Proceedings of Machine Learning Research vol 278:1–10, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Evolution of Bitcoin Trust Communities

Yuemin Cao 15618517686@163.com
The Hong Kong University of Science and Technology, Hongkong, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Bitcoin, a digital currency facilitated by blockchain technology, enables direct exchange and personal ownership of digital assets, verified through mathematical consensus. This paper explores and analyzes transaction data within the Bitcoin network, with a focus on improving the efficiency of entity recognition methods and identifying illegal transaction patterns. We begin by introducing Bitcoin's development background, underlying principles, and transaction processes. We then delve into the structure of Bitcoin transaction data and review recent literature on its analysis, summarizing key technologies and research directions. To address the inefficiencies of traditional heuristic entity recognition methods, we propose an innovative solution that establishes entity relationship sets and utilizes active address data. Our approach introduces specific advancements, including a novel algorithm designed to enhance network connectivity and stability, and a centrality aggregation index that outperforms traditional node centrality indices. This algorithm facilitates quick reconnection to previously successful peer nodes, discovers new nodes upon connection loss, and propagates node information across the network for more stable connections. Additionally, it employs a seed node mechanism to expedite network discovery. Our method leverages a core data structure that maintains a list of peers for initial connections, automated through a seed node process. This bootstrapping mechanism allows Bitcoin clients to efficiently connect to the entire Bitcoin network. For implementation and analysis, we utilize NetworkX, a Python package for manipulating and investigating complex networks. We visualize the network structure using the number of transactions or reviews as node size, average review sentiment as node color, review mistrust as edge length, and a force-directed algorithm for node positioning. Our results demonstrate that the first-order aggregation centrality index performs better than the node centrality index, confirming that incorporating more information about first-order correlation attributes around a node enhances the model's effectiveness. Our proposed model, integrating the centrality aggregation index, achieves a 1% improvement in precision, a 5% improvement in recall, and a 4% improvement in F1 score compared to the original feature set model. We define C as the node centrality feature set, C1 as the first-order aggregated feature set, C2 as the second-order aggregated feature set, and AF as the original feature set. From both model performance and visualization perspectives, the centrality aggregation index enables quick identification of key nodes and enhances the discovery of illegal transaction patterns in the network. By reversing and backtracking the capital flow path, our method can uncover more illegal transaction nodes and provide greater interpretability for the illegal transaction model. Finally, we discuss how to analyze and identify illegal behavior characteristics in Bitcoin transaction data, concluding with an examination of data sources, network construction, and analysis methods. By offering a comprehensive exploration of Bitcoin transaction data and advancing entity recognition methods, this paper provides valuable insights into the evolving landscape of cryptocurrency and blockchain technology. Our proposed innovations result in significant efficiency improvements and enhanced detection of illegal activities within the Bitcoin network.

Keywords: Bitcoin, network analysis, automated analysis and quantification, protocol development.

1. Introduction

With the uprise of information explosion and new technologies in the digital era, a tangible number of digital techniques have been adopted and transformed into automated frameworks for more convenient lifestyles. However, the ubiquitous presence of technologies also created prevalent issues regarding privacy when accessibility to personal information is not adequately protected. Such defects can cast hazardous risks on the users and cause prolonged damage to their autonomous status, especially during activities such as online transactions that concern personal owned property. Therefore, a trust system between individual users within different networks becomes an essential attribute that allows them to evaluate the safety of online activities besides the protection from organizational solutions.

Bitcoin, one of the most popular blockchain technologies, particularly relies on the apparatus of trust due to its decentralized nature where users are free from governmental regulations under cover anonymity. Despite the decentralization, there is a record attached to each user indicating their reputation/truthfulness on the scale from -10 to 10 rated by other users. Due to the uncertainties imposed on the exchanges, the rating system between different parties becomes a dominant factor that intervenes users' decision-making behavior (Jelenc et al., 2013). Moreover, relations behind such behavioral mechanisms are shown as interdependent within a multi-dimensional dynamic, consisting of communities detected based on truthfulness quantified as rating scores. With limited resources to confirm the safety of exchanges, identifying undiscovered trust communities can help Bitcoin users to strategize their decision during transactions in order to minimize cybersecurity risks. However, most of the currently available research has focused on building static models of trust communities in the Bitcoin network, leaving the temporal attribute, which contains information about community evolution, much less discussed. Therefore, considering the achievements and limitations of previous research on community detection in Bitcoin network regarding trust, we recognize that providing insights into underlying interaction and evolution of existing trust communities becomes essential for a more comprehensive understanding on how the value of trust transforms within different models and time intervals, which may contribute to predict the growth or decline of trust weight.

Since the birth of Bitcoin, after more than ten years of development, both the Bitcoin cryptocurrency itself and the blockchain technology it uses have received widespread attention. The point-to-point decentralization of Bitcoin attracts more and more people to join the Bitcoin network, so the transaction data of Bitcoin continues to increase, which has become a valuable resource for studying Bitcoin. People conduct transactions with the help of Bitcoin addresses that have nothing to do with their identity, so the Bitcoin network has a certain degree of anonymity and has become a paradise for many criminals to commit crimes, and all kinds of illegal transactions are full of them. But it is undeniable that Bitcoin transaction data contains great value, waiting for people to dig and explore.

The blockchain is still in a good period of development, but many problems have been exposed during its development, such as being used in illegal transactions such as money laundering and gambling. Bitcoin, as the earliest and most representative blockchain system, has endured the most controversy. Most of these disputes do not deny the technology itself, but stem from the lack of sufficient understanding and supervision of the Bitcoin network. As a research, Bitcoin de-anonymization is to understand and analyze the development status and evolution process of the Bitcoin network, and for the de-anonymization of specific account entities on the blockchain, the

identification work can also assist supervision and investigation and forensics. Therefore, it has high research value.

2. Literature Review

The blockchain first attracted the attention of the public with the rise of Bitcoin. In November 2008, Satoshi Nakamoto published a paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System" in Cypherpunk, expounding an electronic cash system framework of P2P network, encryption algorithm, time stamp, blockchain and other technologies. This design of writing transactions into blocks and concatenating them using cryptographic methods heralds the birth of blockchain technology.

To detect and analyze the trust communities within the Bitcoin network, it is essential to firstly define the notion of trust (Kumar et al., 2016a). Different trust models have been proposed in recent years, focusing on different domains and objectives. With the variant of trust being highly dependent on the context and individual's motive, Leppänen (2010) proposes the framework of technology trust, which is identified as an individual's confidence regarding a specific technology and the motives of the entity behind it. Such a trust framework is extended and compared with social trust between individuals rather than individual and technology in the research of Khairuddin and Sas under the context of Bitcoin exchanges. The above two research on trust models show that while the available information regarding truthfulness within the Bitcoin network is defined as trust between nodes representing individual users, it interacts with the users' trust in the accessibility, security, and productivity of the network itself. Therefore, during the analysis of the trust rating system of Bitcoin, the usability of the network and users' perception of it should also be taken into consideration along with trust variants within communities, as the network mediates the interactions between users.

In regards to community detection in the Bitcoin trust system, Marina and Punceva have found that there are observable community structures with good modularity in terms of trust as rating scores, which also provides insightful information on highly trusted communities and possible strategies during decision making for Bitcoin users (Punčeva et al., 2018). Additionally, in earlier research conducted by B¨unz and Yu in 2015, it is indicated that the size of trust communities formed within the Bitcoin network is correlated to the preexisting Bitcoin communities in real life, through identity attachment. This finding broadens the understanding of trust mechanisms behind communities within the Bitcoin network in terms of potential factors that affect the formation of trust between users and how it is aggregated.

As the existence of trust communities in the Bitcoin network has been confirmed and thoroughly analyzed in the precious research, it is shown that such research interests possess great potential which can be applied in future development of the Bitcoin network regarding cybersecurity and human-computer interaction. Nonetheless, the temporal attribute of communities' formation and evolution is much less represented in the academic field than the static models. Thus, this research is intended to bridge the gap between previous findings and potential application of community detection within the context of user reputation/truthfulness in the Bitcoin network, exploring the pattern of trust community throughout different time intervals.

To better de-anonymize Bitcoin, Jeffrey combs and analyzes the Bitcoin network from two perspectives: macroscopic network characterization and microscopic address entity identification. They analyzed the characteristics of Bitcoin transaction network from the network perspective. Us-

ing the indicators of the complex network as the starting point, they studied the basic indicators of the Bitcoin network. After multiple assumptions, analysis, and verification point of view, they found that the correlation between the transaction volume and the account capital amount in the Bitcoin network remains stable and the retention rate of the savings account has remained in a low range for a long time after 2012, and most of the early savings account users are still in the current ecosystem. Also, they propose a method for entity recognition and classification using association features. The current mainstream entity recognition and classification method in Bitcoin transaction network is to use network graph structure features. Therefore, they took the multi-transaction association characteristics of Bitcoin address itself as an entry point.

Alqassem crawled the label information from the public labeling website to construct the corresponding dataset and used the experiment to prove that their proposed method can effectively improve the performance of entity recognition (Alqassem et al., 2020). On this basis, they also analyzed the importance of features and the correlation of classification. The feature analysis and entity identification system of Bitcoin network with abnormal alarm function can be seamlessly migrated to other Bitcoin-like networks to serve network analysis and characterization. From the results, the proposed recognition method can improve the performance of entity recognition, and the analysis work further deepened the understanding of the Bitcoin network (Sas and Khairuddin, 2015). The realized analysis and identification system can complete the output of main indicators without manual intervention after basic configuration and can be interrupted and restored in time. The system is also of great help to the research on Bitcoin de-anonymity and to reduce the debugging configuration in the analysis of this large amount of data (Kumar et al., 2016b).

3. Methodology

3.1. Algorithm Design

The quantitative algorithm developed in this study follows a structured procedure to enhance network connectivity and stability. The core logic and parameter selection are detailed as follows:

(1) Node Reconnection: Each node maintains a record of recently successful connections. Upon restart, the node quickly reconnects to the previously established network of peer nodes. This mechanism is parameterized by a timeout value (T_reconnect), which determines how long the node waits before attempting reconnection.

(2) New Node Discovery: If a node loses its existing connection, it initiates a discovery process to find new nodes. This process is governed by a discovery interval (T_discovery), which specifies the frequency of discovery attempts, and a maximum number of attempts (N_attempts) to prevent infinite loops.

(3) Message Propagation: Once one or more connections are established, the node sends a message containing its address to neighboring nodes. These adjacent nodes forward the message to their respective neighbors, ensuring that the node information is propagated throughout the network. The propagation depth (D_propagation) controls how far the message travels, balancing network coverage and message overhead.

(4) Peer List Request: A newly accessed node can send a "get-address" message to its neighbors, requesting a list of IP addresses of known peers. This list is filtered based on a trust threshold (T_trust) to exclude potentially malicious nodes. The node then selects peers from this list based on a connection limit (N_connections) to manage resource usage.

(5) Seed Node Assignment: Upon startup, if an active node is not assigned, the client maintains a list of stable, long-running nodes known as seed nodes. These seed nodes facilitate quick discovery of other nodes in the network. The seed node list is updated periodically based on a refresh interval (T_refresh) to ensure its accuracy and reliability.

3.2. Core data structure

The core part of Bitcoin maintains a list of peers to which it can connect at startup. When a full node starts up for the first time, it must be bootstrapped to the network. This process is now automated at the heart of Bitcoin through a short list of seeds. The option "-seednode" can be used to define this behavior. From there, Bitcoin clients can connect to the entire Bitcoin network. The method of bootstrapping involves using the "-addnode" parameter, which allows the user to predefine which server to connect to and disconnect after the peer list is established.

3.3. Implementation and analysis

The implementation of this technique is primarily based on NetworkX, a Python package for data manipulation and investigation of the network structure, dynamics, and functions of complex networks. To better visualize the network structure, we used the number of transactions or reviews as the node size, the average positivity or negativity of nodes' reviews as node color, the mistrust of a single review as edge length, and a force-directed algorithm to determine node positions. This visualization approach enhances the interpretability of the network structure and facilitates the identification of key nodes and illegal transaction patterns.

4. Results

4.1. Dataset Description

In the selected dataset, this reconstructed network has 5881 nodes and 35592 edges among which 89% are positive (Table 1). The original dataset has SOURCE, which refers to node id of source, TARGET, which refers to node id of target, RATING, the source's rating for the target, whose range is from -10 to +10 in steps of 1 and TIME, which is measured as seconds (Table 2).

Table 1: Ratings per month

rating time	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
1/31/2011	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	65	20	8	4	1	NaN	NaN	NaN	NaN	1
8/31/2015	7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2	34	19	8	1	1	2	5	1	NaN	NaN
7/31/2015	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	44	10	4	NaN	4	NaN	NaN	NaN	NaN	2
2/28/2014	52	NaN	NaN	1	NaN	1	1	1	15	32	243	77	38	19	20	4	7	4	NaN	11
12/31/2011	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3	273	72	28	15	19	NaN	4	4	2	10

4.2. Network Visualization

Here, we used multi-edged directed graph to visualize trust ratings. It is noted that larger nodes have more reviews and green nodes in the figure above are positively reviewed on average and red nodes are negatively reviewed on average (Figure 1). Also, the color intensity is degree of trust, which means positive and negative rating. We used simple kernel function to identify the correlation

Table 2: Years with a broader range of values

time	source	target	rating
2014-03-11 00:16:41.623600128	5457	1850	1
2014-03-20 20:22:39.795640064	5287	988	4
2014-03-31 15:22:44.090899968	5479	5363	-10
2014-03-20 17:14:21.703510016	3837	4304	1
2014-03-03 14:50:00.187429888	4993	4540	5
2014-03-24 15:39:50.654170112	5363	2987	-10
2014-03-04 10:16:54.244169984	5383	3988	1
2014-03-20 19:38:55.874870016	1746	2028	-10
2014-03-17 18:47:33.611880192	2486	1018	1
2014-03-21 18:31:16.692719872	5028	3630	2

between multiple variables and identify the most important factors from the density plots (Figure 2).

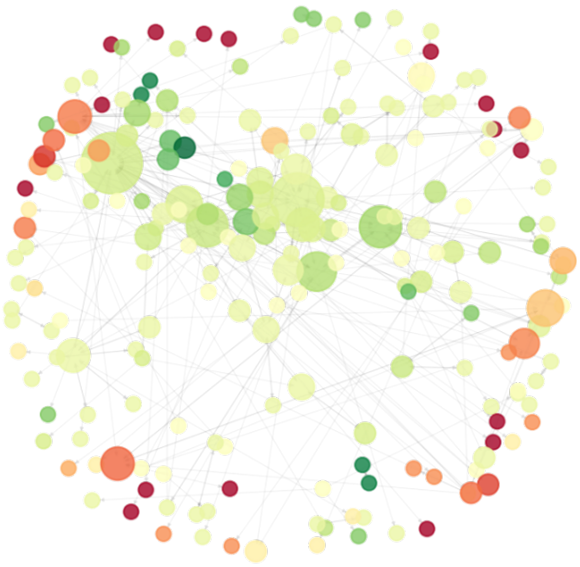

Figure 1: Scatter plot and density plots of multiple variables in the raw dataset

4.3. Quantification and Statistical Analysis

From the figure above, we can see larger node appear to cluster together as expected and there's less likelihood to maintain a high average rating with more ratings (Figure 3). Also, highly negatively rated nodes appear to be pushed to the perimeter and away from positive transaction clusters. Finally, it is also noted that nodes marked in red indicate consistently negatively ratings.

From multiple figures above, we are able conclude that the performance of the first-order aggregation centrality index is better than that of the node centrality index, which proves that the more information about the first-order correlation attributes around the node is obtained, the better the en-

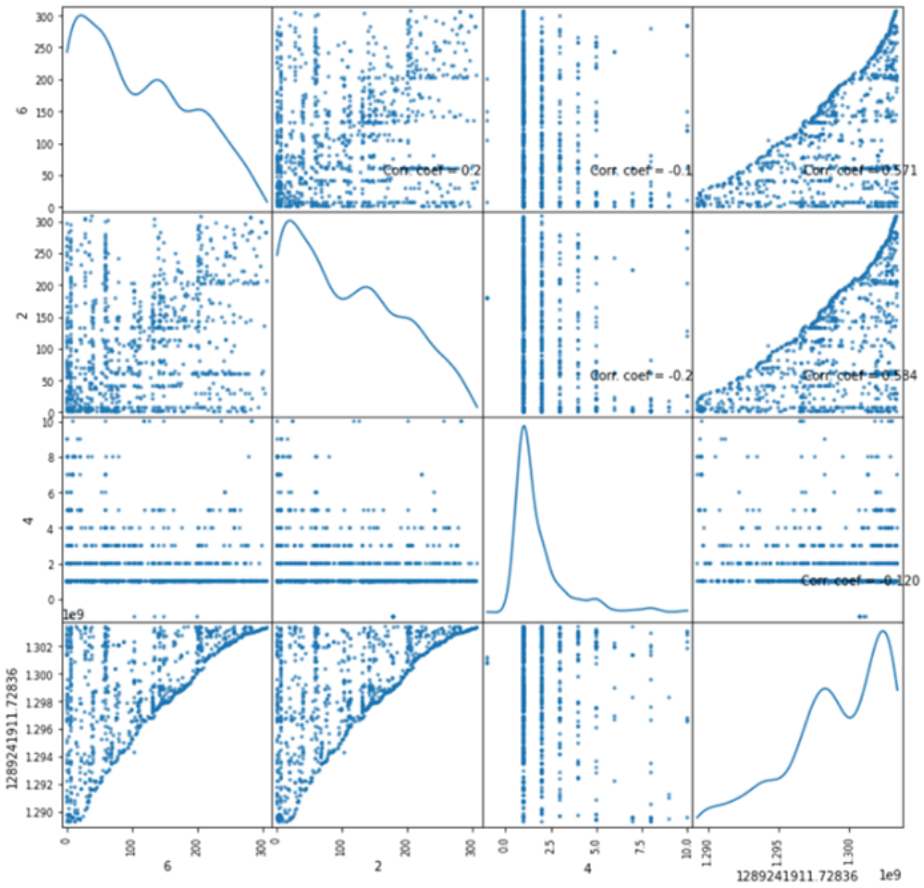

Figure 2: Bitcoin Trust Community Network (sampling on the monthly basis)

hancement effect of the model is (Figure 4). The model integrates the centrality aggregation index, has a 1% improvement in precision compared to the original feature set model, a 5% improvement in recall, and a 4% improvement in F1. Among them, C is the node centrality feature set, C1 is the first-order aggregated feature set, C2 is the second-order aggregated feature set, and AF is the original feature set (Figure 5). From the model perspective, the centrality aggregation index has an improved effect, and from the visualization perspective, the centrality feature index can quickly find key nodes. The network transaction structure discovery can find illegal transaction patterns in the network and can reverse the case and backtrack. The capital flow path can discover more illegal transaction nodes and provide more interpretability for the illegal transaction model (Figure 6).

5. Discussion

Bitcoin is not officially recognized by the government, and its liquidity is highly fragmented and vulnerable to manipulation. A typical example in this regard is, Bitstamp's bitcoin price occupies 50% of the price weight of Bitmex, the world's largest contract exchange, which leads to a series of liquidation of long positions on Bitmex, and the forced liquidation continues to fall off a cliff,

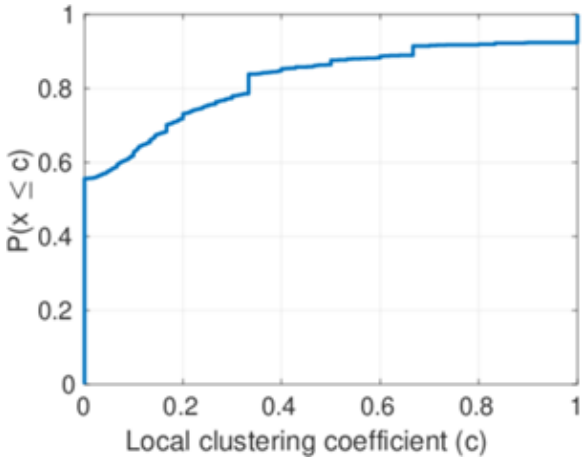

Figure 3: Local clustering coefficient analysis of constructed network

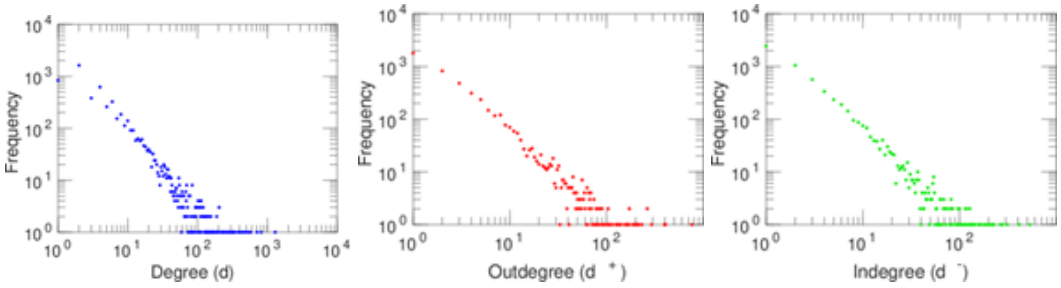

Figure 4: Longitudinal study of frequency vs. degree in the original dataset collected at different time periods

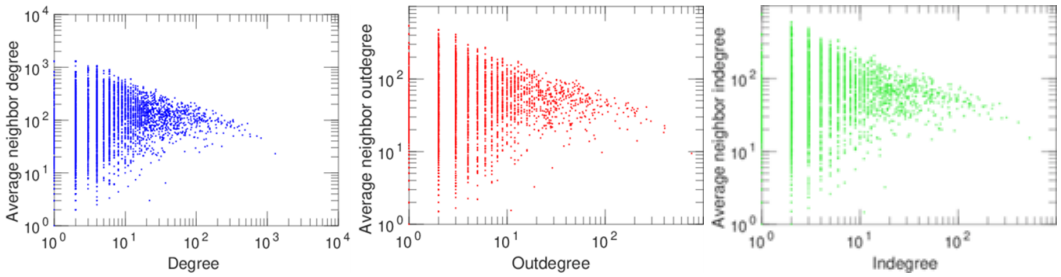

Figure 5: Longitudinal study of average neighbor degree vs. degree in the original dataset collected at different time periods

resulting in At that time, the price of Bitcoin fell by more than 20%. From this point of view, Bitcoin still has a long way to go to become a safe-haven asset.

To understand how social and anti-social trends in Bitcoin's user base affect its development, it is necessary to analyze the Bitcoin system as a network. We explored the local topology and geometry

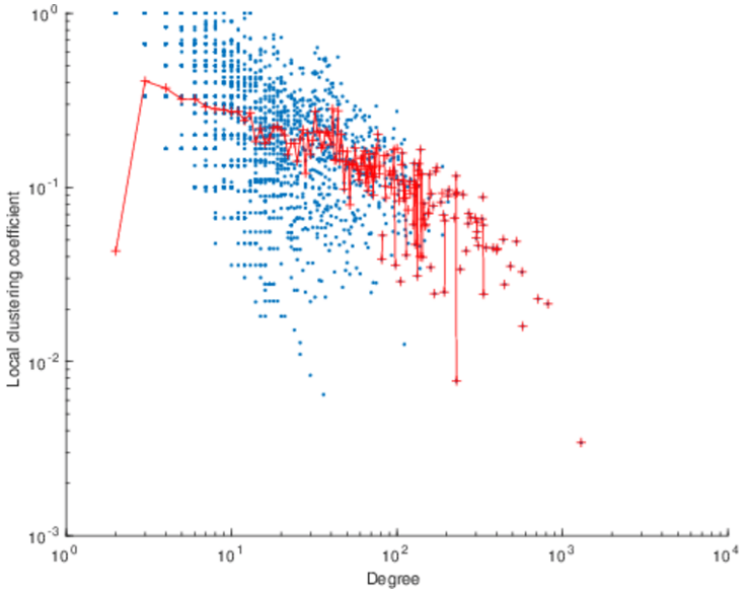

Figure 6: Local clustering coefficient with respect to degree (red: centrality aggregation index)

of the Bitcoin network. To this end, Bitcoin transaction data was processed to build a Bitcoin user graph. Characteristics of user graphs, local and global network properties were analyzed at intervals (Simser, 2015). Small diameters, skewed distributions of transactions, power-law in and out degree distributions, discontinuous graphs, and network analysis indicate the presence of largeconnected components. Therefore, it can be inferred that despite its antisocial tendencies, the Bitcoin network shares similarities with other complex networks (Nelms et al., 2017).

To conclude, after establishing a connection with the peer node, the two parties send each other a message containing the latest block hash value. If a node believes that it has the latest block information or has a longer chain, it will send an inv message containing the hashes of up to 500 latest blocks, indicating that it has a longer chain. The receiving node will request block details, and the remote node sends this information with the command block. After 500 blocks of information have been processed, nodes can request more block information. These blocks are confirmed after being authenticated by the receiving node. Confirmation of new blocks can also be found by miners mining and publishing blocks. Through the previous connection, new blocks are published, and the receiving node can request the details of these blocks.

References

Israa Alqassem, Iyad Rahwan, and Davor Svetinovic. The anti-social system properties: Bitcoin network data analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(1): 21–31, 2020. doi: 10.1109/TSMC.2018.2883678.

David Jelenc, Ramón Hermoso, Jordi Sabater-Mir, and Denis Trček. Decision making matters: A better way to evaluate trust models. *Knowledge-Based Systems*, 52:147–164, 2013. doi: https://doi.org/10.1016/j.knosys.2013.07.016.

Srijan Kumar, Francesca Spezzano, V. S. Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 221–230, 2016a. doi: 10.1109/ICDM.2016.0033.

Srijan Kumar, Francesca Spezzano, V. S. Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 221–230, 2016b. doi: 10.1109/ICDM.2016.0033.

Akseli Leppänen. Technology trust antecedents: Building the platform for technology-enabled performance. 2010.

Taylor Nelms, Bill Maurer, Lana Swartz, and Scott Mainwaring. Social payments: Innovation, trust, bitcoin, and the sharing economy. *Theory, Culture & Society*, 35:026327641774646, 12 2017. doi: 10.1177/0263276417746466.

Magdalena Punčeva, Haute École Arc, and Ninoslav Marina. Bitcoin trust communities. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 1–6, 2018. doi: 10.1109/ICUMT.2018.8631251.

Corina Sas and Irni Eliana Khairuddin. Exploring trust in bitcoin technology: A framework for hci research. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, page 338–342. Association for Computing Machinery, 2015. doi: 10.1145/2838739.2838821.

Jeffrey Simser. Bitcoin and modern alchemy: In code we trust. *Journal of Financial Crime*, 22: 156–169, 05 2015. doi: 10.1108/JFC-11-2013-0067.