Evolution of the External Owned Account Trading Network on Ethereum

1st Yunjie Chen
Institute of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
rothleer@outlook.com

2nd Zhihai Rong
Institute of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
rongzhh@gmail.com

***Abstract*—In this paper, we investigate the evolution of the External owned account Trading Network(ETN) on Ethereum from August 10th,2015 to June 9th,2017 by complex network theory. The result shows that the evolution of the topological properties of the ETN and important events are closely related. The size of the ETN follows important events. The average degree of the ETN is not constant but it fluctuates within a small range of $(2.1, 3.2)$ and the extremely high correlation coefficient between the number of nodes and the number of edges in the ETN which indicates that the relationship is linear. Moreover, following the events happening, the degree distribution of ETN change obviously. Compared with the randomized network, the ETN has a smaller degree correlation coefficient and clustering coefficient. By using the method of k-shell decomposition, we find that the core of the ETN grows stepwise.**

***Index Terms*—Blockchain, Ethereum, Complex network**

I. Introduce

Blockchain technology originated from the seminal paper published by Satoshi Nakamoto in 2008 [1]. It features distributed high redundancy storage, temporal data and unmodifiable and falsifiable, decentralized credit, etc [2]. Many projects have been proposed based on the features of the blockchain, the sharing of management resources for research projects [3], the sharing of electronic medical records [4], the sharing platform for educational resources [5], and so on.

The anonymity of blockchain technology guarantees privacy, but it also brings difficulties to supervision. The well-known virus wannacry uses the anonymous feature of blockchain to extort Bitcoins from users. Therefore, how to supervise the blockchain is an important issue. Ethereum is one of the most famous public blockchain projects. The complex network theory provides a theoretical framework [6] [7] for researching the network composed of complex transaction behaviors among Ethereum users. Chen Ting's seminal work provided a method for reading Ethereum data and found that the degree distribution of the Ethereum transaction network obeyed the power-law distribution [8]. In this paper, we focus on the evolution of the ETN to increase understanding of blockchain and to provide references for the application and supervision of blockchain projects.

II. Datasets And Definitions

A. Datasets

The data of user transactions on Ethereum from [8]. There are two types of accounts on Ethereum, one is an external owned account controlled by users such as real people, organizations, and wallet applications. Such accounts can actively initiate and accept transactions. The other type of account is the Smart Contract Account. The smart contract account can only passively accept transactions from external owned accounts and call smart contracts stored inside the smart contract account. Because smart contracts are not widely used, for the time being [8], we only use data from transactions between external owned accounts. i.e. The sender and receiver are both external owned accounts.

External events will affect the trading decisions of Ethereum users. To explore the impact of external events on the evolution of the Ethereum external account trading network (abbreviation: ETN), we collected events from August 10, 2015, to June 9, 2017, on Ethereum official blog site and the internet [9] [10]. As shown in Table I. We have enumerated important events in chronological order. The content of events implies that events 1, 2, 3, 7, 8, 9 are positive. Events 4, 5, and 6 are negative events. Different events have different effects on ETN. In the following sections, we will focus on the discussion that different events have an impact on the topological properties of Ethereum.

B. Network Definitions

Ether is a digital currency on Ethereum. A transaction is made by sending Ether from one account to another account. We focus on the transactions between external owned accounts that form the ETN. Therefore, in the construction of the network, we ignore transactions whose sender and receiver are the same account. Definition: $G = (V, E)$, where V is the set of external accounts involved in the transaction. E is the set of edges. $E = \{(v_i, v_j)\,|v_i, v_j \in V\}$. An edge (v_i, v_j) indicates that a transaction has occurred between accounts v_i and v_j. Thus ETN is an undirected and unweighted graph. We use a total of 670 days of data from August 10th, 2015 to June 09th, 2017, containing a total of 28217982 transactions, 2079600 nodes, and 4246738 edges. To highlight the evolution

TABLE I
IMPORTANT EVENTS ON ETHEREUM

No.	*Date*	*Event*
1	16.01.01	Ethereum was recognized by the market and its currency value has risen sharply.
2	16.03.14	The first official version *Homeland* released
3	16.04.30	THE DAO project started crowdfunding and it raised more than 100 million U.S. dollars worth of ether.
4	16.06.17	The DAO project has discovered some bugs and was hacked, and transferred a large amount of Ether.
5	16.07.20	For the DAO event. Ethereum developers decided to carry out the hard fork protecting THE DAO project. Violates the decentralized concept. It triggered the division of the Ethereum community.
6	16.09.22	Jeffrey Wilcke, the co-founder of Ethereum, announced that the Ethereum network suffered a DOS attack.
7	16.11.22	Spurious Dragon hard fork solves DOS attacks.
8	17.03.01	The Ethereum Enterprise Alliance was established, with important members including J.P. Morgan, Microsoft, and Intel.
9	17.05.22	The Ethereum Enterprise Alliance added 86 new members.

of the network, we built the network daily. That is a total of 670 networks.

III. RESULTS AND ANALYSIS

A. The Evolution of ETN's Size Follows Events

To increase the understanding of ETN, we investigated the relationship between the size versus time of ETN. Fig 1 shows that, in general, the number of nodes increases with time. However, the number of nodes in the network is also affected by important events. Events 1, 8, and 9 are positive news for external users to participate on Ethereum transactions. Therefore, the size of the network increases significantly as events occur. Events 2, 3, 4, and 7 made the number of nodes fluctuate. Event 5 split Ethereum. Therefore, the number of nodes in ETN stops growing. Event 6 is a DOS attack on the Ethereum network. It may make the service speed of Ethereum to be slow. It may reduce the user's willingness to make a transaction. Therefore, the number of nodes in ETN will decrease in the following time, until event 7 completely fixes the vulnerability. Besides, it can be seen from Fig 1 that whenever an event occurs, the network scale will fluctuate. Especially after event 7. Fig 1 implies that users are sensitive to events.

The size of ETN represents the number of users participating in the transaction at the corresponding time. This reflects the heat of the market. Naturally, as the reputation of Ethereum increases, external users will be attracted to the Ethereum trading market. We call users who have never participated in transactions before as new users. Therefore, we investigated the new user growth rate of ETN, and we found that between event 2 and event 8, the growth rate of new users of Ethereum was linear and increased by about 1,950.37 nodes per day on average. The number of Ethereum users increased super linearly from event 8 to event 9, afterward and before event 2. Therefore, different events have different influences on Ethereum users. Events 1, 8, 9 can attract new users to participate while events 2, 3, 4, 5, 6, 7 affect existing users to participate in transactions.

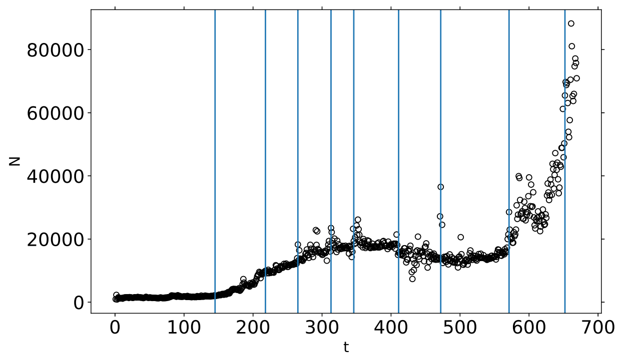

Fig. 1. **the number of nodes N on ETN versus time.** The time labels, from 1 to 670, correspond to each day from August 10th,2015 to June 9th,2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

B. Densification Power-law in the ETN

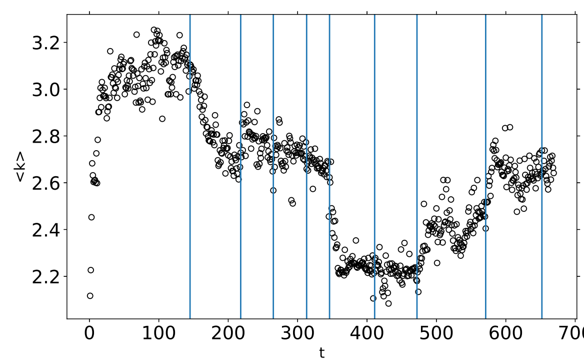

Fig. 2. **Average degree $\langle k \rangle$ of ETN versus time.** The time labels, from 1 to 670, correspond to each day from August 10th,2015 to June 9th,2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

Numerous studies have shown that the evolution of the numerical relation between nodes and edges in many networks obeys the dense power-law $M(t) \sim N^a(t)$ [11]. For exploring the number relation between ETN nodes and edges, We investigate the relationship between the nodes $N(t)$ and the number of edges $M(t)$ evolving. We found that the correlation coefficient of the two $\rho > 0.996$ which implies that the relationship between the number of nodes and edges is likely to be linear i.e. $a = 1$. According to the definition of the average degree

$$\langle k \rangle (t) = \frac{2M(t)}{N(t)}. \quad (1)$$

If the relationship between the number of nodes and edges of ETN is linear, the average degree will be constant. However, Fig 2 shows that the average degree is not a constant but evolves over time and shows that the evolution of the average degree is closely related to the occurrence of events.

The average degree of ETN decreases to about 2.7 when event 1 occurs until event 2 stops, probably because new users are entering Ethereum because of the rise on Ethereum, but these users do not participate in transactions very often, so the average degree decreases. Events 5 and 6 are bad news for Ethereum, so the average degree decreases. Events 7, 8, and 9 are good news, so the average degree increases. Although the average degree is not a constant, Comparing to other networks, the range of fluctuations $(2.1, 3.2)$ is small. Therefore, we consider that the density of ETN tends to be constant over time.

C. The Evolution of ETN's Heterogeneous

Degree distribution $P(k)$ representing the probability of finding a node with k degrees. It is an important property for understanding the network [12]. Numerous studies have shown that the degree distributions of many real networks obey the power-law distribution $P(k) \sim k^{\gamma}$, and the power exponent of the degree distribution γ indicates the heterogeneity of the network. The smaller the γ the more heterogeneous the network is. Therefore, we investigated the evolution of the degree distribution of ETN, as shown in Fig. 3, which obeys power-law distribution.

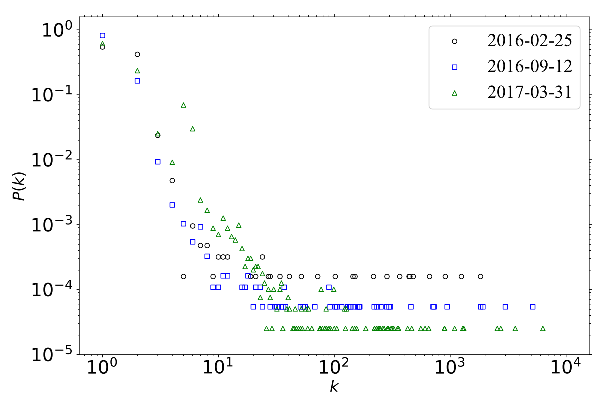

Fig. 3. **The degree distribution of ETN,** extracted in February 25th, 2016, September 12th, 2016 and March 31th, 2017. Three snapshots obey the power-law distribution $P(k) \sim k^{\gamma}$ where γ is in the range of $(2.85, 4.46)$.

To further explore the evolution of ETN's heterogeneity over time. We use the variance of degree equation

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 , \tag{2}$$

to evaluate the heterogeneity of degree sequences. where $\langle \bullet \rangle$ indicates an average value over nodes and k is the degree of a node. The bigger variance is, The more heterogeneity the degree sequence is. From Fig. 4 , we observe that ETN evolution of variance of degree on Ethereum follows the event. Between event 1 and event 2, the heterogeneity of the ETN gradually increases. Between events 2 and 5, the heterogeneity of the network slowly increases. After event 6, σ^2 decreases sharply until event 7 fixes the problem. Between events 7 and 9 the ETN's σ^2 rises.

Besides, the maximum of the network is also a property of the heterogeneity of the network. Fig 5 shows the evolution of the ETN's maximum degree over time, we find that the evolution of k_{max} and σ^2 is consistent with the trend. From the evolution of k_{max} and σ^2, we find that the heterogeneity of the network increases with positive events and decreases with negative events. The evolution of the maximum degree of the network also depicts the evolution of the influence of the most influential nodes in the network. Compared to σ^2, k_{max} has a higher growth rate between event 2 and event 6 which implies that, as the network heterogeneity increasing, the influence of the central node of the network increases.

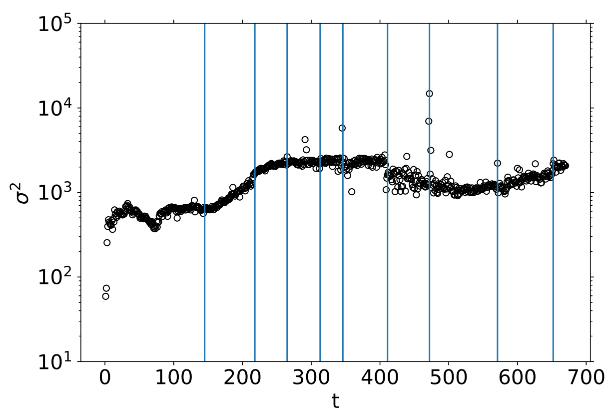

Fig. 4. **The variance of the degree sequence σ^2 on ETN versus time.** The time labels, from 1 to 670, correspond to each day from August 10th,2015 to June 9th,2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

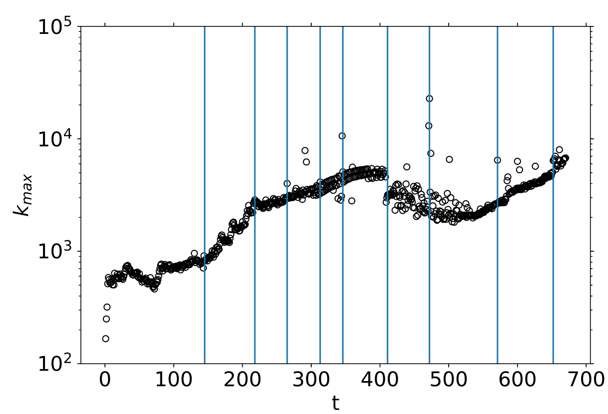

Fig. 5. **The maximum degree k_{max} on ETN versus time.** The time labels, from 1 to 670, correspond to each day from August 10th,2015 to June 9th,2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

D. ETN has a degree disassortative mixing pattern

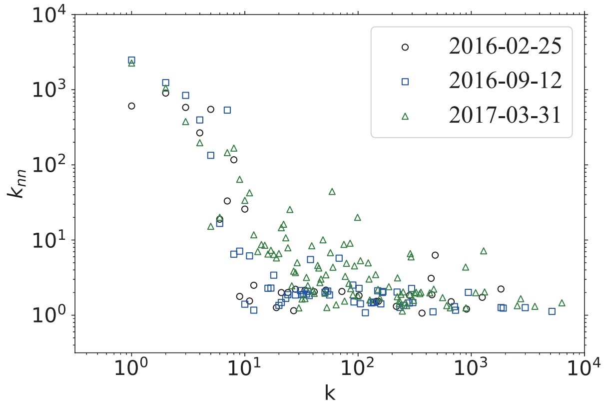

Fig. 6. **Log-log plot of the mean of the average neighbor degree $\langle k_{nn} \rangle$ versus degree k,** extracted on February 25th, 2016, September 12th, 2016, and March 31th, 2017.

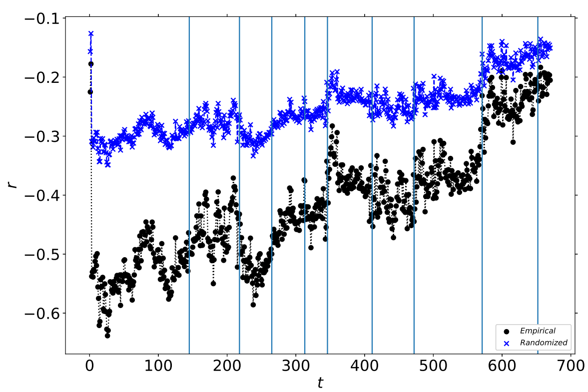

Fig. 7. **The degree correlation coefficient of ETN versus time.** The black cycles denote empirical degree correlation coefficient r(t) of ETN and blue $\times$ denote randomized the degree correlation coefficient $r'(t)$ of the randomized graph which is generated by the configuration model with a given degree sequence. Each result is obtained by 10 randomized graphs. The time labels, from 1 to 670, correspond to each day from August 10th, 2015 to June 9th,2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

we used the mean average neighbor degree [13]:

$$\langle k_{nn} \rangle = \frac{1}{i_k} \sum_{i=1}^{i_k} \langle k_{nn} \rangle_i \ , \tag{3}$$

to research the ETN is degree assortative or disassortative. i_k is the number of nodes with degree k. $\langle k_{nn} \rangle_i$ is average neighbor degree of node i. If a network is degree assortative(disassortative), $\langle k_{nn} \rangle (k)$ will be an increasing(decreasing) function of k. Fig. 6 shows that $\langle k_{nn} \rangle (k)$ is decreasing function of k. It means that, on average, ETN is degree disassortative.

To further explore, whether the nodes in the network are more inclined to connect to nodes with similar degrees or to connect to nodes with different degrees. We use the degree correlation coefficient equation [14]:

$$r = \frac{\langle ij \rangle - \langle i \rangle \langle j \rangle}{\langle i^2 \rangle \langle j^2 \rangle}, \tag{4}$$

where i and j are the degrees of two nodes at the two ends of an edge and $\langle \bullet \rangle$ represents the average over all edges. If the degree correlation coefficient of one network is negative(positive), the network has a disassortative(assortative) mixing pattern which implies hubs tend to connect non-hubs(hubs). However, the degree correlation coefficient is dependent on degree distribution [15] so it is necessary to compare with randomized graphs with the same degree sequence. We obtain randomized graphs by configuration model with a given degree sequence [6]. The steps for generating a randomized graph by configuration model as follows:(i) We give each node i of k_i "stubs". (ii) We randomly choose two of stubs, then connecting them to be one edge. (iii) repeating step(ii) until "stubs" are used up. It may exist some self-edges or multi-edges, or both in the generated graph so each numerical result is calculated by 10 times. Fig. 7 shows that the empirical degree correlation coefficient is smaller than the randomized graph. It indicates that the ETN has a disassortative mixing pattern by degree. Furthermore, the empirical degree correlation coefficient fluctuates more obviously and is affected more significantly by events.

E. The ETN is less clustered than Randomized Network

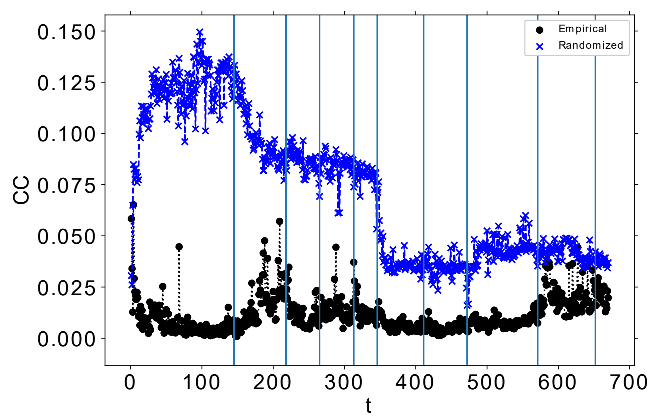

Fig. 8. **The clustering coefficient of ETN versus time.** The black cycles denote empirical clustering coefficient $CC(t)$ of ETN and blue $\times$ denote randomized the clustering coefficient $CC'(t)$ of the randomized graph which is generated by the configuration model with a given degree sequence. Each result is obtained by 10 randomized graphs. The time labels, from 1 to 670, correspond to each day from August 10th, 2015 to June 9th, 2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

The clustering coefficient is used to characterize the probability of one node's neighbors are also neighbors. It indicates how close the neighbors of the node are connected [16]. The clustering coefficient of node i is defined by the equation:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \tag{5}$$

where k_i is the degree of i and E_i is the number of edges between neighbors of node i. The clustering coefficient of a network CC is the average clustering coefficient value over all nodes. i.e.

$$CC = \langle C_i \rangle . \tag{6}$$

As it is shown in Fig. 8, the clustering coefficient CC of ETN is smaller than the randomized graph. It implies that it is hard to create triangles in ETN on a daily scale.

F. The Evolution of ETN's maximum shell

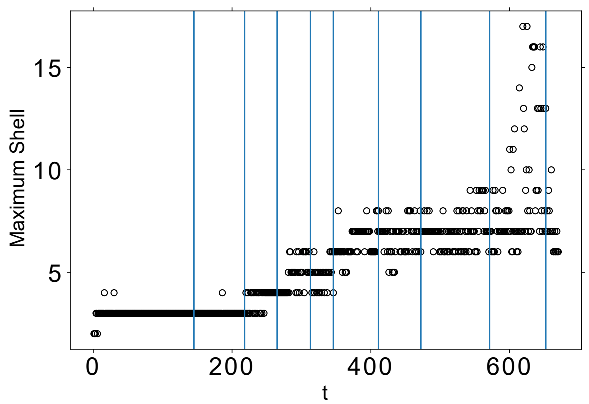

Fig. 9. **The maximum shell of ETN versus the time.** The time labels, from 1 to 670, correspond to each day from August 10th, 2015 to June 9th, 2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

Hierarchy property is a crucial structural of networks. K-shell is a measure to research the hierarchy of networks [17]. The maximum shell is the core of ETN. The steps for the k-shell decomposition method as follows: (i)Repeating deleting nodes with degree k and until there is no node with degree k. (ii) Increasing k by 1. (iii) Repeating step(i), step(ii) until no node left in the network. From Fig. 9, we observe that ETN's maximum shell grows stepwise and it affected obviously by events. Fig. 10 shows that the evolution of the maximum shell's size. It grows rapidly at the start because of the smaller k of maximum shell and following the k of maximum shell it decreases until event 8. After event 8 the size of ETN grow rapidly but at the same time, the k of maximum shell also grows so the size of maximum shell grows a little.

IV. Conclusion and Discussion

In this paper, we investigate the evolution of ETN from August 10th, 2015 to June 9th, 2017. The evolution of topology properties in ETN has a close relationship with events. The positive(negative) events may increase(decrease) size and heterogeneity of ETN. The average degree of ETN fluctuates slightly and the high correlation coefficient between nodes and edges implies that they have a linear relationship. The degree correlation coefficient and the clustering coefficient are small than the randomized graph means that ETN has a degree

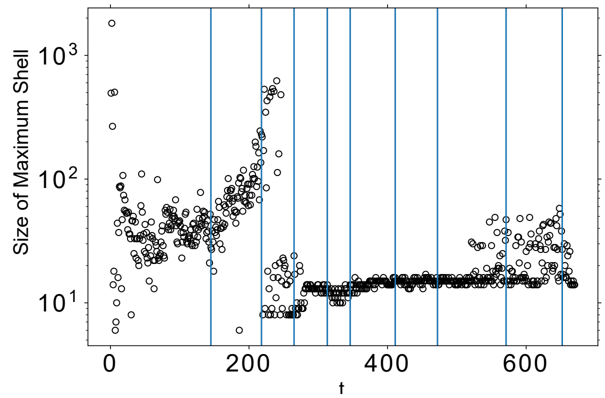

Fig. 10. **The size of the maximum shell of ETN versus the time.** The time labels, from 1 to 670, correspond to each day from August 10th, 2015 to June 9th, 2017. The blue vertical line in the figure represents the time when the events in Table I occurred.

disassortative mixing pattern and hard to create triangles. By k-shell decomposition method, ETN's maximum shell grows stepwise. This research increases understanding of blockchain and providing references for the application and supervision of blockchain projects.

References

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[R]. Manubot, 2019.

[2] Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. Applied Innovation, 2(6-10), 71.

[3] Bai Y, Li Z, Wu K, et al. Researchain: union blockchain based scientific research project management system[C]//2018 Chinese Automation Congress (CAC). IEEE, 2018: 4206-4209.

[4] Ni W, Huang X, Zhang J, et al. HealChain: A Decentralized Data Management System for Mobile Healthcare Using Consortium Blockchain[C]//2019 Chinese Control Conference(CCC). IEEE, 2019: 6333-6338.

[5] Hou, Y., Wang, N., Mei, G., Xu, W., Shao, W., & Liu, Y. (2019, November). Educational Resource Sharing Platform Based on Blockchain Network. In 2019 Chinese Automation Congress (CAC) (pp. 5491-5494). IEEE.

[6] Newman M E J. The structure and function of complex networks[J]. SIAM review, 2003, 45(2): 167-256.

[7] Barabási A L. The architecture of complexity[J]. IEEE Control Systems Magazine, 2007, 27(4): 33-42.

[8] Chen T, Zhu Y, Li Z, et al. Understanding ethereum via graph analysis[C]//IEEE INFOCOM 2018-IEEE conference on computer communications. IEEE, 2018: 1484-1492.

[9] https://blog.ethereum.org/

[10] https://www.coindesk.com/

[11] Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: Densification and shrinking diameters[J]. ACM transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1): 2-es.

[12] Albert R, Jeong H, Barabási A L. Diameter of the world-wide web[J]. nature, 1999, 401(6749): 130-131.

[13] Pastor-Satorras R, Vázquez A, Vespignani A. Dynamical and correlation properties of the internet[J]. Physical review letters, 2001, 87(25): 258701.

[14] Newman M E J. Assortative mixing in networks[J]. Physical review letters, 2002, 89(20): 208701.

[15] Zhou S, Mondragón R J. Structural constraints in complex networks[J]. New Journal of Physics, 2007, 9(6): 173.

[16] Watts D J, Strogatz S H. Collective dynamics of 'small-world'networks[J]. nature, 1998, 393(6684): 440-442.

[17] Pittel B, Spencer J, Wormald N. Sudden emergence of a giantk-core in a random graph[J]. Journal of Combinatorial Theory, Series B, 1996, 67(1): 111-151.