Research on the Important Accounts Identification in Blockchain Trading Networks

Min Zhang[1*], Yongsheng Zhao[1], Zesan Liu[1], Jian Zhang[2], Yunxi Fu[1], Zhenya Wang[1]

mz2016bupt@163.com*

State Grid Information & Telecommunication Group CO., LTD, Beijing, 100052, China[1]

State Grid Tianjin Electric Power Company, Tianjin, 300000, China[2]

Abstract. Important accounts play an important role in guaranteeing the security and stability of blockchain trading networks, given that attacking them can increase the risk of account theft and disrupt the trade order. Consequently, the identification of critical accounts within blockchain trading networks holds paramount significance. However, previous research usually focuses on individual account features while neglecting the impacts of neighbors, leading to biased assessments and inaccurate ranking lists. To overcome these limitations, this paper proposes the NDL algorithm to identify critical accounts in the blockchain trading networks based on complex network methods. Specifically, NDL utilizes degree centrality to compute the attributes of an account itself, and employs the shortest paths to calculate the attributes of its neighbors. By comprehensively considering the influence of accounts and neighbors, NDL effectively distinguishes their importance. Besides, the Susceptible-Infectious-Recovered (SIR) model is employed to estimate the transmission potential of accounts. In addition, Kendall's tau correlation coefficient and monotonicity index are employed to assess the effectiveness and distinguishability of NDL. After conducting thorough experiments on four datasets, the findings demonstrate that NDL outperforms six baseline methods. Specifically, NDL significantly enhances the effectiveness of identifying crucial accounts and improves the distinguishability of the ranking list.

Keywords: Blockchain; Complex Network; Critical Node; Network Security.

1 Introduction

Blockchain technology holds tremendous application value and broad developmental prospect. It guarantees the transparency and immutability of information, strengthens the security and credibility of data, diminishes transaction and trust costs, and facilitates the evolution of the digital economy and the refinement of governance systems. Consequently, blockchain is regarded as one of the most disruptive innovative technologies in the upcoming decade, attracting increasing attention and application on a global scale [1][5].

The identification of important accounts within blockchain trading networks plays a significance role in guaranteeing the security and stability of blockchain networks, reducing the risk of account theft and malicious attacks, defending trade order, safeguarding public interests, and preventing fraudulent behaviour. Therefore, it is necessary to strengthen the identification of important accounts in the blockchain trading network.

ICBBEM 2024, March 29-31, Wuhan, People's Republic of China

DOI 10.4108/eai.29-3-2024.2347430

Complex network methods have found widespread application in the identification of crucial accounts within blockchain trading networks, offering insights into revealing the network structural characteristics and evolutionary processes from both global and local perspective. These methods, grounded in the realm of complex networks, can be broadly classified into six main types. (1) The first approach delves into the local network structure to reveal direct relationships between nodes, exemplified by metrics like degree centrality [6]. (2) The second approach focuses on the global network structure, providing insights into macro properties such as betweenness centrality [7] that reflects the network's overarching characteristics. (3) The third approach considers the network dynamic, capturing and analyzing the evolutionary processes over time, such as VoteRank algorithm [8]. (4) The fourth approach takes account of the node location, revealing the core structure of nodes, as exemplified by K-Shell [9]. (5) The fifth approach assesses the propagation ability of nodes, measuring their influence or effectiveness in information dissemination within a network, such as cascading failure [10]. (6) The sixth approach considers node contractility to measure the integration and connectivity of nodes within the network topology, as exemplified by the contraction algorithm [11]. In addition, complex network methods can also explore the community structure and dynamic behavior patterns in the blockchain trading networks, further revealing the associations and collaborative relationships between important accounts. These research results will contribute to enhancing the security and stability of blockchain treading networks, thereby fostering the healthy development of blockchain technology.

Inspired by the aforementioned research, this paper proposes the NDL algorithm to identify critical accounts within blockchain trading networks. Specifically, the NDL algorithm takes into account both the attributes of node itself and the characteristics of its neighbors, addressing the one-sidedness issues observed in the other research. Experiment results indicate that NDL outperforms the baselines in both effectiveness and distinguishability.

This paper is divided into four distinct segments. The second segment elucidates the methodological approach undertaken in this research. The third segment involves the execution of comprehensive experiments to assess the effectiveness of the NDL algorithm. The fourth and final segment presents the ensuing conclusions.

2 Method

There are two approaches to model blockchain trading networks, including transaction-based method and account-based method. In the transaction-based approach, transactions are abstracted as nodes, and users are represented as edges. In the account-based method, accounts are treated as nodes, and the fund flows are depicted as edges. In this paper, we adopt the account-based method to model the blockchain trading network, which is described as

$$G = (V, E), \tag{1}$$

where $V = \{v_{1,} v_{2,} \dots, v_N\}$ is the set of nodes, v_i is the account, N is the number of nodes, $E = \{e_{ij} | v_i \cap v_j , v_i \neq v_j\}$ *is* the set of edges, e_{ij} denotes fund flow between account v_i and account v_j.

The influence of node within the blockchain trading network depends on the attributes of node itself and the characteristics of its neighbors. Firstly, for each node $i \in G$, we use the degree centrality to represent the influence of itself, namely

$$\Theta(i) = k_i/N - 1,\ \ \Delta(i) = (\Theta(i))^2, \tag{2}$$

where k_i is degree of node i. Secondly, we compute the attributes of neighbors, including the degree centrality and the number of shortest paths, namely

$$\Phi(j) = (\Theta(i) + \Delta(i)) \times (\Theta(j) + \Delta(j))/g_{ij}^2, \tag{3}$$

The sum of attributes of neighbors is denoted as $\Xi(i)$, namely

$$\Xi(i) = \sum_{j \in V} \Phi(j). \tag{4}$$

Thirdly, we compute the influence of node i according to its first-order neighbors, namely

$$\Omega(i) = \sum_{j \in \Gamma(i)} \Xi(j), \quad NDL(i) = \sum_{j \in \Gamma(i)} \Omega(j), \tag{5}$$

where $\Gamma(i)$ is the set of first-order neighbors. The pseudo-code of NDL is shown in Algorithm 1.

Algorithm 1 The NDL Algorithm

Input: blockchain trading network $G = (V, E)$
Output: the influence scores of all nodes
for $i \in V$ **do**
    $\Theta(i) = k_i/N - 1$←Get the degree centrality for node i in G
    $\Delta(i) = \Theta(i)^2$←Get the square of degree centrality for node i in G
end for
for $i \in V$ **do**
    for $j \in V$ **do**
        g_{ij}←Calculate the quantity of shortest paths between node i and node j
        $\Phi(j) = (\Theta(i) + \Delta(i)) \times (\Theta(j) + \Delta(j))/{g_{ij}}^2$←Get the product of node i and node j
        $\Xi(i) += \Phi(j)$←Get the sum of attributes Φ of neighbors on node i
    end for
end for
for $i \in V$ **do**
    for $j \in \Gamma(i)$ **do**
        $\Omega(i) += \Xi(j)$←Get the quantity of influence $\sum$ of first-order neighbors on node i
    end for
end for
for $i \in V$ **do**
    for $j \in \Gamma(i)$ **do**
        $NDL(i) += \Omega(j)$←Get the quantity of influence Ω of first-order neighbors on node i
    end for
end for
return the influence scores of all nodes based on NDL

3 Experiment

3.1 Evaluation Metrics

(1) Kendall's correlation coefficient (τ)is utilized to assess the effectiveness ability of NDL. Assuming X and $Y$are two ranking lists with length N_L, select the *i-th* element to form a pair (x_i, y_i). If x_i>x_j and y_i>y_j or x_i<x_j and y_i<y_j, $(x_i, y_i)$is a concordant pair. Otherwise, (x_i, y_i) is a discordant pair except when x_i=x_j and y_i=y_j. The definition of τ is

$$\tau(X,Y) = \frac{2(N_c - N_d)}{N_L(N_L - 1)}, \tag{6}$$

where N_c and N_d are the quality of concordant pairs and discordant pairs, respectively. The larger the values of τ, the better the effectiveness ability of the algorithm.

(2) Monotonicity index (MI) is utilized to evaluate the distinguishing ability of the NDL algorithm. The definition of MI is

$$MI(X) = \left(1 - \frac{\sum_{\alpha \in L'} N_\alpha (N_\alpha - 1)}{N_L(N_L - 1)}\right)^2, \tag{7}$$

where X is the sorted list with length N_L, α is the sorted value. The larger the values of MI, the better the distinguishing ability of the algorithm.

3.2 Baselines

(1) Degree Centrality (DC) [6]. The DC algorithm considers that nodes with a greater number of neighbors will have significant influence.

$$DC(i) = \frac{k_i}{N - 1}, \tag{8}$$

where k_i is the degree of node i.

(2) Between Centrality (BC) [7]. The BC algorithm considers that nodes occupying pivotal hub positions will exert important influence.

$$BC(i) = \sum_{i \neq u \neq v} \frac{g_{uv}^i}{g_{uv}}, \tag{9}$$

where g_{uv}^i is the number of shortest paths passing through node i.

(3) VoteRank (VR) [8] . The vital spreaders are determined by the VoteRank algorithm based on voting scores, which will not join in the subsequent voting. The VoteRank algorithm has five process. Firstly, the voting score and the quality of votes for each node is 0 and 1, respectively. Secondly, each node and its neighboring nodes will conduct another round of voting to determine their voting scores. Thirdly, the vital spreader is selected based on the

voting scores. Fourthly, the voting power of nodes supporting the vital spreader will diminish. Fifthly, repeat steps two through four until multiple spreaders are obtained.

(4) H-Index (HI) [12] .

$$HI(i) = H(k_{j_1}, k_{j_2}, \ldots, k_{j_{k_i}}), \tag{10}$$

where k_{j_x} is the degree of node j_x, H is an operator, (k_{j_1} ,k_{j_2} ,...,$k_{j_{k(i)}}$) is a set, *HI(i)*>0 is the highest integer that is less than or equal to k_{j_x}, $x = 1,2,\ldots,k_i$.

(5) GLS [13].

$$GLS(i) = \left(k_i \sum_{j\in\Gamma_i} 1.1^{com(i,j)}\right) \times \left(\sum_{j\in\Gamma_i} \frac{k_i}{N-1} \times \frac{k_i}{\sum_{u\in\Gamma_j} k_u}\right), \tag{11}$$

where *com(i, j)* is the number of common neighbors.

(6) RAS [14].

$$RAS(i) = s_{iv'} = \sum_{z\in\Gamma(i)\cap\Gamma(v')} \frac{1}{k_z}, \tag{12}$$

where node v' is virtual and connects to all nodes.

3.3 SIR Model

The Susceptible-Infectious-Recovered (SIR) model is a mathematical framework used to analyze how infectious diseases spread within populations. Specifically, “S” represents nodes in a susceptible state, “I” indicates nodes in an infected state, and 'R' signifies nodes in a recovered state. θ is the infection rate, which represents the probability of each susceptible node being infected after contact with the infected node. β is the recovery rate, which represents the probability of an infected node transitioning to the recovered state. The infection probability threshold θ_c is

$$\theta_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{13}$$

In this paper, we utilize a variable parameter ζ to change the infection rate θ, i.e., $\theta = \zeta \times \theta_c$.

3.4 Datasets

Numerous studies have consistently shown that the degree distribution of blockchain trading networks follow a power-law distribution $P(k) = k^{-\gamma}$ [15]. This phenomenon is attributed to the presence of preferential attachment and cumulative advantage in blockchain trading networks. This observation indicates that, within the network, the majority of nodes exhibit small degrees, while a minority of nodes possess larger degrees. Most nodes tend to engage in transactions with only a few others, whereas a small subset of nodes actively participates in transactions with a larger number of nodes. In this paper, we leverage the Barabási-Albert (BA) network to generate four blockchain transaction networks, aiming to verify the effectiveness and distinguishability. The statistical indicators of four datasets are shown in Table 1, in which M is the number of edges, ρ is the network density, $<k>$ is the average degree, k_{max} is the largest degree, C is the average clustering coefficient, AS is the degree assortativity, HE is the degree heterogeneity.

Table 1. The statistical indicators of four blockchain transaction networks.

	N	M	ρ	$<k>$	k_{max}	C	AS	HE
Dataset-1	200	396	0.0198	3.960	26	0.0919	-0.2147	2.0731
Dataset-2	500	996	0.0079	3.984	64	0.0470	-0.1335	2.4314
Dataset-3	1000	1996	0.0039	3.992	92	0.0243	-0.0952	2.8684
Dataset-4	2000	3996	0.0019	3.996	94	0.0142	-0.0741	2.8595

3.5 Results

Firstly, we verify the distinguishability of the NDL algorithm on the four datasets. As shown in Table 2, these quantities are accurate to six decimal places. The *MI* values of the NDL algorithm is the largest, indicating that the nodes in the sorted list generated by NDL exhibit nearly distinct positions. It is worth noting that nodes with the same sorting position may cause conflicts in the datasets, as they are treated as identical entities. The NDL algorithm achieves an accuracy of 99.9% in distinguishing node sorting positions, enhancing the intuitiveness of evaluating node influence.

Table 2. The monotonicity indices of six baselines and the NDL algorithm on the four datasets. The parameters are β=1, ζ=2. The monotonicity index of NDL is the largest.

	DC	BC	HI	GLS	VR	RAS	NDL
Dataset-1	0.485998	0.991876	0.203632	0.999498	0.033897	0.992877	**0.999699**
Dataset-2	0.480030	0.997517	0.209423	0.999888	0.024340	0.993693	**0.999936**
Dataset-3	0.483021	0.999508	0.204447	0.999968	0.029658	0.993851	**0.999976**
Dataset-4	0.490216	0.999824	0.227401	0.999978	0.051778	0.994367	**0.999994**

Secondly, we verify the effectiveness of NDL across the datasets. As depicted in Figure 1, the x-axis is ζ used to adjust the infection rate θ, while the y-axis is the Kendall's correlation coefficient τ.The red column represents the performance of the NDL algorithm, exhibiting the highest accuracy. This indicates that the NDL algorithm performs better than the baselines, including degree centrality, betweenness centrality, H-Index, GLS, VoteRank and RAS. The reason is that the baselines, such as betweenness centrality and H-Index, have one-sidedness issues, leading to inaccurate ranking lists. Similarly, baselines like GLS and RAS have not fully explored the contributions of node neighbors. In contrast, the NDL algorithm takes a comprehensive approach by considering the influence of both nodes and their neighbors. It leverages degree centrality to calculate the attributes of node itself and utilizes shortest paths to compute the attributes of its neighbors, thereby generating an accurate ranking list.

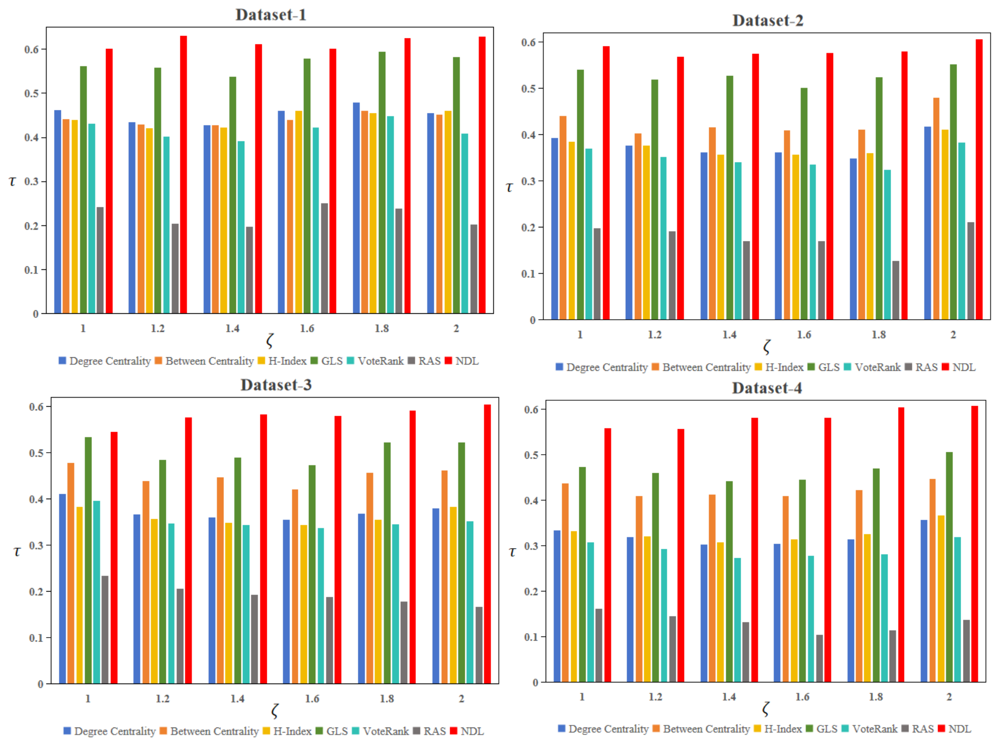

Fig. 1. The Kendall' s tau correlation coefficients are calculated to verify the effectiveness of six baselines and NDL on the four datasets, with varying infection probability. The parameter is β=1. The red column represents the performance of the NDL algorithm, with the highest accuracy.

Thirdly, we compare the execution time of six baselines and NDL on the datasets. Table 3 shows the running time of NDL is moderate. The reason is that the NDL algorithm need to calculate the shortest paths. We will optimize the algorithm to reduce the running time in the future.

Table 3. The running time of six baselines and NDL on four datasets. The parameters are β=1, ζ=2.

	DC	BC	HI	GLS	VR	RAS	NDL
Dataset-1	0.000001	0.048001	0.000001	0.007989	0.011996	0.000001	0.418334
Dataset-2	0.015590	0.328090	0.000001	0.015620	0.140613	0.000001	3.908138
Dataset-3	0.050220	1.577959	0.000001	0.046855	0.874907	0.000001	19.57481
Dataset-4	0.062000	11.82943	0.005000	0.172071	10.00058	0.006921	114.7484

4 Conclusion

In this paper, we propose the NDL algorithm to identify important accounts in the blockchain trading networks based on complex network methods. In contrast to previous studies, NDL takes a more comprehensive consideration of the influence of accounts and their neighbors. Specifically, NDL utilizes degree centrality to compute the attributes of account itself, and employs the shortest paths to compute the attributes of its neighbors. Besides, the SIR model is utilized to assess the propagation effectiveness of each account. To assess the effectiveness and distinguishability of NDL, two metrics of Kendall's tau correlation coefficient and monotonicity index are introduced. By conducting extensive experiments on four datasets, the results indicate that NDL outperforms the six baseline methods, exhibiting the highest values of Kendall's tau correlation coefficient and the values of monotonicity index reach 99.9%. It indicates that NDL can produce a list with superior effectiveness and distinguishability. In addition, the running time of NDL is moderate. NDL holds significant application values, serving as a crucial foundation for identifying important accounts within blockchain trading networks and contributing to their security and stability.

Acknowledgments. The work is funded by the National Key R&D Program of China (2022YFB2403900)

References

[1] Oguntegbe K F, Di P N, Vona R. Communicating responsible management and the role of blockchain technology: social media analytics for the luxury fashion supply chain[J].The TQM Journal, 2024.

[2] Chakraborty S, Jyothi N M, Lekhi S, et al. Multi technology ensembled optimized enterprise architecture for retail industry[C]. 2023 International Conference on Computational Intelligence for Information, Security and Communication Applications, 2023, 137-142 .

[3] Bazel M A, Mohammed F, Ahmed M. Blockchain technology in healthcare big data management: benefits, applications and challenges[C]. 2021 1st International Conference on Emerging Smart Technologies and Applications, 2021, 1-8.

[4] Peng J, Lei Z, Siming Y, et al. Blockchain technology applications in waste management: overview, challenges and opportunities[J]. Journal of Cleaner Production, 2023, 421: 138466.

[5] Abhijeet G, Michael B, Sachin K. Blockchain implementation in pharmaceutical supply chains: a review and conceptual framework[J]. International Journal of Production Research, 2023, 61: 6633-6651.

[6] Xu X, Zhu C, Wang Q Y, et al. Identifying vital nodes in complex networks by adjacency information entropy[J]. Scientific Report, 2020, 10(1): 1-12.

[7] Lv Z W, Zhao N, Xiong F, et al. A novel measure of identifying influential nodes in complex networks [J]. Physica A: Statistical Mechanics and its Applications, 2019, 523: 488-497.

[8] Liu J G, Lin J H, Guo Q, et al. Locating influential nodes via dynamic sensitive centrality[J]. Scientific Report, 2016, 6: 21380.

[9] Yang Q, Wang Y H, Yu S, et al. Identifying influential nodes through an improved k-shell iteration factor model[J]. Expert Systems with Applications, 2024, 238: 122077.

[10] Wang B, Zhang Z, Qi X, et al. Identify critical nodes in network cascading failure based on data analysis[J]. Journal of Network Systems Management, 2020, 28: 21-34.

[11] Doi T, Endres M G. Unified contraction algorithm for multi-baryon correlators on the lattice[J]. Computer Physics Communication, 2013, 184 (1): 117-123.

[12] Khurana P, Sharma K. Impact of h-index on authors ranking: a comparative analysis of scopus and WoS [J]. arXiv.2102.06964, 2021.

[13] Sheng J F, Dai J Y, Wang B, et al. Identifying influential nodes in complex networks based on global and local structure[J]. Physica A: Statistical Mechanics and its Applications, 2020, 541(3): 123262.
[14] Ai J, He T, Su Z. Identifying influential nodes in complex networks based on resource allocation similarity[J]. Physica A: Statistical Mechanics and its Applications, 2023, 627: 129101.
[15] Chen B J. Research of the transaction behavior of ethereum users based on complex network theory[D]. University of Electronic Science and Technology, 2018.