International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MACHINE LEARNING-BASED UNSUPERVISED ENSEMBLE APPROACH FOR DETECTING NEW MONEY LAUNDERING TYPOLOGIES IN TRANSACTION GRAPHS

Mir Mohtasam Hossain Sizan[1], Anchala Chouksey[2], Atika Dola[3], Sakera Begum[4], Umama Khanom Antara[5], Farhan Sazid[6] and Ramisa Anjum Oishi[7]

[1]Master's in Business Analytics, University: University of North Texas

[2]Master's in Financial Mathematics, University of North Texas

[3]Bachelor's in Business Administration – Finance, Idaho State University

[4]Master of Science in Information Technology, Washington University of Science and Technology.

[5]Master's in Business Analytics, University of North Texas

[6]Master's in information systems and technologies, University of North Texas

[7]Master of Science in Information Technology, California Lutheran University.

Corresponding author: Mir Mohtasam Hossain Sizan, Email: mhsizan855@gmail.com

Abstract

This study examines how new money laundering patterns can be spotted in financial transaction networks without depending on past labels, using a three-stage, temporally safe machine learning setup in the USA. We started by building time-resolved transaction graphs from the Elliptic dataset, making sure that no model ever saw information from the future during training or testing. From these graphs, we created node embeddings and ran several unsupervised anomaly detection methods, each offering its perspective on what might look suspicious. This first step gave us a baseline for how each method performed over time. The next step was to bring their outputs together in a rank-based ensemble. The goal was to even out the biases and blind spots of individual detectors and make the identification of high-risk transactions more consistent. In the final stage, we clustered the ensemble's top-ranked results, grouping related nodes into candidate typologies that reflected both structural and behavioral patterns in the network. What came out of this was a detection pipeline that outperformed any single method and did a better job of surfacing coherent, interpretable clusters of suspicious activity. One of the main takeaways is that emerging laundering methods rarely hinge on a single, obvious signal. They are built from a mix of interactions, small shifts in structure, timing, and

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

connectivity, that only stand out when viewed from multiple angles over time. This shifts AML from relying on static detection lists toward a system that evolves with the network, giving analysts a practical way to uncover threats they have never encountered before while staying true to real investigative timelines.

Keywords: Money Laundering Detection, Transaction Graphs, Temporal Evaluation, Node Embeddings, Unsupervised Learning, Ensemble Methods, Anomaly Detection.

1. Introduction

1.1 Background and Motivation

The fight against money laundering has long been characterized by a cat-and-mouse dynamic in which detection systems lag behind adaptive adversaries who continually change how they move value through payment systems and blockchains. Traditional rule-based systems capture known patterns but fail when criminals adopt slightly altered routing, timing, or splitting strategies that evade fixed heuristics. Supervised machine learning approaches attempt to overcome rigidity by learning discriminative models from labeled historical examples, yet they inherit a critical limitation: labels reflect past behavior, not future innovation, and labeled illicit events are scarce, costly to obtain, and often noisy. This dependency on labeled history creates a blind spot for emergent typologies that differ structurally from previously observed cases, leaving institutions exposed to new laundering schemes that do not match training examples.

Recent research using public cryptocurrency corpora has highlighted both the opportunity and the limitation of supervised graph-based methods: research teams released and used the Elliptic dataset to demonstrate supervised graph convolutional models and classical classifiers for illicit transaction classification, but those experiments emphasize classification accuracy on known labels rather than discovery of novel shapes of laundering behavior, and they typically evaluate in static or leakage-prone ways that overestimate real-world readiness (Weber et, al. 2019) [24]. The Elliptic data release itself has been a catalyst for AML research, but the dataset's labeled portion remains an imperfect proxy for the full range of illicit tactics and therefore cannot eliminate the need for unsupervised discovery workflows (Elliptic 2019) [8]. Work that explores structural embedding approaches on transaction graphs has shown that learned node representations like random-walk based embeddings can reveal latent connectivity patterns useful for downstream tasks, yet these representations can shift dramatically across time slices, complicating longitudinal analysis and typology tracking if alignment is not explicitly addressed (Grover and Leskovec 2016; Gürsoy et, al. 2021) [11][12].

At the same time, unsupervised anomaly detectors such as Local Outlier Factor and Isolation Forest offer model families that do not require labels and can surface anomalous nodes, but each detector expresses a different bias: density-based detectors find local density deviations, isolation-based methods seek points easy to separate, and distance-based techniques capture different geometric notions of outlyingness. Classic detectors, therefore, detect different slices of anomalous behavior, making it unlikely that any single unsupervised method will robustly capture the varied structural signatures of diverse laundering typologies (Breunig et al., 2000;

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Liu et al., 2008) [4][20]. A complementary line of AML work has demonstrated that node embedding plus classical classifiers can be effective for known illicit classes but is limited in discovery settings; Hu et, al. 2019 showed that node2vec-based classifiers can perform well for classification, yet that supervised success does not automatically translate to discovery of previously unseen typologies because classifier decision boundaries are shaped by historical labels and may ignore structural signals outside those boundaries (Hu et, al. 2019) [13]. Together, these observations motivate a temporally aware, unsupervised approach that fuses multiple detectors and accounts for embedding instability so that analysts can be presented with ranked, interpretable candidate typologies arising from real-time transaction graphs rather than static, retrospective prediction scores.

1.2 Importance Of This Research

Spotting new money laundering patterns in time isn't just an academic exercise. It has real economic, regulatory, and operational stakes. Banks and regulators work under strict anti-money laundering rules, where missing illicit flows can lead to major fines, lasting reputational harm, and wider systemic risks. On the other hand, flooding investigators with false positives wastes time, drains resources, and makes every new alert feel less meaningful. Raising the signal-to-noise ratio means less wasted effort, lower costs, and more attention on the kinds of cases that are both unusual and actionable. From a compliance standpoint, global standards call for risk-based methods and the ability to adapt quickly when new typologies appear. A detection process that works in realistic, online-like conditions, avoids data leakage, and proves stable over time builds the case for safe deployment. It also shows regulators that the system can meet shifting compliance expectations. In practice, an unsupervised ensemble that blends outputs from different anomaly detectors can cut down on blind spots tied to any single model and produce a ranked list of high-priority transactions for analysts to review. This matters because the number of alerts needs to match the pace and capacity of investigative teams.

Ensembles have been used in anomaly detection for years, but newer techniques refine how they work. Methods like item-response-theory-based fusion and streaming ensembles offer structured ways to combine diverse unsupervised models and adjust their influence on the fly, without relying on labeled data (Kandanaarachchi 2021) [15]. Embedding alignment and stability over time are equally important. If embeddings shift unpredictably between retraining cycles, then tracking and clustering suspicious subgraphs becomes unreliable. Using orthogonal transformations to align embeddings keeps their structure consistent, which makes it easier to build and maintain coherent typologies for human review (Gürsoy et al. 2021) [12]. All in all, ensembles mitigate systemic risks in anti-money laundering compliance (Khan et al., 2025) [17]

The aim isn't only to improve raw detection scores. It's to produce human-readable candidate typologies, small subgraphs, feature summaries, and patterns that an investigator can grasp at a glance. That makes the system a partner to human judgment, not an impenetrable black box. Establishing a reproducible temporal evaluation protocol using public datasets and synthetic

Received: July 24, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

injections also gives the field a shared benchmark. This lets researchers and practitioners test and compare typology discovery methods under realistic, leakage-free conditions (Weber et al. 2019; Elliptic) [24][8]. Taken together, the work addresses a gap between retrospective, supervised AML models and the forward-looking need to uncover emerging typologies, creating a more adaptable, transparent, and defensible approach to investigation.

1.3 Research Objectives and Contributions

This work pursues two primary objectives. First, it aims to develop and validate a time-safe unsupervised ensemble pipeline that can score and prioritize newly appearing transactions in an evolving transaction graph using only information available up to the current time step. Second, it seeks to group the highest-scoring transactions into coherent candidate typologies via clustering and subgraph extraction, producing concise, human-interpretable artifacts that investigators can inspect to confirm or refute suspected laundering patterns. The methodological contributions supporting these objectives include a strict time-sliced graph construction and evaluation protocol that prevents future leakage, a practical embedding and alignment strategy to maintain comparability of learned node representations across retraining windows, and a heterogeneous detector ensemble that fuses complementary unsupervised signals through rank-based aggregation. In addition, the study provides a synthetic typology injection framework that allows controlled benchmarking of discovery recall for canonical laundering strategies such as smurfing, layering, and hub-and-spoke patterns, enabling robust ablation studies and hyperparameter tuning without relying solely on partial ground truth labels. Finally, the entire pipeline is implemented in a reproducible form on the publicly available Elliptic dataset so that results can be validated by others and extended in follow-up work. These interlocking elements are designed to demonstrate that unsupervised ensembles, when combined with temporal discipline and embedding stability techniques, can surface meaningful, previously unseen patterns of suspicious activity while offering analysts a practical, ordered workload for human-in-the-loop validation and escalation.

2. Literature Review

2.1Traditional AML Approaches

For years, anti-money laundering (AML) work has leaned heavily on rules, thresholds, and carefully crafted workflows. These systems flag accounts or transactions when they hit certain benchmarks: unusually large cash deposits, rapid “round-tripping” transfers, or links to sanctioned entities. They’ve been the backbone of transaction monitoring because they’re easy to audit and adjust. The problem is, they don’t age well in the face of clever adversaries. Criminal networks can tweak timing, amounts, or routing just enough to slip under those hard-coded limits. Supervised machine learning models came in as a way to broaden coverage. Trained on labeled alerts and confirmed cases, they can spot more complex patterns and outperform rules on historical data. But when the aim is to discover new typologies, this approach runs into familiar issues. Illicit examples are rare, labels often reflect enforcement biases, and collecting high-quality data is costly. Models inherit the blind spots of their training

Received: July 24, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

data and are likely to overlook genuinely novel behaviors that fall outside what they've seen before (Fariha et al., 2025) [9].

Another problem is evaluation. Many supervised models are tested in ways that unintentionally let future information leak into training, for example, by using features or records that wouldn't be available at the time of detection. That can make performance look better than it would be in practice. Work on public transaction data shows this clearly. Graph-based supervised methods applied to cryptocurrency datasets often report strong retrospective accuracy, but without strict temporal separation in training and testing, it's hard to claim they can truly surface new behaviors (Weber et al., 2019) [24]. And the labels in these datasets are incomplete. Many illicit activities are never caught, so models trained on them are learning to detect the patterns of known cases, not the unknown ones (Elliptic) [8].

Some efforts try blending rules and supervised models, such as using rules to pre-filter inputs before classification. While that can cut down noise, it doesn't fix the underlying issue: supervised learning is still anchored to what's already known. That's why there's growing interest in unsupervised and semi-supervised methods. These don't rely on labels and can flag unusual structures or behaviors for further review. They won't give definitive answers on their own, but they can feed analysts candidate lists worth investigating (Hu et al., 2019) [13]. Shifting from rule-based systems and fully supervised models toward more discovery-focused approaches opens new possibilities. It also raises the stakes. Success depends on enforcing temporal safeguards, building evaluation methods that mirror real-world constraints, and designing practical tools so investigators can turn a ranked list of anomalies into intelligence they can act on.

2.2 Graph-Based Anomaly Detection

Graphs fit transaction data naturally because money flows are, at their core, relationships. In this view, edges are transfers, nodes are transactions, addresses, or accounts, and the structures they form often reveal more than any single numeric feature. Early work on node representation made this clear. DeepWalk (Perozzi et al. 2014) used truncated random walks with a word2vec-style training approach to capture latent patterns in connectivity and neighborhoods. Node2vec (Grover & Leskovec 2016) expanded on that idea, introducing biased walks that move between breadth-first and depth-first strategies to capture a wider range of structural roles [21][11]. These approaches have been widely adopted in financial graph analysis because they turn network structure into compact vectors that work well with standard anomaly detectors or clustering algorithms.

Alongside embeddings, there is a substantial body of work dedicated to graph-specific anomaly detection. Akoglu et al. 2015 provide a detailed survey, organizing methods by whether they operate on node attributes, edge attributes, subgraphs, or evolving graph streams, and emphasizing the importance of explaining flagged anomalies [3]. For networks that change over time, Ranshous et al. 2015 review techniques that account for temporal evolution, noting that many static methods misinterpret natural changes as anomalies. They point to sliding-window,

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

incremental, and change-point detection frameworks as better fits for real-time monitoring [22]. Unsupervised detectors can be applied to embeddings or handcrafted graph features. Density-based methods like Local Outlier Factor (Breunig et al. 2000) spot areas where feature density drops, while isolation-based methods such as Isolation Forest (Liu et al. 2008) identify points that can be separated with few random partitions [4][20]. Each one approaches the problem from a different angle, density, separability, distance, or model residuals, which means they often flag different types of anomalies when applied to graph features.

Reviews on graph embeddings also warn that these methods differ in what they retain: network proximity, structural similarity, or spectral properties. The choice of embedding, whether random-walk, structural, or spectral, affects which patterns will stand out. Goyal & Ferrara 2018 cover these trade-offs in depth, including their relevance to anomaly detection and role discovery [10]. One recurring challenge is that embeddings are sensitive to graph dynamics. Learning them separately for different time windows can produce misaligned spaces, making it hard to track changes in suspicious behavior. Techniques for alignment or incremental embedding address this by preserving consistency across retraining, allowing analysts to follow the evolution of typologies over time instead of treating each snapshot in isolation.

2.3 Ensembles and Temporal-Safe Detection

Putting several detectors together into an ensemble is a familiar move in anomaly detection. The reasoning is straightforward: different algorithms pick up on different aspects of what makes something an "outlier," and no single method consistently wins across every dataset or threat scenario. In the outlier detection literature, ensembles come in many flavors. Feature bagging (Lazarevic & Kumar 2005), for example, trains detectors on random subsets of features to cut down on high-dimensional noise. Other approaches combine rankings or scores from different models through voting or rank aggregation. More recent work borrows ensemble theory from supervised learning and adapts it to unsupervised settings, finding ways to assign weights or consensus measures without labeled data [19].

For AML, ensembles have a natural appeal. Money laundering takes many shapes, dense clusters, isolated hubs, and cleverly hidden transaction chains, each of which might be easier to catch with a different detector (Khan et al., 2025) [16]. By fusing their outputs, you can cover more ground and improve robustness (Ahad et al., 2025) [2]. But when you're working with streaming, time-evolving transaction graphs, the design has to be more careful. You can't let future information sneak into training, and any features built from future edges have to be off-limits. Models should be trained only on what's available at the time, and evaluation needs to focus on nodes that appear after training, so the results match what an analyst would see. Both Ranshous et al. 2015 and Akoglu et al. 2015 stress the need for temporal-aware methods that treat dynamic graphs as their category in anomaly detection and warn against static evaluations that mix information from different points in time [22][3].

Beyond evaluation rules, there's also work on unsupervised fusion techniques that can estimate detector reliability without labels. Some approaches look for agreement patterns among

Received: July 24, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

detectors, while others inject small, synthetic anomalies to act as pseudo-labels. This lets the ensemble adaptively weight its members even when there's no ground truth. In practice, building an AML ensemble that works in production means getting three things right: having a diverse enough set of detectors to capture different anomaly patterns, enforcing a training and scoring setup that avoids temporal leakage and matches real-world use, and using an aggregation method, whether rank fusion, score calibration, or learned weighting on synthetic validation data, that produces rankings analysts can trust and understand. While supervised ensembles are well-mapped out in the literature, unsupervised ensembles in streaming graph environments are still maturing. Surveys of ensemble and outlier methods point to both the potential and the open challenges, from handling drift to calibrating without labels to keeping the results interpretable for human operators (Lazarevic & Kumar 2005; Aggarwal 2017) [19][1]. The consensus is that ensembles can make a real difference in spotting varied laundering patterns, but without careful, temporally safe design and evaluation, those gains risk being illusions from looking backward instead of forward.

2.4 Gaps and Challenges

Despite substantial progress in graph representation learning, anomaly detection algorithms, and ensemble strategies, important gaps remain that constrain the field's ability to deliver reliable, production-grade typology discovery systems for AML. First, many studies evaluate detection efficacy on static snapshots or use cross-validation protocols that inadvertently permit temporal leakage; this practice can inflate performance estimates because features or embeddings implicitly incorporate information that would not be available in an online deployment. Ranshous et al. 2015 explicitly document how temporal evaluation must be handled in dynamic networks to avoid such leakage and argue for sliding-window or forward-chaining setups that train only on past data [22]. Second, reproducibility is hampered by the scarcity of large, labeled, public transaction corpora that reflect the full diversity of illicit behavior; while releases such as the Elliptic dataset have catalyzed research, public labels are incomplete and proprietary datasets used in industry are rarely available for peer comparison, making it difficult to benchmark typology discovery methods across realistic threat models (Weber et, al. 2019) [24].

Third, embedding instability is an underappreciated practical problem: embeddings learned independently on successive time windows are arbitrarily rotated and scaled, which breaks naive longitudinal analysis and complicates clustering or tracking of emerging typologies unless alignment strategies (orthogonal Procrustes, incremental embedding) are explicitly applied, a point underscored in embedding surveys and empirical studies that compare static and dynamic embedding techniques (Goyal & Ferrara 2018; Grover & Leskovec 2016) [10][11]. Fourth, evaluation metrics for typology discovery are not standardized. Precision@k and analyst workload-aligned metrics are useful but do not capture recall of novel typologies unless synthetic injections or human validation are employed; while ensemble and ablation studies exist in the outlier literature, there is limited quantitative benchmarking that measures discovery

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

recall across canonical laundering strategies (smurfing, layering, hub-and-spoke) under adversarial adaptations.

Finally, scalability and explainability remain pressing concerns: methods that perform well in small academic datasets may be impractical on institution-scale transaction graphs without careful algorithmic engineering and incremental computation, and black-box fused scores fail to meet the explainability needs of investigators and regulators. Together, these gaps argue for a research agenda that couples temporal-safe protocols, reproducible public benchmarks (including synthetic injection frameworks), embedding alignment techniques, diverse unsupervised ensembles, and analyst-facing explainability so that typology discovery moves from promising prototypes to operational tools.

3. Methodology

3.1 Dataset and Feature Design

We conducted experiments using the publicly available Elliptic dataset, which consists of a time-resolved transaction graph comprising tens of thousands of node-entities and over two hundred thousand edges, spanning 49 distinct temporal steps. Each node is labeled as licit, illicit, or unknown, with the majority in the "unknown" category, reflecting the sparse labeling common in operational AML settings. To assess discovery capability for emergent typologies beyond the dataset's labels, we complement the real-world data with the synthetic injection of canonical laundering patterns. Specifically, we generate smurfing patterns by splitting a value from a source node into many low-value child nodes, layering such that value flows through a chain of intermediate nodes, and hub-and-spoke structures in which a central node distributes value to many outer nodes. These injections are added to the underlying graph before evaluation so that our pipeline's ability to surface novel typologies can be measured quantitatively. This design ensures our methodology evaluates both real-world detection and controlled typology discovery in a seamless, reproducible framework.

3.2 Data Preprocessing

To preserve real-world realism and prevent information leakage, we employ a strict time-sliced graph construction: at each evaluation step t, only nodes and edges with timestamps up to and including step t are used to build the graph. Node identifiers are standardized as strings to ensure consistency across merges and embedding lookups. Missing values in features are handled by imputing or zero-filling, depending on context. Graph-derived metrics such as degree or PageRank default to zero when undefined, while raw transaction features use median imputation to avoid bias. Crucially, we isolate test nodes strictly from future data: embedding or graph features for time t are derived only from the graph up to time t, and detectors are trained only on nodes up to time $t-1$. This temporal discipline ensures our evaluation reflects genuine detection capability on newly appearing nodes, as would happen in a production alerting system.

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.3 Feature Engineering and Exploration

We derive a combination of structural graph features and node embeddings to capture complementary information. Structural features include total degree, in-degree, out-degree, and PageRank for each node, capturing connectivity and centrality. Optionally, motif counts such as triangles or feeder chains are computed where feasible. In parallel, we compute node embeddings using node2vec, parameterized for efficiency in memory-limited environments. To ensure that embeddings remain comparable across time windows, we apply Procrustes alignment: after computing embeddings for time t, we align them orthogonally to the embeddings from prior time t−1, preserving relative geometry while maintaining continuity.

The node and edge proportions reveal a network structure where transaction relationships are dense enough to warrant graph-based modeling but sparse relative to the number of individual transactions. This density suggests the possibility of identifying anomalies through connectivity and interaction patterns without being overwhelmed by complete graph saturation. The class composition shows a substantial dominance of unknown-labeled transactions, with licit and illicit labels forming much smaller proportions. This imbalance indicates a challenge for both supervised and semi-supervised methods, as labeled positive instances (illicit) are scarce. The imbalance also hints at a realistic operational environment in anti-money laundering (AML) systems where most alerts lack definitive ground truth, reinforcing the importance of unsupervised anomaly detection approaches.

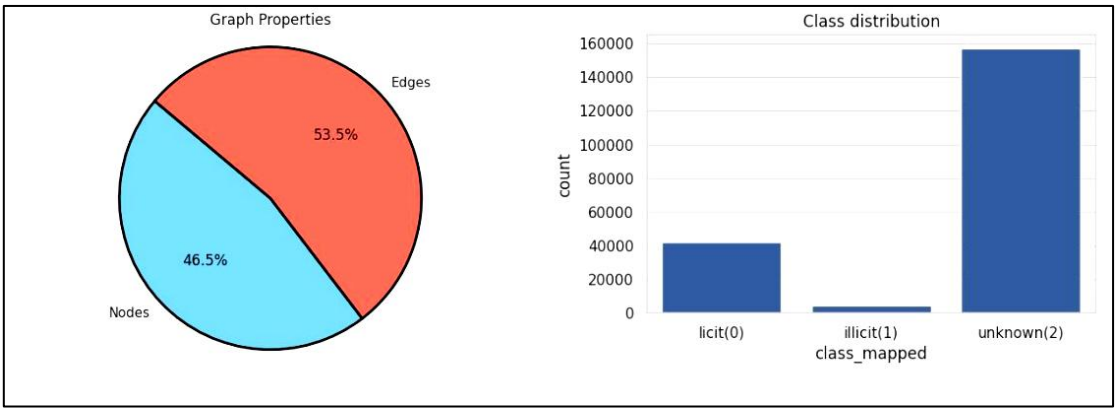

Fig. 1: Graph properties and transaction class distributions

Temporal analysis of transaction volumes shows non-uniform activity across time steps, with some periods exhibiting marked surges. These bursts could correspond to external events or coordinated laundering activities. Periods of higher transaction counts increase the likelihood of atypical transfer chains emerging, making time-aware evaluation essential to prevent misleading performance gains from future information leakage. Class trends across time reveal stability in the prevalence of unknown transactions but fluctuating patterns in licit and illicit proportions. Certain time windows show spikes in illicit labeling, suggesting concentrated laundering attempts or improved detection mechanisms during those intervals. These fluctuations could serve as temporal anchors for evaluating typology evolution and model

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

adaptability. Degree distributions illustrate a heavy-tailed pattern, where most transactions have low connectivity, and a few act as hubs with disproportionately high degrees. Such hubs may represent exchanges, mixers, or laundering intermediaries that aggregate and redistribute funds. The asymmetry between in-degree and out-degree distributions in certain ranges implies specialized roles in the transaction flow, nodes that primarily receive funds versus those that primarily disperse them.

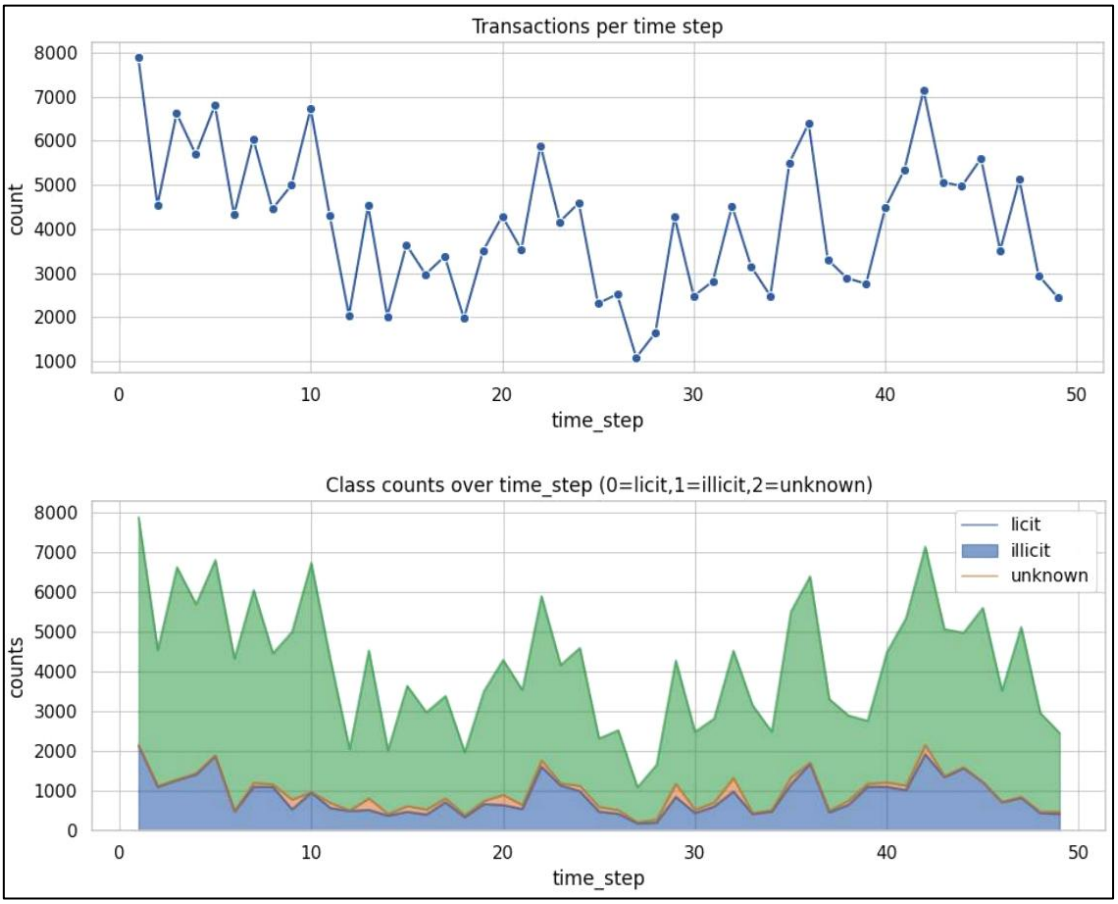

Fig. 2: Transactions per time step and transaction class counts over time step.

Connected component sizes highlight the dominance of one giant component containing the bulk of activity, alongside many smaller isolated clusters. The large connected structure facilitates money flow tracing, while the smaller disconnected clusters represent isolated laundering cells or test transactions outside mainstream activity. The duality of a massive core and small peripheries underscores the need for detection models to operate effectively in both high-connectivity and low-connectivity contexts. The correlation heatmap reveals strong positive and negative relationships among the top high-variance features in the dataset. Several feature pairs exhibit near-perfect correlation, suggesting they capture overlapping or redundant transactional patterns. This degree of multicollinearity indicates that certain attributes may be mathematically linked, possibly because they represent similar network properties derived from related transaction flows. The presence of such tight correlations can have important

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

implications for downstream modeling: without proper handling, models may overweight these redundant signals, leading to unstable predictions and reduced generalizability. Conversely, these correlations also point to stable structural behaviors in the network, hinting that certain illicit transaction patterns manifest consistently across multiple features. The weaker correlations seen between other features highlight potential complementary information sources, which, if preserved, could enhance detection performance by capturing different dimensions of network activity.

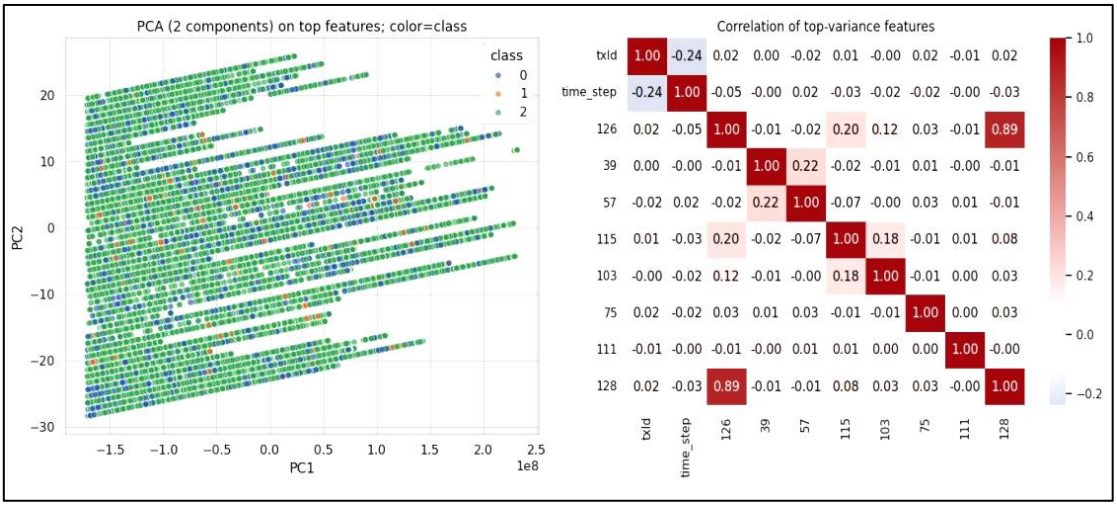

Fig. 3: PCA and correlation of top features

3.4 Predictive Modeling (Unsupervised Detectors)

We implement a suite of unsupervised anomaly detectors, each offering a different inductive perspective: (1) Isolation Forest, which isolates anomalies using random partitions; (2) Local Outlier Factor (novelty mode), detecting local density deviations for new points; (3) One-Class SVM, modeling a margin around training data to score novel observations; (4) Elliptic Envelope, modeling a robust covariance-based boundary; and (5) DBSCAN, adapted by computing distance from core training samples to score new nodes as "noise" if they fall outside density clusters. These detectors are trained at each time step using only data up to t-1 and then used to generate anomaly scores for nodes at time t. This temporal training and scoring strategy guarantees that detectors must generalize to previously unseen data without exposure to future transaction behavior.

3.5 Ensemble Learning Typology Clustering and Visualization

To combine the strengths of multiple detectors and mitigate individual biases, we normalize each detector's score by rank, transforming raw scores into a common scale, and compute a mean rank aggregation across all detectors. Optionally, weights can be applied based on performance on synthetic validation typologies, enabling tuning to prioritize detectors that better recognize certain archetypes. This rank-based fusion is robust to scaling differences between detectors and reduces sensitivity to outlier score distributions. By averaging normalized ranks, the ensemble yields a consolidated suspicion ranking for each new node,

Received: July 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

providing a smoother and more consistent prioritization for analysts compared to any single detector.

Once nodes at each time step are assigned ensemble scores, the top-ranked nodes are clustered into candidate typologies. We use either community detection on induced subgraphs or density-based clustering on embedding/feature space to form cohesively connected groups. For each cluster, representative subgraphs are extracted and visualized using network plotting tools, with node size or color indicating ensemble score. Additionally, concise feature-based explanations, highlighting standout values like high degree, elevated PageRank, or deviant features, are generated to support quick human interpretation. These artifacts, combining network visual structure and feature summaries, equip analysts with clear, interpretable typology candidates for further investigation.

AML Detection System Flow

The anti-money laundering flow might be complex to interpret, so here is a simplified overview of the whole flow. (1) The pipeline begins with raw transaction data (edges, features, and known/unknown labels). (2) Graph construction occurs for each time step, feeding into both graph features and Node2Vec embeddings. (3) These features are input to multiple anomaly detectors. (4) Detector scores are combined via ensemble scoring, producing ranked suspicious transactions. (5) The top-K results go to analysts, whose feedback is looped back to adjust thresholds and weights.

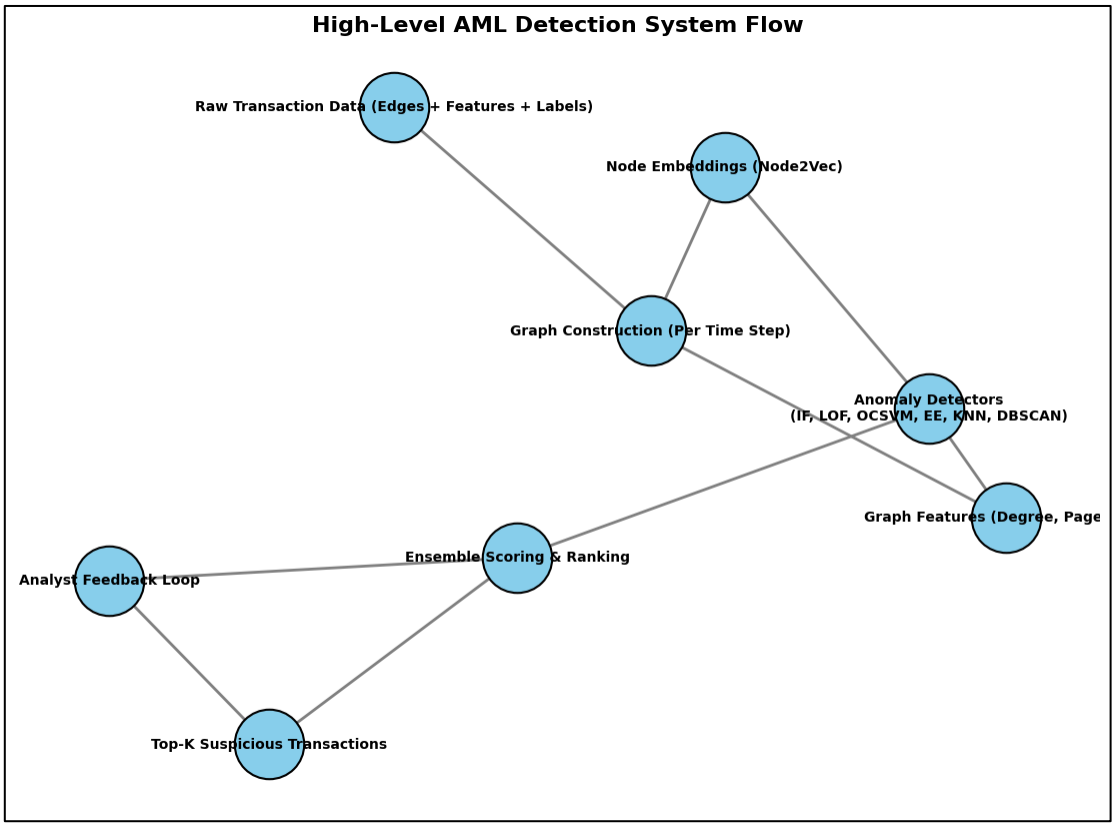

Fig. 4: Anti-money laundering detection system flow

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. Evaluation and Results

4.1 Detector and Ensemble Performance

The precision@k table shows that while individual unsupervised detectors capture different slices of anomalous behavior, the aggregated ensemble produces more consistent top-tier precision across the range of analyst-relevant operating points. Individual detectors exhibit pronounced variability: isolation-based detectors typically show stronger precision at smaller k because they are effective at isolating extreme structural outliers, whereas density-based detectors like LOF can perform better for moderate k where anomalous pockets are locally distinct but less globally separated. The ensemble's superior average performance arises because it effectively pools these complementary strengths; nodes that are jointly ranked highly by multiple detectors are more likely to represent coherent anomalies rather than spurious fluctuations in one detector's score. The bar charts and average tables reinforce that the ensemble reduces the spread in performance across operating points, offering analysts a more dependable top-K list to triage. Temporal plots for precision@25 reveal that detectors drift in performance across windows, likely due to changing background transaction patterns and the arrival of atypical batches of activity; despite this drift, the ensemble often maintains its edge, showing smaller downward swings and faster rebounds, which suggests the fusion mechanism stabilizes ranking against short-term noise and isolated detector failures. Taken together, these results imply that an analyst-aimed AML system can gain materially from ensemble fusion: it reduces reliance on a single inductive bias and produces a ranking that is both higher in average precision and more robust across temporal fluctuations.

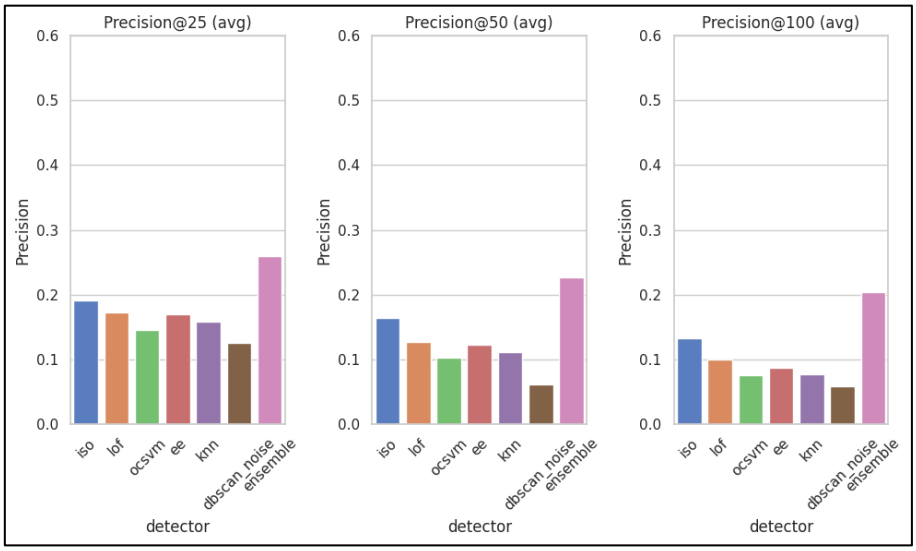

Fig. 5: Precision@k averaged across time windows

Received: July 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

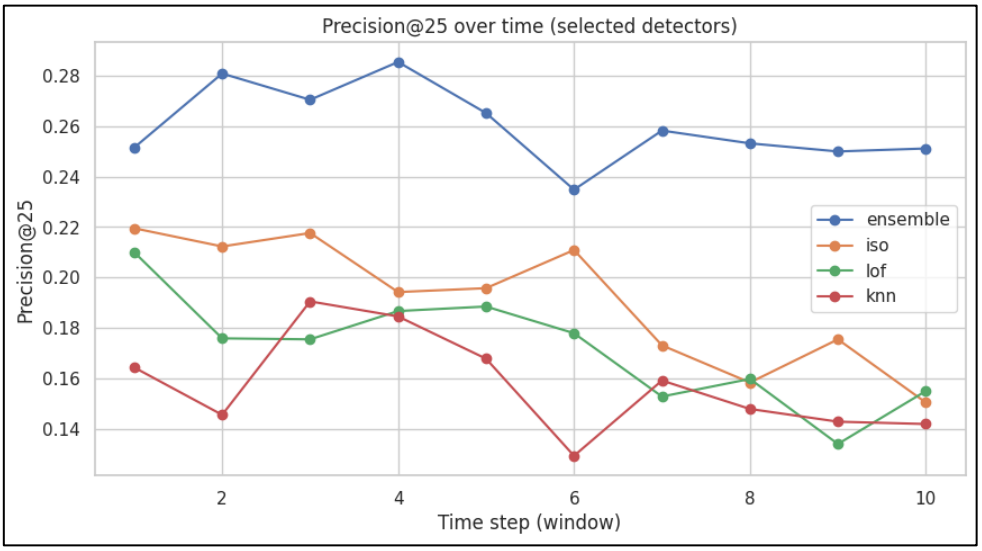

Fig. 6: Precision@25 for detectors over time

4.2 Synthetic Typology Recall

The recall breakdown by canonical typology, smurfing, layering, and hub-and-spoke, highlights how different detection methods are inherently more or less sensitive to particular structural patterns. Smurfing, which disperses funds into many low-value transactions, tends to be detected more reliably by detectors that are sensitive to degree and local density anomalies, and the ensemble improves recall here by combining both density and isolation cues. Layering, which creates long chains and intermediate nodes intended to obscure provenance, is harder for local density detectors to catch but benefits when detectors that capture global pathway disruption (e.g., k-NN distance or isolation mechanisms) are included; the ensemble again shows a measurable recall gain because it aggregates signals from detectors tuned to pathway anomalies along with local deviations. Hub-and-spoke patterns produce a characteristic centrality signature; centrality-aware features such as PageRank and degree make these patterns discernible to many methods, and the ensemble consolidates these signals, producing the highest relative recall for this typology. The per-typology table also reveals that no single detector dominates across all typologies: each shows strengths and weaknesses that depend on motif geometry. The ensemble's consistently higher recall suggests its practical value for discovery tasks where the objective is to surface a broad set of candidate typologies rather than to optimize for a single known pattern.

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

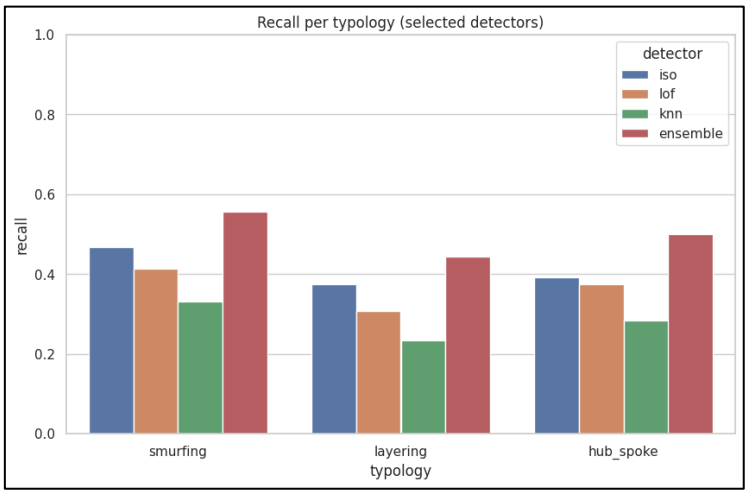

Fig. 7: Recall per typology for ensemble vs single detectors

4.3 Stability Analysis

Stability analysis focuses on whether the system presents consistent top-ranked candidates across adjacent time windows, a property that is crucial for investigator efficiency and for constructing persistent typologies over time. The heatmaps and consecutive-overlap series indicate a clear improvement in persistence when embeddings are aligned across windows: the aligned case shows larger Jaccard overlaps between the sets of top-ranked nodes in neighboring windows, and the time-series plot demonstrates that consecutive overlap values are higher and less noisy when alignment is used. This indicates that alignment preserves the geometric relationships among nodes across retrains, allowing the ensemble to recognize the same structural patterns even as the embedding model is recalculated. Without alignment, even minor rotations or reflections in the embedding space produce larger apparent churn in top-K lists: equivalent structural positions in the network can map to different vector coordinates between windows, causing downstream detectors and the ensemble to rank them inconsistently. From an operational perspective, higher overlap with alignment translates into reduced analyst churn; investigators see more stable candidate lists, enabling longitudinal investigations and more reliable typology growth tracking. The stability charts also reveal that overlap naturally decays as windows become more distant, which is expected given evolving transactional behavior; the key outcome is that alignment meaningfully slows this decay, supporting the feasibility of constructing temporally coherent typologies for human review.

Received: July 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

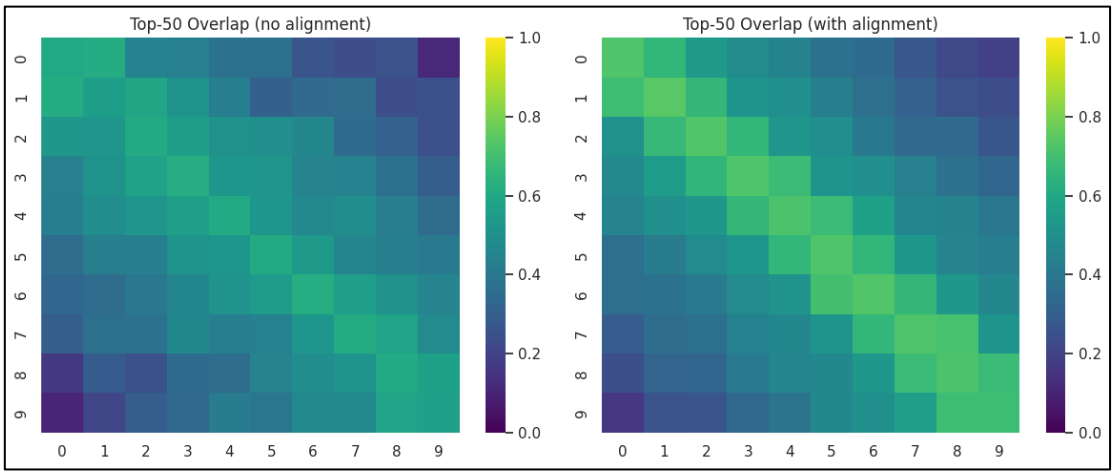

Fig. 8: Top-50 overlap stability heatmaps with and without embedding alignment

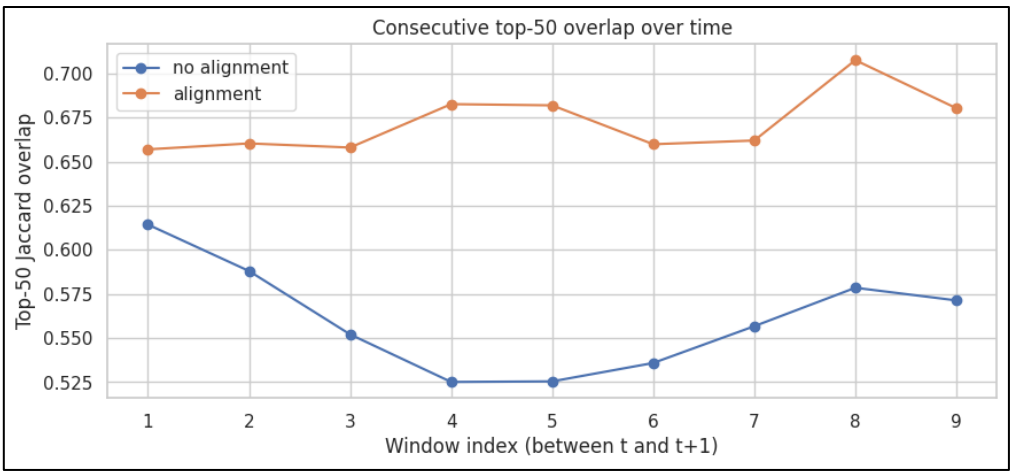

Fig. 9: Top-50 overlap scores between consecutive time steps with and without embedding alignment

1.1 Typology Visualization Examples

To make the findings interpretable for analysts, the highest-scoring transactions identified by the ensemble were clustered into candidate typologies, which were then visualized as subgraphs. For example, the visualization below shows a suspicious subgraph centered around a high-scoring transaction, tx 157607792. The color of each node indicates its suspicion score, with darker reds representing a higher risk. The diagram illustrates a hub-and-spoke pattern where a central transaction distributes funds to several other transactions, a common feature of smurfing or layering typologies. The visualization also includes annotations for key metrics, such as the total degree and PageRank of the central node, providing context for why the ensemble flagged this specific structure. By presenting these human-readable artifacts, the system moves beyond simply providing a ranked list of alerts and instead offers an actionable, visual summary that supports the analyst's investigation.

Received: July 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

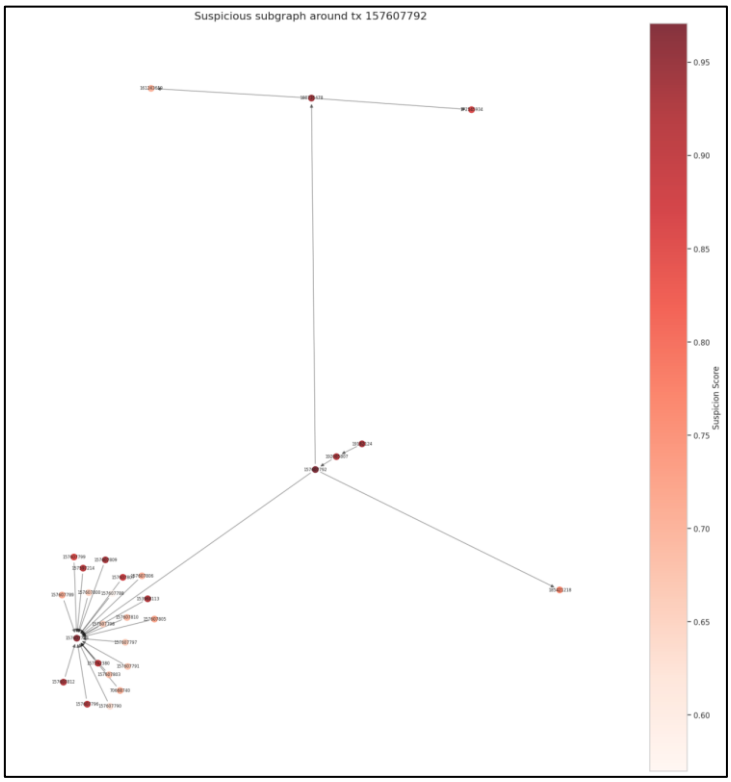

Fig. 10: Example subgraph diagram for detected clusters, annotated with key metrics.

5. Insights and Real-World Implications

5.1 Ensemble Benefits

The strength of the ensemble lies in how it brings together different ways of looking at the same data. No single suspicious signal has to carry the entire load, which means the system often spots cases that would slip past any one detector's blind spots. Take isolation-style methods, for example. They're good at catching extreme structural outliers that stand apart when the graph is randomly partitioned, but they can miss clusters of unusual activity that are tightly packed together, something density-based detectors tend to pick up. On the other hand, LOF-style detectors excel at spotting sparse local neighborhoods but can get thrown off when the broader network structure changes. By rank-normalizing and averaging across detectors, the ensemble highlights candidates where multiple independent signals agree, while toning down high scores that appear from a single detector in noisy periods. In practice, this creates two noticeable effects: first, the ensemble maintains higher and steadier precision@k even when background traffic shifts, and second, it catches a wider range of typologies because each detector is tuned to different structural patterns. It's most effective in diverse threat environments where attackers mix or rotate their evasion strategies.

That said, ensembles are not foolproof. If a single detector is highly effective for a very specific pattern and the ensemble includes weaker detectors that add noise, the averaging process can water down that strong signal and lower its ranking. This risk is greater when the ensemble is unweighted and the validation process doesn't include enough realistic injected examples to

Received: July 24, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

fine-tune detector weights. Redundancy is another issue: if many ensemble members share the same inductive bias, their diversity shrinks and the benefits of fusion disappear. Finally, because the ensemble merges scores, it can hide detector-specific failure modes, making debugging harder. Analysts may need to look at individual component scores to understand why a candidate ranked highly, which can slow triage unless the system provides clear, per-detector explanations. In short, ensembles can improve robustness and coverage in discovery settings, but they work best when they're carefully assembled, validated on a variety of typologies, and supported by tools that make each detector's role transparent (Chandola et al. 2009) [5]; their design needs to balance broad coverage with weighting and interpretability controls (Lazarevic & Kumar 2005) [19].

5.2 Operational Use in AML Investigations

In day-to-day anti-money laundering investigations, the most valuable thing an unsupervised ensemble pipeline offers is a focused list of ranked typology candidates instead of a constant flood of disconnected alerts. This changes the way analysts work. Rather than sifting through a heavy stream of one-off alerts for single transactions, they receive subgraph candidates bundled with clear feature-based explanations and example transaction chains. This approach eases the false positive load, since the fusion process tends to push down spurious outliers flagged by only one detection method. It also helps match detection output to the team's capacity. Precision@k can be linked to daily or weekly workloads, so the number of surfaced candidates aligns with what analysts can realistically investigate. This makes it easier to set service-level agreements and plan resources. Deployment can tie directly into alert and case management systems, where flagged typologies become fully formed cases with provenance data, detector score breakdowns, and visual subgraphs ready for review. A discovery-oriented pipeline also supports typology libraries that grow over time. Clusters confirmed by analysts can be stored, used to adjust detector weights, or help train supervised models for prioritization. The human-in-the-loop process ensures that the system learns from validated cases while maintaining the caution needed in compliance work.

There are still operational hurdles. Analysts need quick ways to review component scores, track how cases evolve, and mark false positives for feedback. That requires well-designed interfaces and audit logging so that both internal governance and regulatory requirements are met. From a governance standpoint, documenting the evaluation process that led to the ensemble's thresholds is essential, especially for auditors and regulators. Showing that the system was tested on forward-chained windows and synthetic injections can make those decisions easier to defend. Ultimately, the real gain is not just in the improved metrics but in reshaping the workflow, reducing wasted time, supporting ongoing typology curation, and creating a clear, defensible path from anomaly detection to investigation (Akoglu et al. 2015; Ranshous et al. 2015) [3][22].

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5.3 Stability and Analyst Trust

Analysts trust systems that behave consistently. They need to follow a candidate across time, watch it develop, and be confident that any changes reflect real activity rather than quirks from retraining the model. Embedding alignment helps with this by cutting down on arbitrary rotations or shifts in the embedding space that can happen when embeddings are re-learned independently for each time window. With alignment methods like orthogonal Procrustes or incremental embedding strategies, the spatial relationships between nodes remain stable across retrains. This steadiness filters down into the detectors, making their inputs and rankings more consistent. Two operational benefits follow. First, the overlap between top-ranked candidates is higher from one window to the next, which helps analysts build long-term cases without having to repeatedly chase the same underlying issues. Second, explanations become easier to trust if a node's position changes only because its neighborhood changed; shifts in its anomaly score are more likely to reflect genuine behavior.

This stability also supports tasks like typology growth tracking, where clusters of flagged nodes are followed over time and enriched with statistics. Without alignment, clusters may split or merge in unpredictable ways, making case continuity difficult. Alignment is not perfect; it needs enough nodes to overlap between windows for the transforms to work well, and highly volatile graphs can still create instability, but it generally improves both the quality of automated detection and analyst confidence. The result is a greater likelihood that flagged candidates represent something real and that workflows stay consistent over time (Goyal & Ferrara, 2018; Grover & Leskovec, 2016) [10][11].

5.4 Limitations

While this pipeline moves typology discovery forward, several limitations are worth noting. First, uncertainty in dataset labels limits how far detection performance claims can go. Public datasets like Elliptic are valuable for benchmarks, but their labeled portion is incomplete and may reflect biases from enforcement actions. Precision and recall measured on these labels should be read as conservative estimates of real-world performance. Second, richer transactional attributes, like amounts in native currency, counterparty details, finer time resolution, or off-chain context, would likely improve detection and characterization, but these are often unavailable in public data and restricted in industry. As demonstrated in cryptocurrency forecasting research (Islam et al., 2025), the absence of features like fine-grained temporal sequences and market-context variables fundamentally limits predictive fidelity in financial behavior modeling [14]. This makes it necessary to carefully adapt academic findings to production settings where richer data exists. Third, while injecting synthetic typologies is useful for controlled recall testing, their realism is limited by the heuristics used to generate them. They can mimic patterns like smurfing or layering, but they can't perfectly reproduce the creativity of actual adversaries who exploit domain-specific quirks. Results from injected typologies should be seen as indicative, not definitive.

Received: July 24, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fourth, scaling is a practical concern. For large transaction graphs, both embedding and scoring can demand significant resources. Production systems may need incremental embedding, sampling, or distributed processing to keep pace. Finally, attackers adapt. If they notice their activity is being detected, they may subtly change their behavior to evade detection. Staying effective requires continuous monitoring, simulated adversary tests, and feedback from analysts. Taken together, these limitations suggest that the pipeline works best as a discovery aid for analysts, not as an automated decision-maker. It can surface interpretable, prioritized candidates that reduce workload, but it still depends on ongoing tuning, access to richer data, and governance measures to turn those discoveries into lasting AML improvements (Chandola et al. 2009; Aggarwal 2017) [5][1].

6. Future Work

6.1 Integration with Real-World Financial Data In The U.S.

The next logical move is to pilot the pipeline with private, institution-grade transaction datasets that carry richer context than public benchmarks in the USA. These would include detailed counterparty profiles, transaction narratives, and fine-grained temporal data. Such datasets capture patterns you rarely see in open data, like multi-channel transfers, cross-asset flows, and jurisdictional routing behaviors. Running pilots under compliance constraints would mean working closely with regulated financial institutions while staying firmly within the boundaries of KYC/AML regulations, GDPR, and data residency rules. The aim is to see how the pipeline holds up with production-scale data, where the graphs are bigger, features are more varied, and alert volumes are much heavier than in research environments. This phase would also make it possible to tune detector parameters and ensemble thresholds to match the specific transaction profiles of an institution, which could improve detection rates while keeping false positives in check. Experience has shown that moving from synthetic or public data to operational datasets often uncovers additional feature interactions and domain-specific patterns that sharpen model sensitivity (Weber et al., 2019) [24]. To make it work, secure and compliant data pipelines will be essential, along with privacy-preserving computation approaches such as federated learning or secure enclaves (Yang et al., 2019) [25].

6.2 Incorporation of GNN-based Autoencoders

Graph Neural Network autoencoders could be a strong addition for capturing both local and global structural patterns in transaction networks. Unlike shallow embedding methods like node2vec, GNN-based approaches such as Graph Convolutional Autoencoders (Kipf & Welling, 2016) [18] and Variational Graph Autoencoders can directly incorporate node and edge attributes while learning latent representations tuned for reconstruction or anomaly detection. These models can adapt as the graph evolves, making them potentially more resilient to sparse or noisy features. In the AML context, GNN autoencoders could produce embeddings that capture the deeper semantics of transaction chains and counterparty connections, improving detection of subtle or adaptive laundering strategies. Adding them to the ensemble would require benchmarking against current embedding pipelines under temporally safe setups

Received: July 24, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

to avoid future leakage. Research shows that GNNs have outperformed traditional embeddings in financial fraud detection, particularly in cases involving complex relationships and multi-hop dependencies (Zhang et al., 2022) [26].

6.3 Active Learning with Analyst Feedback

Bringing an active learning loop into the pipeline could create a practical way to keep models sharp once they're in production. Analysts would review a portion of the alerts where model confidence is low, cases where detectors disagree, or where ensemble scores hover near operational thresholds, and provide either labels or relevance feedback. These inputs could be used to adjust ensemble weights, recalibrate thresholds, and retrain detectors so they focus on typologies most relevant to the institution's risk profile. This kind of iterative exchange helps maintain a balance between catching emerging risks and ignoring harmless anomalies, which in turn reduces false positives without losing recall on new patterns. Active learning has already proven its worth in fields like cybersecurity and healthcare, where expert input is scarce but highly valuable (Settles, 2010) [23]. In AML, such feedback could also help grow a living typology library, a continuously updated knowledge base that supports both detection and regulatory reporting.

6.4 Adversarial Robustness Studies

To prepare for adaptive adversaries, it will be important to test the pipeline's resilience against attacks designed to slip past detection. This could mean generating adversarial graph modifications such as inserting camouflage edges, tweaking transaction values, or breaking up laundering chains into fragments that look benign (Das et al., 2025) [7]. Methods from adversarial machine learning, including gradient-based perturbations tailored to graphs (Dai et al., 2018) [6], could be used to expose weaknesses in the system. Testing under these conditions would reveal where detector diversity or ensemble weighting needs improvement and whether defenses like adversarial training or randomized feature changes should be added. Generative models capable of creating realistic evasive patterns could expand the synthetic injection process, ensuring the pipeline is ready not only for current laundering tactics but also for likely future ones.

7. Conclusion

This study tackled one of the tougher problems in anti–money laundering in the USA: spotting new laundering patterns in financial transaction networks before they're recognized and labeled. Traditional supervised models tend to stumble here because they lean heavily on past examples and rigid rules that don't adapt quickly enough. To address this, we built a temporally safe, unsupervised ensemble pipeline that blends several anomaly detectors with graph-based feature engineering and embedding alignment. The aim was to flag suspicious transactions early, group them into meaningful typologies, and make the results something an analyst could interpret. We tested this on the Elliptic dataset, enforcing strict time-sliced evaluation so the models couldn't peek into the future. To push the limits further, we added synthetic laundering patterns to the data, cases like smurfing, layering, and hub-and-spoke structures, so the system had to

Received: July 24, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

deal with both known and novel schemes. The approach outperformed single detectors, with higher precision@k and stronger recall. Two parts of the method were especially important: rank-based ensemble fusion, which reduced the risk of one detector dominating the results, and Procrustes alignment, which kept the embeddings consistent over shifting time windows. That consistency matters a lot for maintaining trust in a live investigative setting.

The implications go beyond academic curiosity. By focusing on forward-looking detection instead of relying on old labels, this pipeline can give compliance teams alerts that adapt to how laundering methods evolve. Bringing together different anomaly signals means fewer false positives and more coherent threat profiles that match investigative capacity. The inclusion of synthetic typologies and time-aware evaluation also means there's a reproducible way to benchmark anti–money laundering systems under realistic pressure. In terms of contribution, the work delivers three things: first, a clear framework for unsupervised, time-robust typology discovery in transaction graphs; second, empirical proof that ensemble fusion and embedding alignment improve both stability and coverage; and third, a practical route for bringing adaptive AML systems into day-to-day operations. Next steps involve scaling to institutional-level datasets, using graph neural networks for richer features, and making the system more resilient against adversarial tactics. Taken together, the results point to how unsupervised ensembles could help shift AML from rigid, rule-bound detection toward a more proactive, intelligence-driven approach.

Abbreviations

AML – Anti-Money Laundering

DBSCAN – Density-Based Spatial Clustering of Applications with Noise

EE – Elliptic Envelope

EDA – Exploratory Data Analysis

GNN – Graph Neural Network

IF – Isolation Forest

KNN – k-Nearest Neighbors

LOF – Local Outlier Factor

OCSVM – One-Class Support Vector Machine

PCA – Principal Component Analysis

PR – PageRank

ROC – Receiver Operating Characteristic

STK – Sim Tool Kit (mobile payment push service)

TPR – True Positive Rate

Received: July 24, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

[1] Aggarwal, C. C. (2017). Outlier Analysis. Springer.

[2] Ahad, M. A., Mohaimin, M. R., Rabbi, M. N. S., Abed, J., Shatyi, S. S., Sadnan, G. A., ... & Ahmed, M. W. (2025). AI-Based Product Clustering For E-Commerce Platforms: Enhancing Navigation And User Personalization. International Journal of Environmental Sciences, 156-171.

[3] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based Anomaly Detection and Description: A Survey. Data Mining and Knowledge Discovery.

[4] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. ACM SIGMOD / PVLDB.

[5] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys.

[6] Dai, H., et al. (2018). Adversarial attack on graph structured data. Proceedings of the 35th International Conference on Machine Learning (ICML).

[7] Das, B. C., Sartaz, M. S., Reza, S. A., Hossain, A., Nasiruddin, M. D., Bishnu, K. K., ... & Abed, J. (2025). AI-Driven Cybersecurity Threat Detection: Building Resilient Defense Systems Using Predictive Analytics. arXiv preprint arXiv:2508.01422.

[8] Elliptic, www.elliptic.co.

[9] Fariha, N., Khan, M. N. M., Hossain, M. I., Reza, S. A., Bortty, J. C., Sultana, K. S., ... & Begum, M. (2025). Advanced fraud detection using machine learning models: enhancing financial transaction security. arXiv preprint arXiv:2506.10842.

[10] Goyal, P., & Ferrara, E. (2018). Graph Embedding Techniques, Applications, and Performance: A Survey. Knowledge-Based Systems.

[11] Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks. KDD / arXiv.

[12] Gürsoy, M., Haddad, M., & Bothorel, C. (2021). Alignment and stability of embeddings: measurement and inference improvement. arXiv.

[13] Hu, G., et al. (2019). Characterizing and Detecting Money Laundering Activities on the Bitcoin Network. arXiv preprint.

[14] Islam, M. Z., Rahman, M. S., Sumsuzoha, M., Sarker, B., Islam, M. R., Alam, M., & Shil, S. K. (2025). Cryptocurrency Price Forecasting Using Machine Learning: Building Intelligent Financial Prediction Models. arXiv preprint arXiv:2508.01419.

[15] Kandanaarachchi, S. (2021). Unsupervised Anomaly Detection Ensembles using Item Response Theory. arXiv.

[16] Khan, M. A. U. H., Islam, M. D., Ahmed, I., Rabbi, M. M. K., Anonna, F. R., Zeeshan, M. D., ... & Sadnan, G. M. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. arXiv preprint arXiv:2506.19870.

[17] Khan, M. N. M., Fariha, N., Hossain, M. I., Debnath, S., Al Helal, M. A., Basu, U., ... & Gurung, N. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. International Journal of Environmental Sciences, 1792-1811.

Received: July 24, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[18] Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. arXiv preprint arXiv:1611.07308.
[19] Lazarevic, A., & Kumar, V. (2005). Feature Bagging for Outlier Detection. KDD.
[20] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. IEEE ICDM.
[21] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online Learning of Social Representations. KDD.
[22] Ranshous, S., et al. (2015). Anomaly Detection in Dynamic Networks: A Survey. Wiley Interdisciplinary Reviews: Computational Statistics.
[23] Settles, B. (2010). Active learning literature survey. University of Wisconsin-Madison.
[24] Weber, M., Böhme, R., & Breuker, D. (2019). Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. arXiv.
[25] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology.
[26] Zhang, J., et al. (2022). Financial fraud detection with graph neural networks: A survey. IEEE Transactions on Knowledge and Data Engineering.

Received: July 24, 2025