior.

Network analysis based on calculated node centralities revealed key points through which the largest flows of information and resources pass. Identification of nodes with high degree and internodal centrality allowed us to identify central elements of the structure responsible for the distribution of transaction flows. Nodes with high eigenvector centrality demonstrat-

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

ed importance for the entire network, as they exert influence through their connections with other significant nodes.

The approach of using centrality metrics to analyze interaction networks has significant potential in practice. In the financial sector, it can be used to monitor transactions and identify anomalous patterns associated with money laundering or fraud. It can also be useful for optimizing logistics and communications in complex networks, such as transportation systems or corporate supply chains, as it identifies key nodes on which to focus monitoring and management.

The conducted analysis showed the effectiveness of this approach in identifying structural features and important elements of the network. Future research can focus on developing algorithms for automatic detection of suspicious nodes in real time and on integrating the proposed approach with other data analysis methods.

Limitations and Potential Sources of Bias

Despite the positive results, the proposed approach has a number of limitations that must be considered when interpreting the findings.

First, the graph structure and centrality calculations depend on the completeness and quality of the input data. If certain transactions are unregistered or intentionally concealed—for example, through intermediaries or shell companies—key nodes may be excluded from the analysis, leading to distorted centrality values. This issue is particularly relevant in low-transparency cross-border operations. It should also be noted that significant effort is required to normalize and clean the data before applying graph analysis.

Second, the centrality metrics used—such as betweenness and eigenvector centrality—assume that all connections are equally significant. In practice, differences in transaction volume, currency, or participant type (e.g., individuals versus legal entities) can significantly affect a node's actual importance. Incorporating edge weights and attribute-based analysis could improve precision but would increase model complexity and raise data requirements.

A third limitation arises from the static nature of the network snapshot. Suspicious activity may unfold dynamically over time, and without a temporal component, short-term anomalies or episodic schemes may remain undetected.

In addition, high centrality does not necessarily indicate illicit activity. Structurally central positions may also be occupied by legitimate entities such as payment providers, brokers, or financial institutions with high transaction volumes. For this reason, graph-based analysis should be complemented by contextual information such as Know Your Customer (KYC) data, geographic distribution of transactions, client profiles, and other relevant sources.

A promising direction for future research is the development of adaptive models that account for temporal changes in the network, the use of weighted centrality metrics, and the integration of graph analysis with machine learning techniques.

Ethical Considerations and Practical Risks

Graph-based anti-money laundering (AML) systems provide powerful tools for detecting suspicious financial activity, but their application raises significant ethical, legal, and operational concerns that must be addressed to ensure responsible deployment. While centrality metrics offer interpretable insights into network structure, they may incorrectly flag legitimate entities simply due to their position within a transactional graph—particularly in highly connected or intermediary roles. These false positives can lead to reputational harm, unwarranted investigations, service disruptions, and potential legal liability.

Beyond the risk of incorrect classification, there are profound concerns related to data privacy, regulatory compliance, and algorithmic fairness. The analysis of financial networks typically

DOI: 10.37943/22ZSKI6025

involves sensitive client information, including personal identifiers, transaction histories, and behavioral profiles. Mishandling such data—whether through leakage, unauthorized access, or lack of safeguards—can violate privacy regulations and erode public trust. Notably, the General Data Protection Regulation (GDPR, EU 2016/679) and the Sixth Anti-Money Laundering Directive (6AMLD, EU 2018/1673) impose strict requirements for data minimization, lawful processing, and auditability in financial institutions. Similarly, the Financial Action Task Force (FATF) promotes a risk-based approach to ensure proportionality and protection of individual rights.

To mitigate these risks, institutions must implement robust technical and organizational controls. Key measures include:

- Anonymization or pseudonymization of sensitive data prior to analysis;
- End-to-end encryption and role-based access control throughout data pipelines;
- Audit logging of system decisions and investigator actions to ensure accountability;
- Bias monitoring and fairness assessments to prevent disparate treatment of clients based on geography, demographics, or behavioral clustering;
- Regular model validation and explainability testing, especially in high-impact use cases.

It is also critical to integrate human oversight into the decision loop. Suspicious entities identified by centrality metrics—such as those falling in the top 5% for betweenness or eigenvector scores—should undergo contextual review before any enforcement action is taken. Investigators should evaluate transaction volume trends, Know Your Customer (KYC) data, peer group behavior, and jurisdictional risk factors. For instance, an entity flagged for high network centrality may be deprioritized if its counterparties include reputable financial institutions with historically low risk, or escalated if it interacts with known high-risk entities or exhibits a sudden increase in transaction velocity.

To facilitate such decisions, the system architecture should support explainable interfaces, including graph visualizations, annotated paths, and rationales for alerts. For example: "*Node 729 was flagged for appearing in 83% of shortest paths between two high-risk transaction clusters over a 48-hour period, with a 60% increase in transaction volume relative to baseline*." This transparency allows for informed decision-making and reduces the likelihood of unwarranted escalation.

Ultimately, the ethical deployment of graph-based AML technologies requires a balance between analytical power and institutional responsibility. Ensuring data protection, procedural fairness, and explainability is not only a matter of regulatory compliance - it is essential for sustaining the legitimacy and effectiveness of financial surveillance in modern institutions.

Technological Constraints and Real-Time Applicability

One of the key challenges in applying graph-based approaches to AML tasks is the high computational load. Calculating centrality metrics—particularly betweenness and eigenvector centrality—requires substantial resources, which complicates the analysis of large-scale data in real-time environments.

This limitation reduces the scalability of such solutions and calls for the implementation of more efficient technical strategies. Possible approaches include approximate algorithms, data aggregation prior to analysis, parallel processing, stream-based graph computation, and the use of graphics processing units (GPUs) to accelerate calculations. Modern graph analytics libraries, such as cuGraph, can significantly reduce execution time by leveraging hardware-level parallelism.

Recent advancements in real-time anomaly detection have demonstrated that integrating stream-based machine learning models with dynamic network features can significantly en-

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

hance the responsiveness and adaptability of AML systems. For instance, the Graph Feature Preprocessor enables real-time subgraph-based feature extraction for financial crime detection, improving model accuracy and throughput. Similarly, SLADE employs self-supervised learning to detect dynamic anomalies in edge streams without labeled data, facilitating rapid adaptation to evolving patterns in financial transactions [22], [23].

Comparison with Alternative Detection Methods

Despite their high interpretability and structural informativeness, graph-based centrality metrics should be compared with other approaches such as supervised learning (e.g., Random Forest, SVM) and clustering algorithms (e.g., DBSCAN, k-means).

Supervised learning methods demonstrate high accuracy when labeled data is available, but they tend to lose effectiveness when confronted with new or previously unseen fraud schemes. Clustering algorithms can identify hidden groups, but they require precise parameter tuning and are often sensitive to noise, which reduces their stability. In contrast, centrality metrics do not rely on historical labels, making them especially useful in scenarios with limited training data or when detecting emerging anomalies. They also provide insight into the structure of interactions within the network—something that is difficult to capture using tabular features alone. However, in highly dense or noisy graphs, centrality measures may lose precision. The best results are often achieved by combining them with other techniques - for example, in ensemble models that integrate both structural and statistical approaches.

Practical Implementation in AML Systems

The integration of graph analytics into existing AML infrastructures can be approached in several stages. Initially, centrality metrics may be used as a secondary filter after rule-based triggers, flagging only those transactions that appear in high-centrality subgraphs. These results can then be passed into case management tools (e.g., SAS AML, NICE Actimize), with interactive dashboards for investigators. GPU-accelerated graph engines like cuGraph and streaming pipelines using Apache Flink or Kafka enable real-time screening. For institutional adoption, standard operating procedures (SOPs), staff training, and model governance frameworks must be implemented in parallel to technical deployment.

Conclusion

The present study demonstrates the significant analytical power of network-centric approaches for detecting potential money-laundering activities in complex financial systems. By analyzing direct and indirect connections of top-ranked nodes and examining multiple centrality metrics—particularly degree, betweenness, and eigenvector—it becomes evident that certain entities act as pivotal intermediaries in transactional flows. These nodes occupy critical structural positions, evidenced by both extensive direct ties and numerous indirect linkages, thereby suggesting that they facilitate an unusually large volume of monetary transfers or resource exchanges.

Such a finding proves highly relevant for practical applications in finance. The identification of these intermediary nodes, which may not always appear overly active on the surface yet maintain considerable influence through second- and third-tier connections, illuminates the covert architecture of illicit financial operations. In particular, observing uncharacteristic or anomalous degrees of connectivity serves as an early warning signal that the entity in question could be enabling or concealing unauthorized fund transfers.

Moreover, the results confirm that the integration of centrality-based methods can be effectively employed beyond mere detection, for instance in optimizing communication paths within corporate supply chains or managing routing flows in transportation networks. By pin-

pointing nodes that shape network structure and control resource movement, stakeholders can allocate monitoring resources and risk assessments more judiciously.

Nevertheless, opportunities remain for further refinement. Future research may emphasize the development of automated real-time modules that continuously update centrality measures, enabling more immediate detection and intervention. An additional line of inquiry involves the extension of these techniques to encompass emerging digital finance platforms, such as cryptocurrencies and decentralized financial systems, where transaction anonymity and velocity present unique detection challenges. Overall, these findings underscore the potential of graph-based analytics as a critical tool for financial intelligence, compliance, and broader network analysis, reinforcing their value in both academic inquiry and real-world implementation.

Graph-based methods can be integrated into existing financial monitoring infrastructures at multiple levels. During the initial screening stage, they can serve as a "second filter" following rule-based systems - for example, when a predefined rule flags a group of transactions, a localized graph can be constructed, and centrality metrics can be calculated. In the investigation phase, the results of graph analysis (such as central nodes, their connections, and subgraphs) can be automatically transferred to case management systems (e.g., NICE Actimize, SAS Anti-Money Laundering) and visualized through interactive dashboards. Formalized decision protocols can also be introduced—for instance, if a client ranks in the top five by betweenness centrality, a mandatory document review is triggered. To scale such an approach, standardized procedures for graph interpretation and dedicated training for compliance teams are essential.

This is consistent with the objectives set out in the European Union's Sixth Anti-Money Laundering Directive (6AMLD), which highlights the need for enhanced monitoring capabilities and technological innovation in detecting financial crime [21].

References

[1] Lawlor-Forsyth, E., & Gallant, M. M. (2018). Financial institutions and money laundering: A threatening relationship. Journal of Banking Regulation, 19(2), 131-148.

[2] Financial Action Task Force. (2023). *Digital transformation of AML/CFT for operational agencies*. Retrieved May 30, 2025, from https://www.fatf-gafi.org/en/publications/Digitaltransformation/Digital-transformation-aml-cft.html

[3] Lewis, J.B. (2022). Money finds a way: increasing AML regulation garners diminishing returns and increases demand for dark financing. Vand. J. Transnat'l L., 55, 529.

[4] Ahmed, M., Mahmood, A.N., & Islam, M.R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems, 55*, 278–288. https://doi.org/10.1016/j.future.2015.09.018

[5] Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of machine learning-based k-means clustering for financial fraud detection. *Academic Journal of Science and Technology, 10*(1), 33–39.

[6] Savage, D., Wang, Q., Chou, P., Zhang, X., & Yu, X. (2016). Detection of money laundering groups using supervised learning in networks. *arXiv Preprint*, arXiv:1608.00708. https://arxiv.org/abs/1608.00708

[7] Sun, X., Feng, W., Liu, S., Xie, Y., Bhatia, S., Hooi, B., ... & Cheng, X. (2022, February). MonLAD: Money laundering agents detection in transaction streams. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (pp. 976-986).

[8] Pambudi, B. N., Fauziati, S., & Hidayah, I. (2022). A minimum error-based PCA for improving classifier performance in detecting financial fraud. *Jurnal Teknik Elektro, 14*(1), 1–9.

[9] Zhu, M., Zhang, Y., Gong, Y., Xu, C., & Xiang, Y. (2024). Enhancing credit card fraud detection a neural network and smote integrated approach. arXiv preprint arXiv:2405.00026.

[10] Sousa Lima, R. (2022). Social network analysis applied to money laundering detection in Brazil. *Journal of Financial Crime, 29*(4), 1152–1168. https://doi.org/10.1108/JFC-09-2021-0210

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 22, JUNE 2025

[11] Deprez, B., Vanderschueren, T., Verbeke, W., Baesens, B., & Verdonck, T. (2024). Network analytics for anti-money laundering: A systematic literature review and experimental evaluation. *arXiv Preprint*, arXiv:2405.19383. https://arxiv.org/abs/2405.19383
[12] Gerbrands, P., Stancu, A. M., & Nozina, M. (2021). Centrality measures in criminal network analysis for money laundering. *Journal of Illicit Economies and Development, 3*(2), 136–149. https://doi.org/10.31389/jied.94
[13] Smith, K., & Allen, T. (2021). Graph-based detection of anomalous financial transactions using social network analysis. *IEEE Access, 9*, 145932–145945. https://doi.org/10.1109/ACCESS.2021.3123483
[14] Lee, J. H., & Park, S. (2022). Blockchain network analytics for real-time anomaly detection in cryptocurrency flows. *Information Sciences, 590*, 305–315. https://doi.org/10.1016/j.ins.2021.11.033
[15] Wang, C., Zhao, H., Li, R., & Yang, J. (2021). Real-time money laundering detection with dynamic graph representation learning. *Expert Systems with Applications, 185*, 115583. https://doi.org/10.1016/j.eswa.2021.115583
[16] Kim, J., & Choi, J. Y. (2023). High-frequency trading fraud detection through dynamic network centralities. *ACM Transactions on Management Information Systems, 14*(2), 1–19. https://doi.org/10.1145/3594296
[17] Zhao, P., & Liu, Q. (2021). Cross-border remittance analysis using multi-layer social network metrics. *IEEE Transactions on Computational Social Systems, 8*(4), 757–767. https://doi.org/10.1109/TCSS.2021.3067635
[18] Verma, R., & Devi, V. (2022). Enhanced SNA-based method for illicit financial flow detection. *IEEE Transactions on Engineering Management, 69*(4), 1191–1202. https://doi.org/10.1109/TEM.2021.3082284
[19] Moreno, P., Ortega, D., & Sánchez, G. (2023). Visual analytics approach to exploring large-scale financial transaction networks. *Information Visualization, 22*(2), 124–137. https://doi.org/10.1177/14738716221138902
[20] Jha, S., Alsubaie, N., Alarifi, A., & Mehedi, H.M. (2023). Towards hybrid AI strategies for proactive anti-money laundering solutions. *Journal of Financial Innovation and Technology, 2*(1), 37–52.
[21] European Commission. (2021). *Directive (EU) 2018/1673 on combating money laundering by criminal law (6AMLD)*. Retrieved May 30, 2025, from https://eur-lex.europa.eu/eli/dir/2018/1673/oj
[22] Blanuša, J., Cravero Baraja, M., Anghel, A., von Niederhäusern, L., Altman, E., Pozidis, H., & Atasu, K. (2024). Graph Feature Preprocessor: Real-time subgraph-based feature extraction for financial crime detection. *arXiv Preprint*, arXiv:2402.08593. https://arxiv.org/abs/2402.08593
[23] Lee, J., Kim, S., & Shin, K. (2024). SLADE: Detecting dynamic anomalies in edge streams without labels via self-supervised learning. *arXiv Preprint*, arXiv:2402.11933. https://arxiv.org/abs/2402.11933