Enhancing Illegal Activities in Crypto Currency Using Graph Neural Networks

Mrs. V. Sandya[1], R. Pranathi[1], G. Sathvika[1], D. Revanth Sai[1]
Department of Computer Science and Engineering (Data Science)
CMR Technical Campus, Hyderabad, India
Mails: Sandya.vooradi@gmail.com, reddyvpranathi@gmail.com, sathvikagajelli7204@gmail.com, dagudurevanthsai5317@gmail.com

Abstract- Cryptocurrency money laundering is the equivalent of a serious offense since it assists criminals to conceal their actions, in addition to disrupting markets and the financial sector. Researchers are developing effective Anti-Money Laundering (AML) systems to combat this vice. These systems are beneficial to the society because they mitigate the damage committed by crime. This paper explores the possibility of detecting Bitcoin transactions using the Graph Neural Networks (GNNs). Particular types of GNNs, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), Chebysev convolutional neural networks, and Graph SAGE networks are used in the study. We tested various sets of features after examining the data. We discovered that GNN convolutions with a final linear layer and skip connections are more effective in case of the best results, and with Chebyshev and GATv2 convolutions.

Keywords: Anti-Money Laundering, Deep Learning, Graph Neural Networks, Classification.

I. INTRODUCTION

Money laundering of cryptocurrencies is a massive challenge nowadays. The ease with which criminals conceal movements of money of illegal activity is due to anonymity and decentralization of the exchange. The first step that criminals follow is to convert ordinary money into cryptocurrency. These coins are then transferred between them, through a convoluted line of dealings, and dealings, exchanges, and dealings of this kind. This is aimed at concealing the money trail in order to make it difficult to trace it.

In the past years, individuals have been examining how this issue can be prevented or mitigated. One of the concepts is to establish legal regulations that compel some corporations to carry out money laundering (MLA) and terrorist financing (CFT) checks. The other concept is machine learning with the aim of identifying suspicious transactions.

The initial research provided proof that simple machine-learning algorithms such as Decision Trees and Random Forest can successfully classify transactions reasonably well. These findings indicate that even further advanced methods that take into account the peculiarities of crypto platforms may be even better.

Utilization of Unspent Transaction Output (UTO) is applicable in most crypto platforms. In this system, one or more represented UTXOs in the transaction will serve as the inputs in a transaction, and there will be spending of the coins contained by those UTXOs. The transaction then forms new UTXOs which can be utilized on the later. This circulation pattern of money can be represented as an acyclic and directed graph which is analysable.

Graph Neural Networks (GNNs) are the best next fit since a transaction graph will have relationships between numerous nodes. GNNs are capable of capturing the relational data in a graph-structured data, and are better suited to crypto transaction network analysis.

We experimented with the designs of GNN in our work. Initial, a two-layer graph convolutional baseline. Second, we used a last linear layer in order to sharpen the learnt features. Third, we added a linear layer and skip connections that allowed information to go through indirect routes and directly through to later layers in order to enhance accuracy.

Findings indicate that the Chebyshev convolution and GATv2 variant is better than other methods, as it generates better recall and F1 scores. The experiments affirm that the inclusion of skip connections and last layer provide giant performance improvement.

In general, these results demonstrate the benefit of using GNNs to identify and prevent money laundering of cryptocurrency, and serve as an indication of a more efficient AML tool in the future.

II. LITERATURE SURVEY

A. Early Graph-Based Approaches for Cryptocurrency AML

The release of Elliptic by Weber et al. represented a breakthrough of the ongoing cryptocurrency AML literature with one of the largest scale, labelled transaction graphs available to customers to study. Using Graph Convolutional Networks (GCNs) demonstrated that relationship-based learning can substantially enhance detection of illegal Bitcoin transactions through analysing relationship structure and transaction dynamics which are absent in models based on standard machine-learning. This paper demonstrated that strong patterns were found in transaction networks, such as money flowing along many steps and activity clusters and the process of learning on a graph is most relevant to AML problems. Nevertheless, the study also demonstrated certain issues, including excessive asymmetry between regular and criminal transactions and a lack of some issues with deep models interpretation which complicate their application in the daily compliance practice.

B. Multigraph Modelling and Enhanced Structural Representation.

The relationships among blockchain networks allow money to flow in more than one direction, and along multiple paths, so simple graph neural networks are insufficient to understand the money laundering process as repeated money could enter and exit the networks or money could enter and exit them at the same time. This gap was filled by Ding and others, who developed DIAM which is a system that is aware of these complex network connections. It employs order and direction encoders that display the direction and order of transactions. The special component of the model is the multigraph discrepancy module which aids the model to identify typical monetary laundering trends such as the peel chains, mixers, and loops in the complex graphs. Experiments demonstrate that DIAM provides significantly better results than traditional graph neural networks and other algorithms on multiple sets of cryptocurrency data, which confirms that it is important to consider the specifics of each connection. The results indicate that the anti-money laundering detection becomes significantly enhanced when the graph structure is explicitly designed with the consideration of the numerous paths the funds can take in a blockchain.

C. Self-Supervised and Contrastive Learning for Label-Scarce AML Data.

The fact that the number of labelled illicit transactions is low and the data of block chains is anonymous led Lu and colleagues to develop GCPAL. GCPAL is a contrastive graph pretraining algorithm which trains powerful embeddings without requiring a large number of labelled examples [5]. It forms augmented displays of the graph, and it employs contrastive objectives, which enhances the node representation and leaves downstream classifiers to perform properly even in the case of low label abundance. This is extremely practical in anti-money laundering (AML) where fraud is uncommon and their confirmation in the case needs a professional competence. On the same note, it was demonstrated through Inspection-L by Lo and his peers that self-directed learning is effective. They concluded that supervised GNN embeddings only combined with the conventional machine-learning models outperform only supervised deep-learning pipelines only when working in AML settings where the data is imbalanced [6]. Collectively, these investigations demonstrate that self-supervision along with contrastive learning is required to address the real-life blockchain issues like unclean metadata, markedly scarce labelled samples, and so on.

D. Hybrid GNN–Machine Learning Pipelines.

These fusion methods are combinations of GNN embedding with normal machine-learning classifiers. Lo et al. demonstrated that such models as Random Forest and XGBoost could utilize the Inspection in the form of embeddings in order to classify AML instances with high accuracy and reasons provided. They performed better than pure GNN models, particularly in case of imbalanced data or in case of a changing pattern of transactions. The DIAM model by Ding and colleagues can also be applied in hybrids and produce rich graph representations that can guide classifiers to make reliable and explainable predictions. The hybrids provide the AML experts with graph insights in addition to clear explanations that are required by regulators in the process of audits and reports. They are quite well fitting in any day-to-day AML processes, which demand explainability.

E. Theoretical Advances in GNNs for Directed Multigraphs.

The important theoretical concepts provided by Egressy et al. included the development of strong GNNs which operate on directed multigraphs, which are similar to the way blockchain transaction networks are organized. They have introduced new features like port numbering, reverse message passing, and ego-ID encodings that have enabled GNNs and made them quite more competent. The models use these changes to identify the various subgraphs and detect complex laundering patterns in spots [9]. These outcomes provide the basis to develop AML models that are capable of identifying complex, multi-edge laundering behaviors that other GNNs do not find easily [10].

III. PROPOSED METHODOLOGY

A. Transaction Graph Representation.

The process works on the creation of a transaction graph using the Elliptic data. Each transaction is a node, each money transfer between two individuals or firms is a direction of one node to the other.

This demonstration indicates that the fraudulent operation is not only a single occurrence but a trend in numerous transactions interconnected to each other.

With the help of a graph, the system will be able to examine both the immediate connections of the relationships (who sent to whom) and more extended connections that are capable of presenting money laundering.

This graph model is more realistic in terms of money flowing within the blockchain to allow the classifier to learn small scale and macro-scale flow.

With this framework, the system has a better opportunity to identify the small and dynamic illegal transactions that would otherwise be difficult to locate had the data were nothing more than just a simple table.

B. Integration of Multiple GNN Architectures.

The system incorporates four designs of Graph Neural Network to extract helpful data regarding the transaction graph: GCN, GAT, Chebyshev, and GraphSAGE. Each design is determined to be able to learn various portions of the graph. GCN makes spectral intensities less irregular and provides information diffusion in a short timeframe whereas GAT incorporates attention methods that enable the model to prefer significant neighbours. Chebyshev makes use of spectral filters, which selects many hop signals without a lot of computation. GraphSAGE allows us to learn some nodes that we have never seen. We supply skip connections such that the features of the early layers remain rocked in the model and do not experience the recurring issue of deeper layers rendering node embeddings all too similar. An extra linear layer follows the encoder of the GNN as well to complete the fine-tuning of the embeddings and prepare them to be more useful in classification. The alterations enable the learning framework to be more robust and sturdy, allowing the GNNs to be able to identify multifaceted money-laundering patterns of numerous transaction paths.

C. Feature Refinement and Reduction.

To ensure the proposed AML system works well, feature engineering would be appreciated. We examine mixed characteristics of the transaction graph and see whether they are too close. In case two features are correlated to more than 0.9 we delete one of them. This eliminates redundant features in order to retain the best ones to train. Removing the correlated features also results in a faster and less overfitting model, thus better on other set of patterns in transactions. This enhancement directly will increase the performative stability of the model when it operates within dynamic blockchain environments.

D. Performance Benefits and System Robustness.

The technique performs well in detection of illegal transactions and the scores are very high up to 0.906 better than the older machine-learning tools which are not assisted by hand-made features.

It detects more illegal transactions and is less liberal with suspicious behavior showing it is more cautious with complex networks requiring multiple steps to decode.

In the graph neural networks, the model is more reliable in that adding skip connections and final linear layer stabilizes the internal representations of the model as the network scale varies.

The shortening of features also makes the system faster and allows them to process high transaction networks without being inaccurate.

With all these upgrades, the system is better equipped to protect it against new laundering techniques that can enable it to continue functioning even when the criminals employ highly advanced patterns.

The approach is a total scale-able and high personage solution to anti-money-laundering.

E. System Architecture

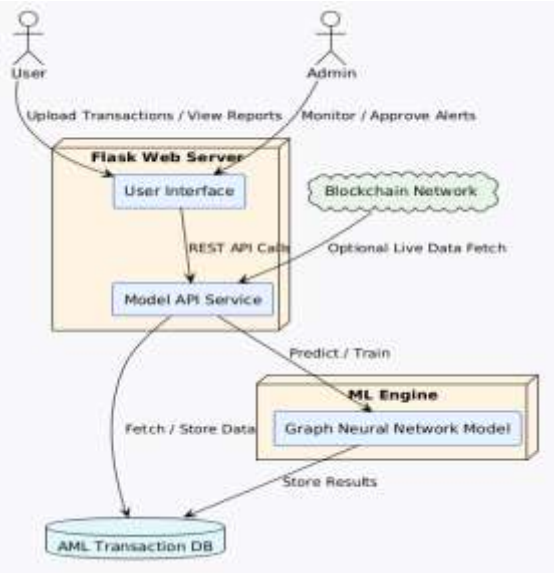

Fig 1: System Architecture

The system is designed in a modularized pipeline that connects the users, administrative modules, machine-learning services, and the transaction database to it in such a way that it can identify money-laundering in a short period of time.

The interface is Web based and is created in Flask. They will be able to post transaction information and view the reports that the system produces. The interface communicates with the backend through calls of REST API. Such calls are requests to the Model API Service where the input is processed, predictions made, and the data is transferred between components.

A blockchain network is also able to be linked to the architecture. This allows the system to fetch the recent transaction information on demand. The Model API Service is the service communicating directly with the ML Engine, which executes the Graph Neural Network model which trains and predicts illicit transactions.

IV. RESULT AND DISCUSSION

A. User Login and Authentication.

The module of login verifies a username and a password and redirects the users to their own dashboard after logging in. A photograph depicts the fact that actual users take a short amount of time to access the AML, and the error reports on incorrect passwords are precise. This is to ensure that the login process is not broken and that only authorized individuals are granted access to the AML dashboard. This successful log-in process also demonstrates that the back-end system is compatible with the login screen which ensures the security of the permissions and overall system of every user.

B. Admin Dashboard Access and Alert Monitoring.

The admin module will provide an overview of system performance, flagged transactions and on user requests. The screenshot displays the dashboard of the admin when you sign in is able to prove that the links and the login is functioning well. Admin staff members will be able to give new users, review warnings regarding suspicious activities and examine real-time performance measures. This is demonstrating that the system can assist in monitoring by the admins and allows the AML analysts to have a close observance of high-risk activities effectively.

C. Transaction Dataset Upload and Preprocessing.

The module of upload and pre-processing ensures that only correct and formatted data will be inputted into the AML. The screenshot demonstrates that the data is uploaded successfully and that the unsupported format is rejected, which is evidence that the file checks perform positively. Once it has been uploaded, it is pre-processed (some crucial tasks such as data normalization, feature cleaning, missing value processing, and data separation into training, validation, and testing sets are performed at this stage). TC-06 through TC-09 test cases verify every pre-test and ensure that the module has always generated clean and systematized data that can be used to construct the graph. The degree of this reliability is necessary since the correctness of the GNN training heavily depends on the quality and consistency of the input data.

D. Graph Construction and Visualization.

The data of transaction is converted into directed graph using the graph construction module. Wallets or a transaction are the nodes and the money flow is displayed in the edges.

It is a definite picture of transaction clusters which is demonstrated in the screenshot that the relationships and structure were represented accurately by the system.

With proper construction of nodes, edges and khop subgraphs (tested with TC-10 to TC-13), the system will present the analysts with an overview of layout of the network in detail.

Forensic investigation requires these visualizations. They assist investigators in identifying abnormal money flows, closely knit organizations and wallets that made use of possibly suspicious loops.

The exception as well facilitates the model better in visual form, allowing analysts to verify their findings by looking right into the chart and confirming that something looks odd.

E. GNN-Based Prediction and Classification

According to the results depicted in the screenshot, the hybrid GNN model such as the hybrid GNN aml xgb model can properly identify the Suspicious or Legitimate transactions besides providing a confidence score. The system relies on the information of new GNN models such as GCN, GraphSAGE, and Chebyshev and GATv2, which is a major advantage compared to the previous models in ML and demonstrates the importance of graph data. The prediction results are also presented clearly in the screenshot providing the investigators with a quick and useful information.

F. AML Alert Generation.

The alert generation module is the automatic search of suspicious transactions and displays on the user or admin dashboard. The screenshot displays an alert, which has the transaction ID, the risk score, and classification. This system ensures that some potential illegitimate act is displayed immediately, and this way, investigators are able to perform fast. Test cases TC-26 through TC-28 affirm that alert is being generated, displayed and filtered. Making sure that the system does not raise alerts concerning regular transactions is also done to ensure that the work of analysts is accurate and not overloaded. Combining alert logs with graph displays allows the investigators to glimpse the connection between the flagged activities easily.

Discussion

The outcomes indicate that our GNN-based AML system is significant in terms of identifying illegal transactions on cryptocurrencies.

We convert the data to a graph of transactions and apply such models as GCN, GAT, Chebyshev, and GraphSAGE. Such models manage to capture relationships and long-term trends which are not normally available to the methods of normal ML.

We also included terms such as skip connections, a final linear layer and reduction of features, and this made the model more stable and representing information.

To make it more user-friendly, we offer exportable reports and a proper dashboard.

In general, this method provides a good and realistic means of identifying AML within cryptocurrency systems.

V. CONCLUSION

This research is a novel method of identifying money laundering in cryptocurrency. It also applies Graph Neural Networks (GNNs) to learn on the basis of transaction data and an aml_xgb tool that makes the final decisions. The system is able to infer patterns that other models fail to notice by constructing a graph of transactions by utilizing blockchain records. It consumes useful node embeddings of state-of-the-art types of GNNs (Chebyshev, GATv2, GraphSAGE, and GCN) and provides hand-crafted features. This method enhances accuracy, recall and F1 -score of the system in detection of illegal transactions. Integrating XGBoost (aml_xgb) increases the performance, particularly in the situations when the ratio of normal cases to the illegal cases in data is higher, and provide clear explanations. Its structure is developed in a modular manner, encompassing data cleaning, graph building, embedding generation, classification, warning messages, and dashboards, hence can be expanded successfully, be easily utilized, and efficient to operate. In general, the model can enhance AML operations within the digital finance sector and assist investigators and regulators to identify suspicious activity in real-time.

VI. FUTURE SCOPE

The tools of money-laundering in future are going to be completely powered by AI and they will operate in real time and will collaborate with each other to track sophisticated crime.

The systems will allow AI models, in particular, Graph Neural Networks and large anomaly detectors, to learn. They will direct automatically to new ways of laundering rather than just adhering to constant rules.

It will be necessary to see transactions as a graph in real time. The system will identify weird behaviour within milliseconds among banks, fintech, and crypto.

Due to the current shift of laundering across numerous blockchains, novel systems will incorporate the cross-chain analysis, pool wallets and monitor DeFi operations.

The global regulators will need more powerful data sharing and report consistency. This will aid the banks and other institutions to identify coordinated laundering networks better.

REFERENCES

[1] G. Pavlidis, “The birth of the new anti-money laundering authority: Harnessing the power of EU-wide supervision,” *J. Financial Crime*, vol. 31, no. 2, pp. 322–330, Mar. 2024.
[2] S. Marasi and S. Ferretti, “Anti-money laundering in cryptocurrencies through graph neural networks: A comparative study,” in *Proc. IEEE 21st Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2024, pp. 272–277.
[3] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, “Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics,” *Electron. Markets*, vol. 33, no. 1, p. 37, Dec. 2023.
[4] W. W. Lo, G. K. Kulatilleke, M. Sarhan, S. Layeghy, and M. Portmann, “Inspection-L: Self-supervised GNN node embeddings for money laundering detection in Bitcoin,” *Appl. Intell.*, vol. 53, no. 16, pp. 19406–19417, Aug. 2023.
[5] P. Gerbrands, B. Unger, M. Getzner, and J. Ferwerda, “The effect of anti-money laundering policies: An empirical network analysis,” *EPJ Data Sci.*, vol. 11, no. 1, p. 15, Dec. 2022.
[6] M. M. Rathore, S. Chaurasia, and D. Shukla, “Mixers detection in Bitcoin network: A step towards detecting money laundering in crypto-currencies,” in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 5775–5782.
[7]] J. Wu, J. Liu, Y. Zhao, and Z. Zheng, “Analysis of cryptocurrency transactions from a network perspective: An overview,” *J. Netw. Comput. Appl.*, vol. 190, Art. no. 103139, Sep. 2021.
[8] V. Chang et al., “How blockchain can impact financial services—The overview, challenges and recommendations from expert interviewees,” *Technol. Forecast. Soc. Change*, vol. 158, Art. no. 120166, Sep. 2020.
[9] A. Pareja et al., “EvolveGCN: Evolving graph convolutional networks for dynamic graphs,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, pp. 5363–5370, Apr. 2020.
[10] M. Jullum, A. Løland, R. B. Huseby, G. Ånonsen, and J. Lorentzen, “Detecting money laundering transactions with machine learning,” *J. Money Laundering Control*, vol. 23, no. 1, pp. 173–186, Jan. 2020.
[11] M. Weber et al., “Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics,” in *Proc. KDD Workshop Anomaly Detection Finance*, 2019, pp. 1–7.