International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 03 | Mar 2026 www.irjet.net p-ISSN: 2395-0072

From Detection to Investigation: Calibrated Synthetic Graphs for Reliable AML using Graph Neural Network

Mohammed Yasar[1], B. Swathi[2]

[1]Student, Dept. of CSE(AI&ML), Andhra Loyola Institute of Engineering & Technology, India -12
[2]Assistant Professor, Dept. of CSE(AI&ML), Andhra Loyola Institute of Engineering & Technology, India -12

Abstract - *Anti-Money Laundering (AML) systems increasingly rely on graph-based learning to detect complex financial crime patterns. However, existing research often evaluates models on synthetic datasets that contain strong structural biases, leading to unrealistically high performance. This paper presents a calibrated AML pipeline that combines synthetic data generation, Graph Neural Network (GNN) detection, and a post-hoc investigation layer. We introduce a systematic methodology for reducing structural determinism in synthetic transaction graphs through degree stratification, structural noise injection, cross-contamination, and embedded path construction. These techniques progressively transform the dataset from trivial to realistic, reducing inflated AUC scores from 0.99 to 0.95 while restoring meaningful classification performance (F1 = 0.55). A 2-layer Graph Attention Network (GAT) is trained on structural features and evaluated against a statistical baseline, achieving significant improvements in detection performance. Additionally, an investigation layer converts alerts into structured case reports using motif detection, temporal analysis, and similarity retrieval. The results demonstrate that proper dataset calibration is essential for reliable evaluation of graph-based AML systems and that integrating detection with explainability produces actionable intelligence for analysts.*

Keywords: **Anti-Money Laundering, Graph Neural Networks, Synthetic Data, Graph Attention Network, Fraud Detection, Explainable AI, Financial Networks**

1. INTRODUCTION

Modern financial systems are increasingly supported by advanced digital infrastructure, including large-scale transaction processing platforms, real-time monitoring systems, and big data analytics frameworks. Among these, transaction networks form the backbone of global financial operations. However, one of the most critical challenges faced by these systems is the detection of financial crimes such as money laundering. This challenge arises due to the complex, multi-hop nature of transactions and the ability of malicious actors to disguise illicit flows within large volumes of legitimate activity. Therefore, there is a growing need for intelligent Anti-Money Laundering (AML) solutions, which can be effectively addressed using graph-based machine learning techniques that capture relational and structural dependencies.

Statistics indicate that a significant proportion of illicit financial activity goes undetected due to limitations in traditional monitoring systems. Existing approaches primarily rely on rule-based detection or tabular machine learning models, which fail to capture the interconnected nature of financial transactions. Currently, most systems operate on isolated transaction records without leveraging the full transaction graph, leading to high false positive rates and missed detection of coordinated laundering patterns. To address these limitations, there is a need for a scalable and intelligent AML framework that utilises advanced techniques such as Graph Neural Networks (GNNs) to model transaction flows [1][2]. The proposed system focuses on identifying suspicious accounts by learning structural and temporal patterns within transaction graphs. It leverages a Graph Attention Network (GAT) architecture, which enhances representation learning by assigning importance weights to neighbouring nodes, allowing the system to capture subtle relational signals indicative of laundering behaviour.

2. LITERATURE SURVEY

[1] The objective of the paper "Semi-Supervised Classification with Graph Convolutional Networks" by Kipf and Welling was to introduce Graph Convolutional Networks (GCNs) for learning on graph-structured data. The study proposed a spectral-based convolutional approach that aggregates feature information from neighbouring nodes to perform node classification. The results demonstrated that GCNs can effectively capture structural relationships in graph data and achieve strong performance on benchmark datasets. However, the model assigns equal importance to all neighbouring nodes during aggregation, which limits its ability to distinguish between relevant and irrelevant connections in complex financial networks.

[2] The aim of the paper "Graph Attention Networks" by Veličković et al. was to improve graph representation learning by incorporating attention mechanisms into graph neural networks. The proposed model computes attention coefficients to assign different weights to neighbouring nodes, allowing the network to focus on more informative connections. The results showed improved performance over traditional GCNs, particularly in heterogeneous graph environments. However, a limitation of this approach is the increased computational complexity due to multi-head attention mechanisms, which can impact scalability in large transaction networks.

[3] The objective of the paper "Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics" by Weber et al. was to apply graph-based learning techniques to detect illicit transactions in cryptocurrency networks. The study utilised transaction graphs and applied graph convolutional models to classify suspicious entities. The results demonstrated that graph-based approaches significantly outperform traditional methods in capturing laundering behaviour. However, the dataset used in the study exhibited structural biases, which may lead to overestimation of model performance in real-world scenarios.

[4] The objective of the paper "Temporal Graph Networks for Deep Learning on Dynamic Graphs" by Rossi et al. was to extend graph neural networks to handle temporal dynamics in evolving networks. The proposed model incorporates time-aware embeddings to capture changes in node interactions over time. The results showed improved performance in dynamic environments such as communication and financial networks. However, the model introduces additional complexity in training and requires careful handling of temporal dependencies, which can be challenging in large-scale AML systems.

3. PROPOSED WORK

The proposed AML system is designed as a multi-stage pipeline that integrates data generation, detection, and investigation into a unified framework. Each component is carefully designed to address specific limitations observed in existing approaches.

3.1 Synthetic Data Generation and Calibration

The system begins with the generation of a synthetic transaction network intended to simulate realistic financial activity. Constructing realistic graph datasets requires careful structural calibration and sampling strategies [9]. Unlike naive synthetic datasets that exhibit overly clean or isolated structures, the proposed approach introduces several calibration mechanisms to improve realism.

These include degree-stratified sampling to prevent hub dominance, controlled transaction density to avoid unrealistic clustering, and the embedding of laundering paths within normal transaction flows. Additionally, community structures are adjusted to ensure that laundering entities are not trivially separable from normal accounts.

As a result of these calibration steps, the final dataset achieves a balanced structural composition, with over 397,000 transactions and approximately 39,500 nodes. The test graph alone contains more than 31,000 nodes and 37,000 edges, providing a sufficiently complex environment for evaluating detection performance.

3.2 GNN-Based Detection Layer

The detection component utilises a Graph Attention Network (GAT) architecture [2], chosen for its ability to dynamically weigh the importance of neighbouring nodes during feature aggregation. The model consists of two layers, with multiple attention heads in the first layer to capture diverse structural patterns. Learning structural representations in graphs has been widely explored using embedding techniques such as DeepWalk and struc2vec [7][8].

Each node in the graph is represented using a set of ten structural features that capture both connectivity and transactional behaviour. These features include degree-based metrics, centrality measures such as PageRank and betweenness centrality, clustering coefficients, and transaction volume characteristics.

The model is trained using a weighted binary cross-entropy loss function, which is commonly used in imbalanced classification tasks [10]. to address class imbalance, as laundering accounts typically represent a small fraction of the overall dataset. Optimisation is performed using the Adam optimiser, with early stopping applied to prevent overfitting.

The choice of Graph Attention Network (GAT) is motivated by the heterogeneous nature of financial transaction graphs, where not all neighbouring nodes contribute equally to the behaviour of an account. In laundering scenarios, certain connections—such as intermediary accounts in layering patterns—carry significantly more importance than others.

Unlike Graph Convolutional Networks (GCNs), which assign uniform importance to all neighbours, GAT dynamically learns attention weights, enabling the model to prioritise structurally and behaviorally relevant connections. This is particularly important in AML, where illicit patterns are often embedded within noisy and highly connected networks.

3.3 Investigation Layer

A key contribution of this work is the introduction of an investigation layer that operates on top of the detection outputs. Instead of treating model predictions as standalone alerts, this layer transforms them into structured case reports that provide contextual and interpretable insights.

For each flagged account, a local subgraph is extracted using a breadth-first search strategy, with a strict upper bound on the number of nodes to ensure computational efficiency. This subgraph represents the immediate transactional neighbourhood of the account and serves as the basis for further analysis.

Motif detection techniques are then applied to identify common laundering patterns such as circular money flows, which align with graph-based anomaly detection frameworks [6]. Layering structures involving multi-hop paths, and smurfing behaviour characterised by aggregation of small transactions. In parallel, temporal analysis is performed to detect anomalies [10]. Such as bursts of activity within short time windows and rapid transaction chains with minimal delays between successive transfers.

The findings from these analyses are used to generate deterministic explanations, ensuring that each case report clearly reflects the underlying signals detected by the system. Additionally, a similarity retrieval mechanism is implemented using vector representations of cases, allowing the system to identify and present previously observed cases with similar characteristics.

This investigation layer significantly enhances the practical usability of the system by bridging the gap between detection and human analysis**.**

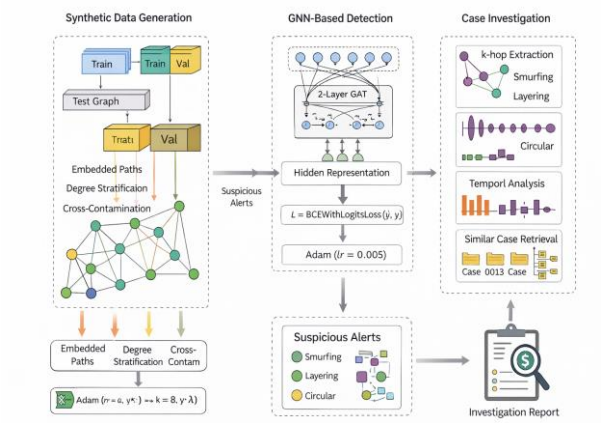

Fig -1: System Architecture

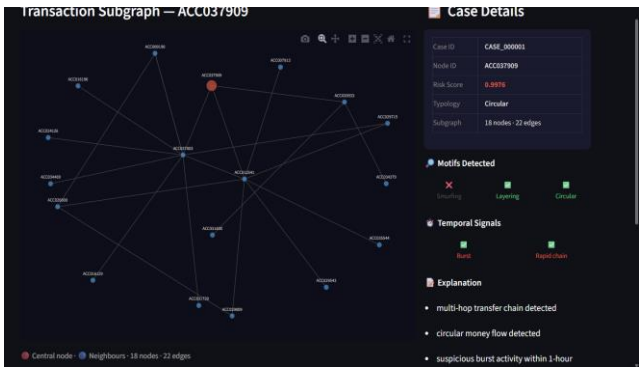

Fig -2: Investigation Dashboard

4. RESEARCH & METHODOLOGY

The research methodology adopted in this work follows a systematic and iterative approach that combines data engineering, graph-based learning, and post-hoc analytical validation. The objective is not only to build a high-performing detection model but also to ensure that the learned patterns are meaningful, unbiased, and applicable to realistic Anti-Money Laundering (AML) scenarios. The methodology is structured into three key phases: dataset construction and calibration, model development and evaluation, and investigation-driven validation**.**

4.1 Dataset Construction and Experimental Setup

The foundation of this research lies in the creation of a synthetic transaction dataset designed to stimulate real-world financial networks, as realistic graph construction is a key challenge in anomaly detection systems [9], while avoiding the common pitfalls of overly simplified or biased data. The dataset is generated through a controlled pipeline that incorporates multiple laundering typologies, including smurfing, layering, and circular transactions, alongside a dominant proportion of legitimate activity.

To ensure robustness, the dataset is partitioned into training, validation, and test splits. This separation is strictly enforced throughout the pipeline, particularly during graph construction, to prevent information leakage between stages. The final dataset consists of approximately **397,580 transactions**, distributed across **39,500 nodes**, with the test graph containing over **31,700 nodes and 37,500 edges**. This scale enables meaningful evaluation of graph-based models under conditions that resemble real financial systems.

A critical aspect of the methodology is the calibration of structural properties within the dataset. Multiple iterations are performed to eliminate trivial signals that could artificially inflate model performance. These include controlling node degree distributions, regulating transaction density, and embedding laundering behaviour within legitimate activity. The effectiveness of these calibrations is assessed using a Community Ambiguity Report, which measures metrics such as internal edge ratios, external neighbour rates, and transaction density ratios. This ensures that the dataset presents a challenging yet realistic classification task.

4.2 Model Development and Training Strategy

The detection model is developed using a Graph Attention Network (GAT) architecture, which extends graph convolutional approaches by incorporating attention mechanisms [2]. Implemented in PyTorch Geometric. The choice of GAT is motivated by its ability to assign adaptive importance to neighbouring nodes, allowing the model to focus on relevant transaction patterns within complex graph structures.

The model operates at the node level, where each node represents an account and edges represent transactions. Input features are derived from structural and transactional properties of the graph, including centrality measures and flow-based attributes. These features are normalised and fed into the network to ensure stable training.

Training is conducted using a weighted binary cross-entropy loss function to address class imbalance, as laundering nodes constitute a small fraction of the dataset. A positive class weight is introduced to penalise misclassification of laundering nodes more heavily than legitimate ones. The optimisation process uses the Adam optimiser with a learning rate of 0.005, and training is performed for a maximum of 150 epochs with early stopping based on validation loss.

To ensure reproducibility and stability, multiple training runs are conducted, and performance is evaluated using standard metrics, including precision, recall, F1-score, and Area Under the ROC Curve (AUC). A baseline statistical model is also implemented to provide a point of comparison, enabling a clear assessment of the benefits introduced by the GNN-based approach.

To establish a meaningful comparison, a statistical baseline model is implemented using logistic regression trained on the same node-level feature set used by the GNN. The features include degree-based metrics, centrality measures, and transaction flow attributes.

The baseline model operates independently on node features without leveraging graph connectivity or neighbourhood information. This setup ensures that performance improvements observed in the GNN model can be attributed to its ability to capture relational dependencies rather than differences in feature representation.

4.3 Iterative Calibration and Bias Mitigation

A key methodological contribution of this work is the iterative refinement process applied to both the dataset and the model. Rather than relying on a single static dataset, the system undergoes multiple calibration phases, each designed to identify and eliminate sources of bias that could lead to misleading performance**.**

In early iterations, the model exhibited near-perfect performance due to the presence of easily identifiable structural patterns, such as highly cohesive laundering communities and hub-dominated activity. To address this, several corrective mechanisms were introduced, including degree-stratified sampling, cross-contamination between legitimate and laundering nodes, and the introduction of sham motifs to mimic illicit structures within legitimate activity.

Further refinements involved embedding laundering paths within normal transaction flows and controlling the frequency and placement of intermediary nodes. These changes significantly increased structural ambiguity, forcing the model to rely on deeper relational patterns rather than superficial cues. The effectiveness of each calibration step was evaluated through changes in model performance across successive experimental phases, ensuring that improvements were driven by genuine learning rather than dataset artefacts**.**

Table -1: Improvement Through Calibration

Sprint Evaluation			
Sprint	**AUC**	**F1**	**Insight**
3A	0.9925	0.6043	Overfitting
3B	0.9710	0.5820	Improved sampling
3C	0.9421	0.0612	Data leakage issue
3D	0.9745	0.0844	Partial recovery
3E	0.9583	0.5515	Final stable model

4.4 Evaluation and Validation Framework

The evaluation framework is designed to assess both the predictive performance of the detection model and the practical usability of the system in an AML context. Performance metrics are computed on the test set to ensure unbiased evaluation, with particular emphasis on F1-score and recall, as these metrics are critical for identifying illicit activity in imbalanced datasets, a widely studied challenge in fraud and anomaly detection tasks [10].

The final model achieves a precision of 0.5515, a recall of 0.5515, an F1-score of 0.5515, and an AUC of 0.9583, significantly outperforming the baseline model, which achieves an F1-score of 0.1713 and an AUC of 0.7167. These results demonstrate the effectiveness of the graph-based approach in capturing complex laundering patterns.

Beyond quantitative metrics, the methodology incorporates qualitative validation through the investigation layer. Each detected alert is converted into a case report, and the consistency of detected motifs and temporal patterns is analysed across cases. The distribution of typologies and anomalies is examined to ensure alignment with expected laundering behaviours, providing an additional layer of validation that goes beyond numerical performance.

4.5 Investigation-Centric Validation and Case Analysis

The final phase of the methodology focuses on validating the system from an investigative perspective. Rather than treating detection as the end goal, the system evaluates how effectively detected alerts can be transformed into meaningful and interpretable cases.

Each alert generated by the GNN is processed through the investigation pipeline, resulting in a one-to-one mapping between alerts and case reports. For the final dataset, 449 alerts produce 449 structured cases, each containing detailed information about subgraph structure, detected motifs, temporal anomalies, and explanatory reasoning.

The distribution of detected typologies shows a dominance of circular patterns, followed by layering and smurfing behaviours, reflecting the diversity of laundering strategies embedded within the dataset. Temporal anomaly analysis further reveals a significant presence of burst activity and rapid transaction chains, indicating coordinated and time-sensitive financial movements.

Additionally, the integration of a similarity retrieval mechanism allows cases to be compared based on structural and temporal characteristics, similar to representation-based anomaly detection approaches [6][10]. This not only validates the consistency of detected patterns but also enhances the system's practical applicability by enabling analysts to identify recurring laundering strategies.

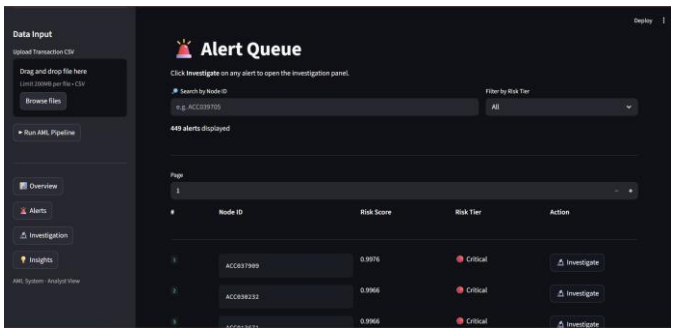

Fig -3: Alert Queue For Investigation

5. RESULTS AND DISCUSSION

The performance of the proposed AML system is evaluated across multiple dimensions, including classification accuracy, robustness to structural ambiguity, and the quality of investigation outputs. The evaluation is conducted on the test split to ensure that the results reflect true generalisation rather than memorisation of training patterns.

5.1 Detection Performance Comparison

Table -2: Baseline vs GNN Performance

Metric	Baseline	GNN
Precision	0.3187	0.5515
Recall	0.1172	0.5515
F1 Score	0.1713	0.5515
AUC	0.7167	0.9583

The results clearly demonstrate that the GNN significantly outperforms the baseline across all evaluation metrics. While the baseline achieves an F1-score of 0.1713, the GNN improves this to 0.5515, indicating a substantial increase in the model's ability to correctly identify laundering nodes. Similarly, the AUC improves from 0.7167 to 0.9583, showing that the GNN effectively captures complex relational patterns that are not accessible to traditional methods.

5.2 Investigation Output Analysis

Fig -4: Dashboard Interface

The system generates a total of 449 alerts and corresponding case reports, each containing detailed structural and temporal analysis. The distribution of detected typologies shows that circular patterns dominate, followed by layering and smurfing behaviours.

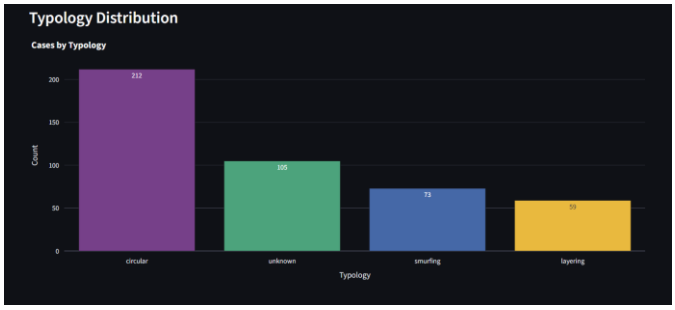

Chart -1: Topology Distribution

This distribution aligns with the embedded campaign design, where circular and multi-hop patterns are more prevalent. The presence of "unknown" cases indicates that the system is capable of identifying suspicious behaviour even when it does not strictly match predefined typologies.

5.3 Temporal Pattern Analysis

Temporal behaviour plays a crucial role in identifying coordinated financial activity. The system analyses transaction timestamps within subgraphs to detect anomalies such as bursts and rapid chains.

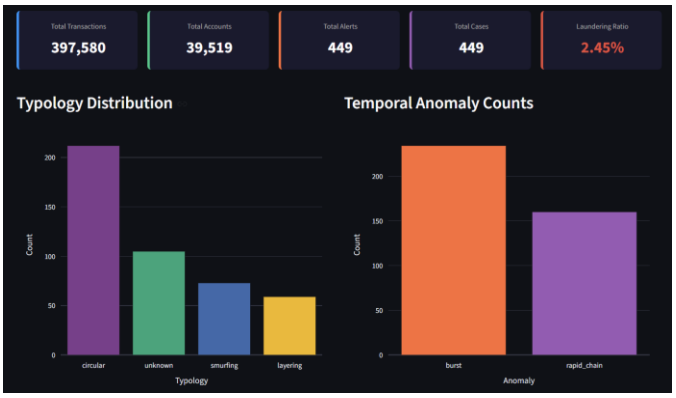

Chart -2: Temporal Patterns

5.4 Risk & Subgraph Distribution Analysis

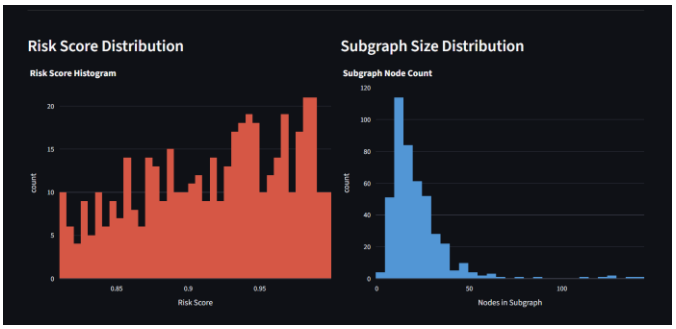

Chart -3: Risk & Subgraph Distribution

The risk score distribution shows that most detected nodes fall within a high-risk range, confirming the effectiveness of the model's scoring mechanism. At the same time, the variation in subgraph sizes indicates that the system is capable of handling both small localised patterns and larger transaction networks.

5.5 Explainability Validation

To evaluate the effectiveness of the investigation layer, qualitative analysis is performed on the generated case reports. Each case includes detected motifs, temporal anomalies, and structural context derived from subgraph extraction.

The consistency of detected patterns across cases is analysed, and the distribution of typologies aligns with the embedded laundering strategies in the dataset. Circular patterns dominate, followed by layering and smurfing behaviours, indicating that the system successfully identifies diverse laundering techniques.

Additionally, similarity-based case retrieval produces clusters of structurally similar cases, providing further validation that the system captures meaningful and repeatable patterns rather than random anomalies.

6. CONCLUSION

This work presents a comprehensive Anti-Money Laundering (AML) framework that integrates synthetic data generation, graph-based detection, and explainable investigation into a unified pipeline. Unlike conventional approaches that rely heavily on static rules or simplistic datasets, the proposed system emphasises realism, structural ambiguity, and interpretability as core design principles.

A major contribution of this research lies in the construction of a calibrated synthetic dataset that avoids trivial structural signatures. Through iterative refinement, including degree stratification, cross-contamination, and embedded laundering paths, the dataset evolves into a challenging benchmark where illicit activity is deeply integrated within legitimate transaction flows. This ensures that model performance reflects genuine pattern recognition rather than exploitation of artificial artefacts.

The detection layer, built using a Graph Attention Network (GAT), demonstrates strong capability in capturing complex relational dependencies within the transaction graph. The final model achieves a significant improvement over the baseline, with an F1-score of 0.5515 and an AUC of 0.9583, indicating robust performance under realistic conditions. More importantly, the model avoids the pitfall of near-perfect accuracy observed in earlier iterations, confirming that the learned representations are meaningful and generalizable.

Beyond detection, the introduction of an investigation layer transforms the system into a practical analytical tool. By generating structured case reports that include subgraph analysis, motif detection, temporal anomalies, and deterministic explanations, the system bridges the gap between automated prediction and human decision-making. The addition of similar case retrieval further enhances analytical efficiency by enabling pattern-based reasoning across cases.

Overall, the proposed system demonstrates that effective AML solutions require a combination of realistic data modelling, advanced graph learning techniques, and interpretable outputs. The integration of these components results in a balanced framework that is both technically robust and practically applicable.

7. FUTURE WORK

While the proposed framework demonstrates strong performance on calibrated synthetic data, future work will focus on validating the system using real-world financial transaction datasets to assess its robustness under practical conditions. Incorporating temporal graph learning techniques, such as dynamic or Temporal Graph Neural Networks, can further enhance the model's ability to capture evolving laundering behaviours. Additionally, improving scalability for large-scale financial networks and integrating advanced explainability methods, including attention visualisation and feature attribution, can strengthen the system's applicability in real-world AML operations.

REFERENCES

[1] Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks*. International Conference on Learning Representations (ICLR).

[2] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks*. International Conference on Learning Representations (ICLR).

[3] Weber, M., Domeniconi, G., Chen, J., Weidele, D. K., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). *Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics*. Proceedings of the 25th ACM SIGKDD Conference.

[4] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). *Temporal Graph Networks for Deep Learning on Dynamic Graphs*. ICML Workshop on Graph Representation Learning.

[5] Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). *Enhancing Graph Neural Network-Based Fraud Detectors Against Camouflaged Fraudsters*. Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM).

[6] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based anomaly detection and description: A survey. Data Mining and Knowledge Discovery, 29(3), 626–688.

[7] Ribeiro, L. F. R., Saverese, P. H. P., & Figueiredo, D. R. (2017). struc2vec: Learning node representations from structural identity. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[8] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[9] Ahmed, N. K., Neville, J., Kompella, R., & Kolda, T. G. (2015). Network sampling: From static to streaming graphs. ACM Transactions on Knowledge Discovery from Data (TKDD), 8(2).

[10] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.