This is the Pre-Published Version.
The following publication B. Tao, I. W. -H. Ho and H. -N. Dai, "Complex Network Analysis of the Bitcoin Blockchain Network," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, Korea, 2021 is available at https://doi.org/10.1109/ISCAS51556.2021.9401533.

Complex Network Analysis of the Bitcoin Blockchain Network

Bishenghui Tao* , Ivan Wang-Hei Ho†, Hong-Ning Dai*
*Faculty of Information Technology, Macau University of Science and Technology
2009853YII30001@student.must.edu.mo; hndai@ieee.org
†Department of Electronic and Information Engineering, The Hong Kong Polytechnic University
ivanwh.ho@polyu.edu.hk

***Abstract*—In this paper, we conduct a complex-network analysis of the Bitcoin network. In particular, we design a new sampling method namely random walk with flying-back (RWFB) to conduct effective data sampling. We then conduct a comprehensive analysis of the Bitcoin network in terms of the degree distribution, clustering coefficient, the shortest path length, the assortativity, and the rich-club coefficient. There are several important observations from the Bitcoin network, such as small-world phenomenon and non-rich-club effect. This work brings up an in-depth understanding of the current Bitcoin blockchain network and offers implications for future directions in malicious activity and fraud detection in cryptocurrency blockchain networks.**

***Index Terms*—Blockchain, complex network, Bitcoin, network analysis**

I. Introduction

Recent years, Bitcoin has been widely accepted as one of the most representative cryptocurrencies. In 2009, the founder of Bitcoin Satoshi Nakamoto put forward this brand new digital currency [1]. Due to its innovative characteristics such as anonymity, low transaction fee, decentralized, and 24-7 accessibility, Bitcoin received extensive attention in the past few years [2].

Although there are a number of different studies on Bitcoin, there are few studies on the analysis of the Bitcoin network. It is crucial to analyze the Bitcoin blockchain from the network perspective since it can lead us to a further understanding of incumbent blockchain systems. There are several attempts in this field. The work [3] focused on the information dissemination and delay of the blockchain network while [4] and [5] initially explored the Bitcoin transaction network with a limited graph analysis. Moreover, other studies analyzed other cryptocurrencies such as Ethereum [6] or BCH [7].

However, the research on the Bitcoin transaction network is still quite limited. For example, network characteristics such as assortativity and connection tendency need to be further investigated. To this end, we conduct a multi-dimensional analysis of the Bitcoin transaction network from different perspectives. The main contributions of this paper are summarized as follows.

This work was supported in part by the General Research Fund (Project No. 15201118) established under the University Grant Committee (UGC) of the Hong Kong Special Administrative Region (HKSAR), China.

- We design a new sampling method called random walk with flying-back (RWFB) to improve data sampling in contrast to conventional sampling methods.
- We conduct a multi-dimensional analysis of the Bitcoin transaction network, covering the degree distribution, clustering coefficient and the shortest-path length, assortativity analysis, and rich-club coefficient.

II. Network construction

When users transfer bitcoin from one address to another address, transactions are made. Bitcoin transactions have multiple input addresses and multiple output addresses. In other words, one transaction may have multiple input addresses (more than one) and multiple output addresses. Thus, a transaction having inputs from n addresses and outputs to m addresses will be represented as $n \times m$ edges, in which each input address is linked with each output address [8].

We then convert a Bitcoin transaction network into a weighted directed graph denoted by $G = (V, E, W)$, where V is a set of nodes, E is a set of edges and W is a set of weights. Each edge is represented as $e_{ij} = (i, j, w_{ij})$, where i is the input address, j is the output address, and w_{ij} is the weight value. The set E of a graph with N nodes can be represented as an $N \times N$ matrix, which is essentially an adjacency matrix denoted by $\mathbf{A}$. For any element a_{ij} in $\mathbf{A}$, we have $a_{ij} = w_{ij}$ if there exists a link with weight w_{ij} between i and j; $a_{ij} = 0$ otherwise [9]. In particular, we have

$$a_{ij} = \begin{cases} w_{ij} & \text{if } e_{ij} \text{ is defined;} \\ 0 & \text{if } e_{ij} \text{ is not defined.} \end{cases} \tag{1}$$

We conduct an analysis on the Bitcoin blockchain network through collecting and extracting Bitcoin network data that are publicly available. In this work, we obtain the transaction and address data by the ELTE Bitcoin Project [10] from January 2017 to January 2018 via synchronizing real-time Bitcoin blockchain.

III. Bitcoin network sampling

We then conduct Bitcoin network sampling to extract representative samples for the further analysis.

A. Random-walk based graph sampling

Since the Bitcoin transaction network is a large massive graph with millions of nodes, it is necessary to obtain a representative sample for simplifying the analysis. Some pioneering studies [6] and [11] show that random-walk based sampling methods can well preserve structural properties of this scale-free network with a power-law degree distribution. Therefore, we introduce a graph sampling method based on a random-walk to represent the Bitcoin blockchain network.

In [12], the next-hop node j is randomly chosen from the neighbors of the current node i. The probability of moving from i to j is denoted by $P_{i,j}^{RW}$. Node i have k_i neighbors, i.e., the degree of node i. We thus have

$$P_{i,j}^{RW} = \begin{cases} \frac{1}{k_i} & \text{if } j \text{ is a neighbor of } i; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The sampled graph G_{RW} is constructed by appending every current node i and every edge (i, j) with a weighted edge: $G_{RW} = (V_i, E_{i,j}, w_{i,j})$.

However, conventional random-walk methods [13] cannot accurately sample the Bitcoin network since they only choose one neighbor in every step, consequently leading to inaccurate graph properties, e.g., a lower average degree. To address this problem, we propose an improved sampling method, namely random walk with flying-back (RWFB). In particular, RWFB considers a flying-back probability when sampling the Bitcoin network. At every step of the walk, RWFB flies back to the start node with the flying-back probability p; it chooses a random neighbor to move to with a probability of $1 - p$. Therefore, we have the RWFB probability (P_i^{RWFB}) defined as follows

$$P_i^{RWFB} = \begin{cases} \frac{(1-p)}{k_i} & \text{move to neighbor } j \text{ of node } i, \\ p & \text{fly back to } i, \end{cases} . \tag{3}$$

Thereafter, we redefine the sampled graph with flying-back function as $G_{RWFB} = (V_i, E_{i,j}, w_{i,j})$.

B. Evaluation of sampling methods

We evaluate the Bitcoin transaction graph by comparing four sampling methods: Random walk (RWS), Random Walk with Flying-Back (RWFB), Random Node Selection (RN), and Random Edge Selection (RE) [14]. It is worth mentioning that the RN method randomly selects nodes while RE uniformly selects edges at random.

For each method, we evaluate its distribution of degree, clustering coefficient, betweenness and closeness [9]. The results are compared with the original graph by the Kolmogorov-Smirnov (K-S) D-statistic [14], which measures the agreement between two distributions. The whole results and an average D-statistic value (AVG) of all sampling methods are presented in Table I. The well-perform score value in each column is highlighted in bold. We observe that the proposed RWFB outperforms other existing sample methods in most of the given metrics, implying that the sampled graph of RWFB has the greatest approximation to the original graph. Therefore, we use the RWFB method to sample the Bitcoin network in the following.

TABLE I
COMPARISON OF SAMPLING METHODS BY K-S D-STATISTIC

	Degree	Clustering	Betweenness	Closeness	AVG
RWFB	**0.120**	**0.045**	**0.091**	**0.429**	**0.171**
RWS	0.293	0.046	0.536	0.618	0.373
RN	0.895	0.053	0.151	0.433	0.383
RE	0.275	1	**0.067**	0.549	0.473

IV. COMPLEX NETWORK ANALYSIS

We then conduct the analysis on the Bitcoin network via the complex network approach.

A. Degree distribution

It is crucial to investigate node degree distribution of the Bitcoin network. The number of adjacent edges of a node is defined as degree denoted by k in the complex network theory. In the Bitcoin network, the degree k is calculated for every Bitcoin address after the summation of both incoming and outgoing transactions. Moreover, we also introduce the degree distribution denoted by $P(k)$ on degree k. The degree distribution $P(k)$ is the probability that a randomly-selected node has the degree equal to k [9].

Fig. 1 shows the degree distributions of the Bitcoin blockchain network in log-log plots (logarithmic scale in both horizontal and vertical axes). In particular, Fig. 1(a) plots the total degree distribution while Fig. 1(b) and Fig. 1(c) plot the in-degree and out-degree distributions, respectively. We can observe from the figures that all of the degree distributions follow the power-law distribution with the heavy-tail. It implies that the Bitcoin network is a scale-free network, in which only a few nodes have a large number of connections [15] while the majority of nodes are of low degrees.

B. Clustering coefficient and the shortest-path length

In order to measure the tendency of the nodes cliquishness, Watts and Strogatz introduced the clustering coefficient [16]. The clustering coefficient can measure the network from a geometric point of view. We denote the clustering coefficient, the degree, the number of complete triangles of node i by C_i, k_i, $|\Delta_i|$, respectively [5]. We then have the node clustering coefficient C_i as follows,

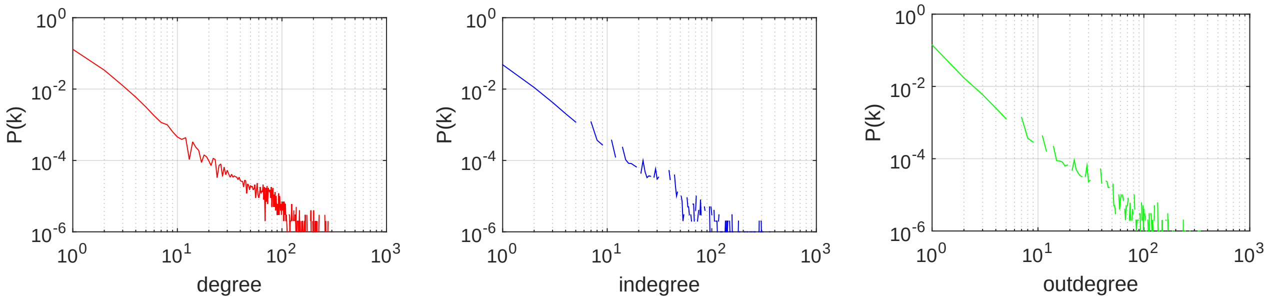
$$C_i = \frac{2\,|\Delta_i|}{k_i\,(k_i - 1)}. \tag{4}$$

We denote the average clustering coefficient of the Bitcoin network by C, which is given as follows,

$$C = \frac{1}{N} \sum_i \frac{|\Delta_i|}{k_i(k_i - 1)/2}, \tag{5}$$

where N is the number of nodes.

We calculate the clustering coefficient of every node and present the results in Fig. 2. The Bitcoin blockchain graph average clustering coefficient $C_B = 0.0071$. It is observed from Fig. 2 that the nodes with a high degree tend to have a



(a) Degree (b) Indegree (c) Outdegree

Fig. 1. Node degree distribution of the Bitcoin network

small clustering coefficient, implying that there are few links or connections among the neighbors of a high-degree node, especially for the nodes with more than 1,000 neighbors.

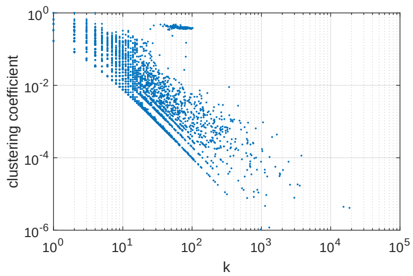


Fig. 2. Clustering coefficient of nodes

Meanwhile, the average shortest-path length measures the transportation efficiency in the complex network theory, which is defined as the average number of hops along the shortest path between any two nodes in the connected subgraph of the whole graph. Since the Bitcoin system is established on top of a peer-to-peer (P2P) network, any two nodes can directly trade with each other. In other words, the average shortest path should be about 2.0. However, according to recent studies [17], there are still some money laundering activities in the Bitcoin network. The transactions may not directly occur between any two parties in these malicious activities. So it is of great significance to investigate the path length for detecting the malicious behaviors. We denote the average shortest path length by L, which is expressed by $L = \sum_{i,j \in V(G)} \frac{l(i,j)}{N(N-1)}$, where $V(G)$ is the set of nodes in graph G and $l(i,j)$ is the shortest path from i to j. In the Bitcoin transaction graph, we denote the average shortest-path length of the connected subgraph by L_B, and we find that $L_B = 3.833$, which implies that there are many indirect transactions.

Small-world effect of the Bitcoin network. We observe that the Bitcoin network conforms to the small-world network model according to the average clustering coefficient and average shortest-path length. To find the evidence for this observation, it is necessary to compare the Bitcoin network with a random graph having the same degrees and edges as the Bitcoin network. Thus, we construct a random network with the same nodes and edges as the Bitcoin network. We denote the average clustering coefficient and the average shortest-path length of the random network by C_R and L_R, respectively. We have $C_R = 0.00026$ and $L_R = 9.00$. Comparing with the Bitcoin blockchain network, we have $C_B > C_R$ and $L_B < L_R$, indicating that the Bitcoin network is indeed a small-world network according to the network characteristics as in [16]. This effect implies that Bitcoin can be moved among the majority by a few steps.

C. Disassortativity

From the results of the clustering coefficient and degree distribution, we observe that nodes with higher degrees tend to have fewer connections. In order to further investigate the connection tendency of the Bitcoin network, we introduce *assortativity* analysis. The vast gap between the number of high-degree nodes and that of low-degree nodes indicates the high heterogeneity of the Bitcoin network. Thus, we adopt the Pearson correlation coefficient denoted by r to characterize the network assortativity [9]. The total number of links in the graph is denoted by M. We then have Pearson correlation coefficient r as follows,

$$r = \frac{M^{-1}\sum_{e_{ij}\in E(G)} k_i k_j - [M^{-1}\sum_{e_{ij}\in E(G)} 1/2(k_i + k_j)]^2}{M^{-1}\sum_{e_{ij}\in E(G)} 1/2(k_i^2 + k_j^2) - [M^{-1}\sum_{e_{ij}\in E(G)} 1/2(k_i + k_j)]^2}, \tag{6}$$

where k_i is the out-degree of the node at the beginning of link $e_{ij} \in E(G)$ and k_j is the in-degree of the node at the end of link $e_{ij} \in E(G)$.

If the Pearson correlation coefficient $r < 0$, it is disassortative; if $r = 0$, node degrees in the graph are uncorrelated on average. We find that the Pearson correlation coefficient $r = -0.023$ indicates the graph is disassortative. The disassortative effect means that high-degree nodes prefer to connect to low-degree nodes while low-degree nodes also prefer to connect to high-degree nodes.

However, the negative value of r cannot fully indicate the disassortativity. We then adopt the following measure to describe the average closest-neighbor node in-degree of node i that has out-degree k_i^{out},

$$k^{cn-in}\left(k^{out}\right) = \sum_{I=1,k_i^{out}=k^{out}}^{N} \left(\frac{k_i^{cn-out}}{N}\right) P\left(k^{out}\right), \tag{7}$$

where $k_i^{cn-out} = \sum_{j=1}^{N} a_{ij} k_j^{in}/k_i^{out}$, a_{ij} is the (i,j)-th entry of the adjacency matrix in Eq. (1), and $P\left(k^{out}\right)$ is the out-degree distribution function. If the value of $k^{cn-in}\left(k^{out}\right)$

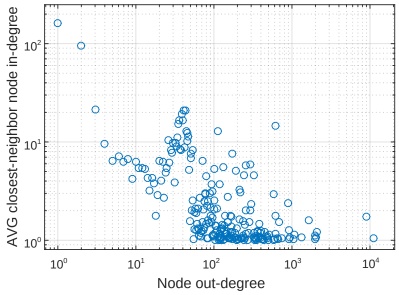


Fig. 3. Average in-degree of the closest neighbors as a function of node out-degree.

shows a downward trend with respect to the variable k^{out}, then the graph is disassortative, which is shown in Fig. 3. When the bitcoin network is disassortative, low-degree nodes show the connection preference of high-degree nodes. This effect can also be explained by the preferential attachment, i.e., newly-joined nodes prefer to connect to high-degree nodes [8].

D. Rich-club ordering

The Bitcoin blockchain network is disassortative, implying that low-degree nodes tend to connect with high-degree nodes. Meanwhile, it is also necessary to explore the connectivity between high-degree nodes. In a complex network, the *rich club* refers to the phenomenon of a tight connection between high-degree nodes. In other words, the higher-degree nodes are regarded as the rich nodes while the rich nodes in the network are more likely to gather into *clubs* (subgraphs) comparing with those nodes with fewer edges. We denote the rich-club coefficient $\phi(k)$, which is defined as follows,

$$\phi(k)=\frac{2E_{>k}}{N_{>k}\left(N_{>k}-1\right)}, \tag{8}$$

where $N_{>k}\left(N_{>k}-1\right)/2$ is the maximum possible edges of all $N_{>k}$ nodes whose degree is higher than k; similarly, $E_{>k}$ denotes the number of edges among $N_{>k}$ nodes. It is worth noting that the rich-club coefficient can be regarded as a more specific way than the assortativity coefficient since the rich-club coefficient focuses on the possibility of connection to a node over degree k. For example, a network with several rich nodes and some low-degree nodes exhibits disassortativity because the rich nodes are not directly connected; however, it still shows the rich-club phenomenon because the rich nodes are closely connected in the same subgraph.

We observe from Fig. 4 that the rich-club coefficient does not monotonically increase with the increment of k, implying no obvious rich-club phenomenon. However, in some discussions in [18], a single analysis of the rich-club coefficient cannot well reflect the rich-club effect of large networks. Thus, it is also necessary to compare it with a random network with the same degree distribution for a more accurate evaluation. Consequently, we adopt the normalized rich-club coefficient $\phi_{\text{norm}}(k)$ to evaluate the rich-club effect:

$$\phi_{\text{norm}}(k)=\frac{\phi(k)}{\phi_{\text{rand}}(k)}, \tag{9}$$

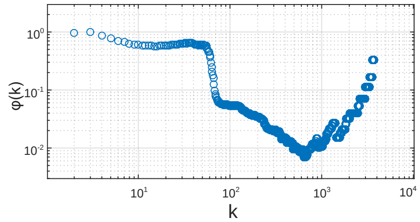


Fig. 4. Rich club coefficient.

where $\phi(k)$ is the rich-club coefficient of the original network and $\phi_{\text{rand}}(k)$ is the rich-club coefficient of a random network with the same degree distribution. Fig. 5 plots the results whereas the actual rich-club ordering depends on whether $\phi_{\text{norm}}(k)>1$. Fig. 5 shows that there is an absence of rich-club ordering on most k values, except for some discrete values. This phenomenon may be caused by the small random network rich-club coefficient with few k values, thereby leading to the small denominator. In general, the Bitcoin blockchain network exhibits the non-rich-club phenomenon, implying that the high-degree central nodes in this network tend to disconnect with each other and are distributed in different connective subgraphs. This effect can be explained by the fact that the central nodes are more likely to be some exchanges or large institutions which have their own relatively-fixed customer groups. Therefore, the rich nodes are not closely connected to each other.

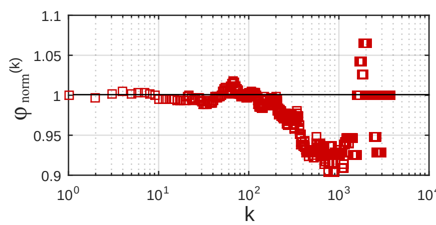


Fig. 5. The normalization of rich club coefficient.

V. Conclusion

In this paper, we have conducted a complex network analysis on the Bitcoin blockchain network. In particular, we have designed a novel sampling method, namely random walk with flying-back properties (RWFB), and got several important observations. Firstly, the degree distribution of the Bitcoin blockchain network conforms to a power-law distribution, approximated to a scale-free network. Secondly, the Bitcoin blockchain network is a small-world network after analyzing the average clustering coefficient and the shortest-path length. Thirdly, regarding to the disassortativity of the Bitcoin network, low-degree nodes prefer connecting to nodes with a higher degree. Meanwhile, we have also identified the preferential attachment of newly added nodes. Last but not least, the current Bitcoin blockchain network does not demonstrate the rich-club phenomenon. In other words, the central nodes are not closely connected in this graph. Such findings can help us better understand the structural behavior of cryptocurrency blockchain networks, and provide insights into the detection of fraud and malicious activities in the fintech era.

References

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

[2] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, pp. 352–375, 2018.

[3] C. Decker and R. Wattenhofer, "Information propagation in the bitcoin network," in *IEEE Thirteenth International Conference on Peer-to-peer Computing*, pp. 1–10, 2013.

[4] A. Baumann, B. Fabian, and M. Lischke, "Exploring the bitcoin network.," in *WEBIST (1)*, pp. 369–374, 2014.

[5] M. Lischke and B. Fabian, "Analyzing the bitcoin network: The first four years," *Future Internet*, vol. 8, no. 1, p. 7, 2016.

[6] D. Lin, J. Wu, Q. Yuan, and Z. Zheng, "Modeling and understanding ethereum transaction records via a complex network approach," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2020.

[7] M. A. Javarone and C. S. Wright, "From bitcoin to bitcoin cash: a network analysis," in *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, pp. 77–81, 2018.

[8] D. Kondor, M. Pósfai, I. Csabai, and G. Vattay, "Do the rich get richer? an empirical analysis of the bitcoin transaction network," *PloS one*, vol. 9, no. 2, p. e86197, 2014.

[9] G. Chen, X. Wang, and X. Li, *Fundamentals of complex networks: models, structures and dynamics*. John Wiley & Sons, 2014.

[10] D. Kondor, I. Csabai, J. Szüle, M. Pósfai, and G. Vattay, "Inferring the interplay between network structure and market effects in bitcoin," *New Journal of Physics*, vol. 16, no. 12, p. 125003, 2014.

[11] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, 2014.

[12] F. Spitzer, *Principles of random walk*, vol. 34. Springer Science & Business Media, 2013.

[13] S. Yoon, S. Lee, S.-H. Yook, and Y. Kim, "Statistical properties of sampled networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 73, no. 4 Pt 2, p. 046114, 2006.

[14] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 631–636, 2006.

[15] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[16] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[17] J. Wu, J. Liu, W. Chen, H. Huang, Z. Zheng, and Y. Zhang, "Detecting mixing services via mining bitcoin transaction network with hybrid motifs," *arXiv preprint arXiv:2001.05233*, 2020.

[18] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature physics*, vol. 2, no. 2, pp. 110–115, 2006.