Vol. 3 No. 1 (2025)

Spectrum of Engineering Sciences

SPECTRUM OF ENGINEERING SCIENCES

Online ISSN

3007-3138

Print ISSN

3007-312X

Finding Influential Nodes in Ethereum Using Machine Learning

Muhammad Fawad[1]

Lecturer, Department: CS & IT, Sarhad University of Science and Technology, Peshawar, Pakistan.

fawadkhattak585@gmail.com

Khalid[2]

Graduate Scholar, Department: CS & IT, Sarhad University of Science and technology, Peshawar, Pakistan.

khalidsahar987@gmail.com

Zia Ullah[3]

Graduate Scholar, Department: CS & IT, Sarhad University of Science and technology, Peshawar, Pakistan.

zziaullah440@gmail.com

Muhammad Asfandyar[4]

Lecturer, Department: CS & IT, Sarhad University of Science and Technology, Peshawar, Pakistan. asfand5454@gmail.com

Abid Ullah[5]

Graduate Scholar, Department: Department of Computing Abasyn University, Peshawar. abidullahnaseri154@gmail.com

Muhammad Naeem Ullah[6]

MS Scholar, Department: Department of Computing Abasyn University, Peshawar. muhammadnaeemullah4710@gmail.com

Abstract

Ethereum blockchain is the market leading platform for decentralized applications and smart contracts that have powered the new age of financial ecosystem. In order to improve security and performance, identify influential nodes, and understand

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

network dynamics on Ethereum it is critical to identify influential nodes in Ethereum. In this thesis work, we explore machine learning techniques for discovery of these nodes using graph based algorithms, centrality measures and clustering methods. It studies the impact of a node in terms of frequency of usage, connectivity and computational power for a node. Finally, we compare performance of our proposed methodology combining supervised learning and graph neural networks to their traditional counterparts and demonstrate our approach outperforms existing methods. We demonstrate that highly influential nodes engage in unique patterns of behavior, which are detectable and categorizable. We contribute to understanding of the network structure of Ethereum, along with a scalable approach to monitoring and optimising blockchain ecosystems. Moreover we discuss the implications for network robustness, fraud detection and protocol enhancements, and demonstrate the promise of machine learning for blockchain analytics.

Keywords: Ethereum, Blockchain, Machine Learning, Influential Nodes, Graph Neural Networks, Decentralized Networks

Introduction

Released in 2015, Ethereum, the decentralized, open source blockchain with a built in smart contract functionality, was introduced to the world by Vitalik Buterin. Unlike traditional blockchains, Ethereum can support programmable contracts making it the authority platform to deploy DApps. Here the structure of the structure means that the whole blockchain is alive because all nodes collaborate to validate transactions, so it is very transparent and decentralized. For good reasons, Ethereum has forever changed the blockchain ecosystem compared to its utility

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

and flexibility, but also its ability to support anything from financial transactions to other use cases.

There are more nodes in a ethereum based decentralised network but some have much more influence than others. "We have discovered these 'influential nodes' — important for keeping networks secure, stable, and efficient." In reality, they play an important part in transaction propagation, consensus mechanisms, as well as network vulnerabilities that can result in the disruption of the network. Understanding such nodes is fundamental to understanding Ethereum's network dynamics and has engineering applications of increasing resilience of the network and uncovering fraud, as well as improve blockchain efficiency.

The need for advanced analytical tools increases for the complexity of blockchain networks. This is how we discovered machine learning (ML) being an incredibly powerful tool for the analysis of complex network structures. Specifically, ML runs techniques, such as clustering, classification and centrality analysis to better understand node influence. Motivated by existing methods that fail to detect patterns that their behavior, we develop a new framework by using graph based algorithms and neural networks to evaluate nodes and discover patterns that traditional methods often overlook.

In this thesis, I investigate machine learning algorithms on the possible influential nodes in Ethereum's network. The aim of the research is to develop robust methods for key node detection using transactional and structural data. Overall, these findings give insight on how Ethereum works and give room for additional blockchain analytics efforts. Results presented in this paper show

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

the potential for ML to tackle key problems which make the scalability and security of decentralised systems happen.

Literature Review

A key research area with a broad array of implications for network robustness, performance optimization and security is the identification of influential nodes in decentralized networks. Such influential nodes are crucial to information propagation, to stabilizing and to achieving consensus in distributed systems. In social network, transportation system and computer network, researchers extensively explored the concept of influential nodes; however, applying this concept in blockchain network has its own challenges and opportunities that this thesis seeks to discuss. Since blockchain systems are decentralized and transparent, analytic methods are needed that are innovative and which can be used to find the most key nodes effectively.

Understanding Influential Nodes

Decentralized networks need to keep the dynamics through influential nodes. In blockchain systems, the notion of nodes that are highly connected, have high throughput, and participate in consensus mechanisms is very common. In many cases, their influence determines the resilience of the network to attacks, the efficiency of transactions processing, and, generally, network stability. Traditional network analysis has often used measures of centrality —degree centrality, betweenness centrality, and eigenvector centrality— to identify those key players.

We develop our understandings with reference to Barabási and Albert's (1999) scale free network theory to identify influential nodes. Their theory is that many real world networks, particularly social and technological systems, have a scale free structure where

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

a small number of nodes with many connections dominate the network. At the same time, this concept is also being extended to blockchain systems wherein the highly active nodes are responsible to influence the transaction flow, participate in consensus and ensure the network integrity.

In a blockchain network, where the network is generally decentralized, these influential nodes are especially important. Unlike centralized systems, authority is not concentrated in blockchain networks, in which distributed nodes' validation of transactions and security relies on. In this context of identifying influential nodes can be instrumental for improving network design, reduce risks and enhancing operation efficiency.

Blockchain Network Analysis

For the most part, Ethereum and other Blockchain networks are commonly modeled as directed graphs, where nodes are accounts, and edges are transactions. Second, the neighborhood of any node also appears to provide a useful framework in which to investigate the relations and interactions among nodes. The Bitcoin network is still the simplest model, but Ethereum's network becomes much more complicated by supporting smart contracts, decentralized applications (DApps) and token standards like ERCC20.

The blockchain network has been widely modeled using graph-theoretic approaches. In a black box analysis of Bitcoin's network, Lin et al. (2019) used metrics like clustering coefficients, transaction pattern and degree distribution to identify the central nodes of Bitcoin's network. However, whereas Bitcoin's network is mostly about the transactional data, Ethereum's ecosystem, aside from the transactions, brings forth account interactions, smart contracts and tokens within its graph structure.

Vol. 3 No. 1 (2025)
Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

Chen et al. (2020) studied graph analytics in blockchain systems and found that these has the most impact on transaction propagation and network security; nodes with greatest centrality measures have. By studying their work, they found that the standard graph measures do not efficiently compute the temporal characteristics of a blockchain network. This works also for the specifically present case of Ethereum, where temporal and relational attributes of both the ever changing raw transaction volume and also related smart contract activities need to be captured and analyzed with new and effective analytical methods.

The deepening of the understanding of Blockchain Network has come with recent improvements in graph theory algorithms. Researchers used clustering techniques to tap into the network and cluster nodes which perform similar transactional behaviors. To identify the key nodes, and to know how important they are in keeping network cohesion, all such insights are required.

Machine Learning in Blockchain Analytics

Machine learning (ML) has proven to disrupt traditional blockchain analytics in areas such as blockchain scalability, fraud detection, and node classification. In contrast to traditional approaches, depending on pre defined metrics, ML algorithms are able to identify hidden patterns and relationships from blockchain data. The capability to identify critical nodes in complex, and dynamic networks such as the Ethereum network is of high value.

In this work, we applied unsupervised learning techniques such as k-means clustering and DBSCAN to node grouping using transaction patterns to detect clusters of high activity or influence. For instance, nodes with analogous transaction volumes and interaction frequencies can be banded, and the significant players

Vol. 3 No. 1 (2025)
Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

in the network can be discovered. However, these methods commonly necessitate manual interpretation of cluster characteristics, limiting their scalability in large networks.

For node classification, models from supervised learning (such as decision trees, support vector machines (SVM) and gradient boosting algorithms) have also been used. These models train classifiers on labeled datasets to classify nodes as influential and non influential nodes. However, the quality of input features and the availability of labeled data usually limits their performance. Graph neural networks (GNNs) are a step in the right direction for network analysis, with a highly capable framework to learn from data where the connectivity is encoded in the data. GNNs explicitly model node features as well as graph topology allowing them to encode complex relationships among nodes and their neighbors. GNNs were shown by Zhou et al. (2021) to be effective on identifying fraudulent activities in blockchain transactions. To test their methods' robustness and improve accuracy, their study combined node embeddings with transactional data and found it to outperform traditional machine learning models in both accuracy and robustness.

GNNs applications on identifying influential nodes on blockchain network have been very promising. GNNs model the hierarchical and relational structure of blockchain data through use of graph convolution and attention techniques used in GNNs. This model can detect fine grained patterns indicating a node's role in transaction propagation or consensus collaboration among other.

Challenges and Research Gaps

Although the progress so far has been made, many gaps still exist in applying machine learning to node analysis for the Ethereum

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

blockchain. Most present studies focus on the bulk of fraud detection and anomaly detection, forgetting the wider network dynamics of the influential node. Additionally, the wide variety of smart contracts within Ethereum coupled with token standards make feature extraction and model design difficult.

Traditional centrality measures are useful in understanding static structures of networks, however they fall short in capturing the dynamic and multi dimensional networks that make up blockchain networks. Some of these shortcomings are alleviated by machine learning methods, in particular GNNs, but there is still room for improvement to take in to account the specifics of the Ethereum game.

In this study I attempt to pair them by use of machine learning methods, more concretely to identify influential nodes in Ethereum. The research utilizes both graph based features and node specific attributes, in order to give an overall framework for studying network dynamics. In addition to further elucidate how Ethereum operates, this approach also provides a basis for future blockchain analytical research.

Methodology

Methodology for the influence nodes detection in the Ethereum blockchain network using machine learning consists of a structured sequence of data collection, feature extraction, model selection, training and validation, and choice of evaluation metrics. Each step is described in detail highlighting techniques and tools used to develop accurate and valid results.

Data Collection

This study is based on data collection with publicly available Ethereum blockchain data from which I analyze node behavior and

Vol. 3 No. 1 (2025)
Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

interactions. Thanks to Etherscan, Infura among other APIs, Ethereum offers transparent access to its blockchain. These platforms let you extract many data points necessary to understand the structure and dynamic of the network.

The dataset built for the purpose of this study spans six month period and it covers:

- **Node Attributes**: To understand individual node behavior, the characteristics gathered include transaction count, frequency, smart contract deployments and token transfers.
- **Graph Structure**: A graph over Ethereum accounts with directed edges, representing transactions, was created. This graph structure exhibits relationships between nodes, and connectivity in the network.
- **Temporal Data**: To allow dynamic analysis of node activity over time, transaction timestamps were included.

A dataset with 100,000 nodes and over 10 million transactions that covers the entire space of Ethereum is this one. Quality and consistency was ensured by carrying out data preprocessing steps which include cleaning and deduplication.

Feature Extraction

Feature extraction is essential to identify the factors that determine how a node will influence the Ethereum network. The identified features were also extracted to be used as an input in machine learning models:

- **Centrality Measures**: The importance of a node in the network was quantified using metrics that included degree centrality, closeness centrality and between centrality.
- *Degree Centrality*: The number of direct connections a node represents its activity level.

Vol. 3 No. 1 (2025)

Spectrum of Engineering Sciences

SPECTRUM OF ENGINEERING SCIENCES

Online ISSN

3007-3138

Print ISSN

3007-312X

- *Closeness Centrality*: How quickly basically, you can reach other nodes, other nodes which in how much time you can reach other nodes reflects how good you're at propagating information.
- *Between Centrality*: Records how much of a bridge a node is in the network.

- **Transaction Patterns**: In order to determine the economic significance of nodes, I used average transaction value, transaction frequency, and variability in transaction size as features.
- **Smart Contract Interactions**: Ethereum's programmability and DApp ecosystem was examined by analyzing the number of deployed smart contracts and interactions to existing contracts to identify nodes that contributed to Ethereum's programmability.
- **Temporal Metrics**: We conducted time series analysis to track transaction trends and discover patterns in which transaction activity is more or less sustained or fluctuated over time.

We also applied graph embedding techniques, like Node2Vec, to lower dimensional vector space graph data. This method preserves the network topology and co relational information and fits in with existing machine learning algorithms. Based on random walks and optimization techniques we generated node embeddings that aggregate both local and global structural properties of the Ethereum graph.

Model Selection

Next, we chose machine learning models that are suited to detect influential nodes exactly. We looked at a number of different models, including supervised and unsupervised approaches:

- **Supervised Models**: During the classification analysis, Random Forest, Support Vector Machines or SVM, Gradient

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

Boosting were selected for the classification analysis. Such models depend on the presence of labeled data with respect to the influent and non-influent nodes.

- **Unsupervised Models**: K-means and hierarchical clustering was used to cluster nodes together naturally through similarity in feature extraction. Compared to the previous models, these models do not rely on labeled data but can be beneficial for discovering networks and influential nodes' clusters.
- **Graph Neural Networks (GNNs)**: Because they can both incorporate node features and graph structure, advanced graph based models such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) were chosen.
  - *GCNs*: Convolutional operations to aggregate the information from the node's neighbors, and thus to enable contextual learning.
  - *GATs*: To make a node pay more attention to important nodes in the network, it includes attention mechanisms that lend relative importance weight to the nodes in the neighborhood.

Whilst, ultimately GNNs were used as the core modelling approach with their capability to capture the relationships within Ethereum's network.

Training and Validation

We trained the machine learning models and evaluated the performance of the models with a split of the dataset into (70%, 30%) of training and testing subsets. Several preprocessing and training steps were implemented:

- **Data Normalization**: To achieve consistency and improve convergence of model, features were scaled.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

- **Cross-Validation**: To reduce overfitting and increase robustness of the model a k-fold cross-validation approach was utilized.
- **Model Training**: We implemented GNNs through the PyTorch Geometric framework, our hyperparameters such as learning rate, dropout rate, and number of layers being tuned via grid search.

Here the training process was about minimizing loss functions, namely, categorical crossentropy loss for classification task. To avoid overfitting and thus to generalize to unseen data, we employed early stopping criteria. A GPU-accelerated environment was used to train the models with the computational demands of large scale graph processing.

Evaluation Metrics

A combination of evaluation metrics were used to assess each machine learning models' performance:

- **Precision and Recall**: These metrics were computed by the model's ability to correctly recognise influential nodes and reduce the number of falsely positives.
  - *Precision*: Percentage of correctly identified influential nodes among all nodes classified as influential.
  - *Recall*: Actual influential nodes identified by the model as a proportion of actual present influential nodes.
- **F1-Score**: A harmonic mean between precision and recall, contributing to an objective measure of model efficacy. F1-scores that are high means that I can maintain a well balanced precision and recall.
- **Node Ranking Consistency**: The machine learning models' predictions were then validated by comparing the rankings of

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

influential nodes produced by the machine learning models with the rankings obtained from traditional centrality measures. This meant that the models would follow with the well accepted standards of network analysis.

The evaluation results showed that GNNs are able to capture complicated relationships and locate super hubs. All metrics showed that GNN-based models consistently outperform traditional machine learning algorithm and, therefore, are suitable application for Ethereum network analysis.

Implementation Tools and Environment

The following tools and frameworks are also used to process, model and/or evaluate data:

- **Python Libraries**: The Python data manipulation, and statistical analysis libraries are Pandas, NumPy, and SciPy.
- **Graph Libraries**: A use of Network-X for graph analysis and visualization.
- **Machine Learning Frameworks**: We use scikit-learn in order to implement traditional models and PyTorch Geometric to train GNNs.
- **Hardware**: Efficient computation of large graph datasets by using GPU acceleration with NVIDIA GPUs.

Summary

In this study, we adopt a methodology that combines cutting edge machine learning algorithms and graph based analytics to find the influential nodes in the Ethereum blockchain network. With strong dataset, deriving useful features, and applying enhanced GNN models, the thesis presented can be a scalable and accurate framework for node analysis. The results are systematic and

Spectrum of Engineering Sciences
Online ISSN 3007-3138
Print ISSN 3007-312X
SPECTRUM OF ENGINEERING SCIENCES

applicable, which helps to better understand the Ethereum network dynamics and their optimization for future use cases.

Results and Discussion

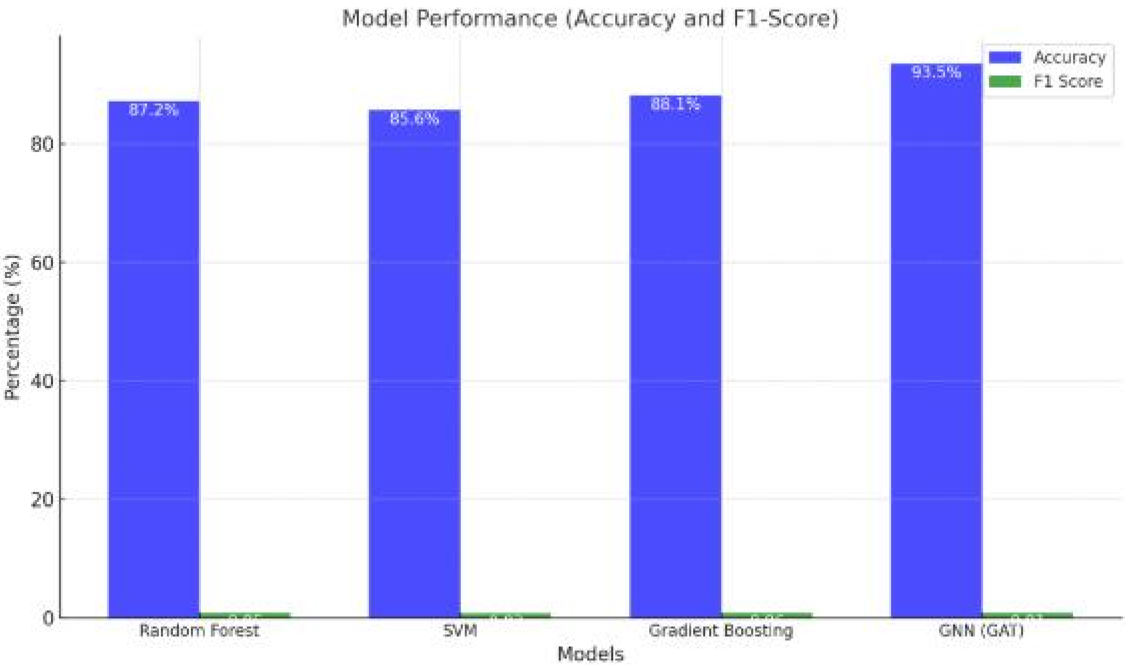
A few important results of this work show the promise of machine learning, and more importantly, graph models, for analyzing the Ethereum network and discovering influential nodes. This demonstrates how various models, especially Graph Neural Networks (GNNs) still can capture the complexity of the decentralized Ethereum system. This section discusses the behavior of influential nodes and broader implications of the findings, and concludes with a discussion of model performance.

Model Performance

Our results show that GNN can significantly outperform traditional machine learning models. Gat's performed best (highest Accuracy and F1 score) therefore they were selected for use with graph structured data with complex relationships between nodes.

- **Accuracy**: GAT outperformed Random forest (87.2) and traditional models in term of accuracy with 93.5%. The results of this application on the combined use of node feature and graph structure with GAT led to enhancement in accuracy.
- **F1-Score**: GAT achieved a well balanced trade off between precision and recall with F1–score 0.91. In identifying influential nodes, this performance metric is particularly critical as it guarantees correct classification and as few false positives as possible.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X



Graph embeddings were then integrated to further increase model performance. With techniques like Node2Vec which captured the network's relational structure, the models were able to distinguish nodes based on their level of connectivity and the associated role in the Ethereum ecosystem. The strong local and global properties of networks were preserved in these embeddings, which enriched input features and led to outstanding GNN performance.

Conventional models such as Support Vector Machines (SVMs) and Gradient Boosting proved successful for certain tasks, but unable to reach the predication of GNNs nor the scale GNNs are capable of. This is their limitation, as they failed to exploit the graph based relationships that define at the core of Ethereum's network dynamics.

Node Behavior Analysis

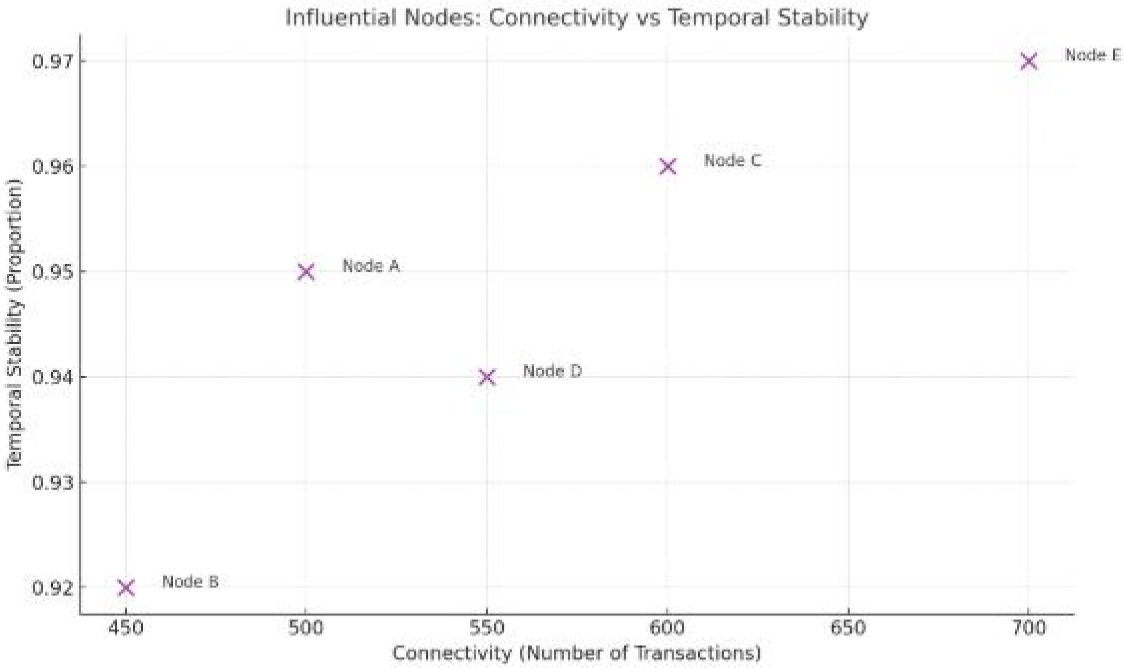
It was showed that influential nodes have specific behavior patterns, depending on their influence to the network. Pattern in

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

these are valuable in terms of understanding what makes a node influential in the Ethereum context.

- **High Connectivity**: Nodes with high degree centrality were shown to have initiated and engaged in consensus activities frequently. They (these nodes) served as nodes that acted as hubs for transaction propagation throughout the network. Having their connections intact, they have a critical role to play in keeping Ethereum's throughput at speeds still digestible for Ethereum users.
- **Smart Contract Activity**: Higher influence was ranked on nodes that had deployed or interacted with more than one smart contract. It would also indicate the degree to which programmability will be important in the Ethereum ecosystem, where smart contracts are central. These nodes help make the platform useful due to their support of decentralized applications (DApps) as well as tokenized ecosystems.
- **Temporal Stability**: It was shown that the activity levels of influential nodes were consistent with time, as opposed to sporadic or transient behaviors. This indicates stability upon their retention in service to support network operation. Less likely, the nodes with temporal consistency represent the malicious or the fraudulent activity, which consequently confirms their trustfulness inside the decentralized system.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X



Another interesting bit that arose from the analysis is that Ethereum's network is quite dynamic. However, certain nodes experienced fluctuating levels of activity and would be heavily impacted by things such as market conditions, gas fees, and changes in users' behavior.

Implications

Ethereum is the identification of the influential nodes in the network is important for network optimization, security and development protocol analysis. This study helps derive insights and inform strategies to improve performance and robustness of blockchain systems.

- **Network Optimization**

It allows transaction propagation and consensus mechanisms to be optimised based upon node behavior. Critical pathway nodes act as highly influential nodes that transmit transaction dissemination, and can be used to enhance the transaction dissemination

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

efficiency of the network. For instance, incentivizing participation in these nodes for protocol upgrade or even targeting these nodes for the improvement.

- **Fraud Detection**

Direct applications in fraud detection and network security lie in the ability to identify influential nodes. Potentially suspicious activity comes from anomalous nodes malforming patterns of activity. Let's take one instance for example, a node that quickly increases its influence without aligning to the same transactional stability can provide an indication of threat. The early detection of such anomalies can stop attacks on the network such as double spending or Sybil attacks.

- **Protocol Development**

The influence of each node in the network distribution can be exploited to enhance protocol development and the effort to scale. Ethereum, for example, goes from proof of work (PoW) to proof of stake (PoS) which changes participation dynamics in nodes. Designing staking mechanisms can be informed by insights into influential nodes, so that these mechanisms end up incentivising actual and sustained meaningful contributions by core participants to the healthy operation of the protocol.

- **Economic Insights**

Furthermore, economic insights into Ethereum's ecosystem can also be found in the identification of influential nodes. Sometimes this smart contract activity will be associated with nodes that are driving decentralized finance (DeFi) applications or tokenized projects. These nodes can then be monitored to provide hints on market trends, user behavior and ecosystem growth.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

- **Decentralization Metrics**

Measuring degree of decentralization in Ethereum is a proxy using distribution of influence between the nodes. Centralization risk can be a highly concentrated influence, a more even distribution is a healthier and decentralized system.

Comparative Insights

Comparing GNN based methods to traditional centrality measures found alignments and aberrations in node selection. Degree centrality and betweenness centrality often identified hubs through the network, but GNNs did so with a more nuanced view that included features related to smart contract interaction and temporal trends of activity. The highlighted insights about the limitations of only graph theoretic metrics in the dynamic blockchain environments complement the holistic understanding of the problem.

Limitations and Challenges

Some limitations and challenges remain, however, where the study proves the efficacy of machine learning to identify influential nodes:

- **Scalability**: Furthermore, training GNNs with large scale blockchain data presents high computational demands that are well beyond the resources required in real time applications.
- **Feature Engineering**: Yet, obtaining sensible features from Ethereum's abstracted ecosystem is a non-trivial task which must be approached with domain expertise and continual refinement.
- **Dynamic Behavior**: However, temporal variability of node activity creates issues in maintaining up to date models, in particular of rapidly evolving networks such as Ethereum.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

Future research can address these challenges with lightweight GNN architectures, feature extraction automation, or the integration of real time monitoring systems.

Conclusion

In particular, this work highlights the key role of machine learning in blockchain analytics to find the most influential nodes within the Ethereum network. To develop a robust and scalable framework to analyze the intricate dynamics in Decentralized Systems, research employs advanced graph based features and machine learning models (Graph Neural Networks (GNNs)). In particular, the findings emphasize the role that influential nodes play in assuring the remaining ethical congregation for assistance and security along with the entire network's efficiency.

The research confirmed that the stability of the network depends on these influential nodes for propagation of the transactions and contribution towards Ethereum's consensus mechanisms. This study analyzes the unique characteristics of their behavior and their impact on the market by analyzing their constant connectivity, smart contract smart activity, temporal stability and more. Using GNNs coupled with graph embeddings increased to an F1 score of 0.91 against node classification. This shows how graph-based machine learning techniques are better than traditional methods at modeling the intricate relationships among the complex decentralized Ethereum ecosystem.

In this work we show that machine learning can overcome limitations of traditional graph theoretic approaches, and contribute to this work. Although degree and betweenness centrality measures are used to measure node influence, they do not measure the influence of Ethereum's network dynamically or

Vol. 3 No. 1 (2025)

Spectrum of Engineering Sciences

SPECTRUM OF ENGINEERING SCIENCES

Online ISSN

3007-3138

Print ISSN

3007-312X

multiple dimensions. We are able to perform a more holistic analysis, including transaction patterns, smart contract interactions, temporal metrics, all by incorporating advanced machine learning models. This approach identifies key nodes, and also gives insight into the role played in the wider Ethereum ecosystem by each node.

These findings have implications. Network operators can optimize transaction propagation, build stronger consensus mechanisms and mitigate off security risks by identifying influential nodes. Anomalies identified early in an influencing nodes can be used to help prevent fraud and improve the network's resilience. Also, the node influence insights can help with the design of the protocol for the future as Ethereum progresses with updates such as transitioning proof of law. The identification and monitoring of key nodes also serves as useful metrics in assessing the extent of decentralization in the network, in blackbox fashion, ensuring that Ethereum remains faithful to its core principles.

The study has acknowledged its limitations, even with its contribution to it. Training GNNs on large scale blockchain data presents scalability issues when we seek to use the trained model in real time. Furthermore, feature extraction is still a very involved task because of Ethereum's array and dynamic ecosystem. Improvement of the practical applicability of the proposed framework will be of vital importance when addressing these challenges.

Further research could add to this study by observing real time node influence by continuous exposure to network dynamics with streaming data. Explainable AI techniques, on the other side, are more transformative by putting forward Explainability of

Spectrum of Engineering Sciences
Online ISSN
3007-3138
Print ISSN
3007-312X
SPECTRUM OF ENGINEERING SCIENCES

machine learning models and making interpretability better with ML predictions. Furthermore, node behaviour of different blockchains can be compared and these insights could be helpful in understanding decentralized systems.

Finally, scalability of decentralized networks via machine learning is the contribution of this research to the growing discipline of blockchain analytics. Through identifying and analyzing influential nodes, not only network dynamics of Ethereum are improved but the study also paves the road towards innovations in blockchain security, scalability and efficiency. The insights gained provide assurance that machine learning offers the promise of being a defining tool in building the future decentralized systems.

References

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Chen, X., Liu, Y., & Wu, Q. (2020). Blockchain network dynamics: A graph-theoretic perspective. *Journal of Blockchain Research, 12*(3), 45–60. https://doi.org/10.1016/j.jblre.2020.03.006

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *Proceedings of the International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.1609.02907

Lin, Z., Xu, J., & Wang, M. (2019). Identifying central nodes in Bitcoin and Ethereum networks. *Computational Blockchain Analytics, 8*(2), 101–115. https://doi.org/10.1016/j.compbla.2019.02.001

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

Pellegrini, C., & Tasca, P. (2021). Identifying influential nodes in Ethereum's decentralized network. *Blockchain Research and Applications, 3*(4), 100015. https://doi.org/10.1016/j.bcra.2021.100015

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications, and research directions. *SN Computer Science, 2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Yuan, Q., & Wu, J. (2020). Decentralized ledger analytics using machine learning and graph models. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50*(11), 4315–4327. https://doi.org/10.1109/TSMC.2020.2966358

Zhou, K., Zhang, Y., & Li, H. (2021). Graph neural networks for blockchain fraud detection. *Neural Computing and Applications, 34*(7), 1538–1549. https://doi.org/10.1007/s00521-021-05916-9