A Centrality-based Analysis of the Bitcoin Transaction Network

Arianna Groetsema Margo Kersey

August 19, 2019

Abstract

The Bitcoin network presents a unique opportunity to explore an anonymized worldwide payment system where each transaction is publicly available and accessible. Using Bitcoin transaction data from 2011, we will examine how this network propagates and what makes this particular network unique. In this paper, we will delve into the characteristics of different transactions in order to identify the main actors within the network and the correlations between the tides of public opinion and the classes of users propagating the network. In 2011, Bitcoin saw a drastic change in the number of users and the value Bitcoin itself. We would like to identify the key players in the network and evaluate how their influence changes over time. Touted as a truly decentralized network, we will determine the validity of this claim and the realities surrounding the main actors within this system in order to quantify how amenable this currency system is to the machinations of large entities. Using these metrics, we want to determine if the Bitcoin network can sustain itself as a lasting stable currency.

1 Introduction

Developed and released by an anonymous entity operating under the name Satoshi Nakamoto in 2009, the Bitcoin blockchain is a decentralized digital cryptocurrency. Bitcoin is a currency system isolated from government and centralized banks' control. The Bitcoin network is a peer to peer system that relies on participating nodes to generate currency and maintain the network. Transactions between users in the network are verified and recorded on the blockchain, a public distributed ledger. As such, an entire log of every transaction is stored and available for public access. While users' identities stay anonymous, each has a public address that is recorded along with each transaction.

Over the past few years, Bitcoin along with many other cryptocurrencies have garnered widespread media attention around the world resulting in their revolutionary ability to provide a method of trade divorced from potential corruption and manipulation by governments as well as protection from particular uncertainties. The blockchain architecture underlying this technology has provided a new wave of solutions for a variety of problems from supply chains, authentications, manufacturing, and identity management.

Some aspects of the network we would like to examine are the number of transactions per user, the distribution of these transactions, taking into account the volume of transactions and the quantity exchanged per transaction, as well as the most central nodes. We would like to analyze whether the centrality of nodes will tell us if there exists influential individuals or subgroups within the network of users. This would hopefully give us insight about whether the currency will be sustainable long term. We propose that the main actors in the network will be involved in transactions with the highest movement of Bitcoin. If we can predict that any given entity is increasing in influence over time, then we can say that the network may susceptible to a 51% attack. This is when a majority of the network is controlled by an entity, giving that entity power of consensus (verifying transactions).

The goal of this paper is to determine the distribution of transactions amongst actors in the network, and therefore observe changes in actors' centrality measures over time to analyze the influence of individual actors in the network. Specifically, we will break down the infamous 2011 spike and fall from Bitcoin history into four time periods, and interpret the centralities we see in each period. We will be measuring and analyzing in-degree, out-degree, and betweenness centrality because we believe these measures have the most relevant meaning in the context of the network.

2 Methods

2.1 Data

Bitcoin transaction data is readily available for public access through various sources. We downloaded the Bitcoin Transaction Network Dataset published in 2017 by Micheal Fire and Carlos Guestrin [1]. This set contains over 16 million Bitcoin transactions spanning 2009 to 2013.

Each edge of the network is a transaction marked with the resolution time of the transaction, and nodes are public key addresses representing an actor in the network. This is a directed, unweighted network, with timestamps for every transaction. Since this network contains record of every Bitcoin transaction during this period, this was the ideal data set to use in order to obtain an accurate view of the structure of the network.

2.2 Matlab

The network has more than 16 million edges and 6 million nodes, and the transactions were not in any kind of recognizable order, so we planned to filter through the network to extract a 20 day period of transaction data. After many failed attempts, we ended up breaking down the file into 161 separate `.csv` files each with 100,000 rows to make the computations more manageable. For each file, we converted the rows of transactions to a `timetable` in order to extract the period we were interested in, and concatenated the data into one large timetable organized by ascending date. From here, we were better equipped to work with the data, and began dividing it up and calculating centrality measures. We created four, five-day periods such that we could capture the effects of the rise and fall of Bitcoin value and transaction volume on the network. By analyzing the network from 5/27/2011 to 6/16/2011, we could see how the node centrality was affected by the surge of growth, and what this might say about any single entity's ability to gain power in a network of this kind.

We created a separate `digraph` for each period, and calculated the out-degree, in-degree, and between-

Figure 1: *Value of Bitcoin (blue) and Volume of Transactions (red and green) Over Time. Chart modified from `www.finance.yahoo.com`*

ness centralities of each node using a Matlab centrality function of the form `C = centrality(G,type)`. We were able to use the same function to calculate all the types of centrality that we were interested in. We decided to compile the top ten nodes with the highest centrality in each time period for each type of centrality. Next, we normalized the centrality values by dividing each value by the total number of transactions in their respective periods. Since centrality measures are only relative to the other nodes in the group, normalizing the values allows us to make conclusions about how a node's centrality value changes over time.

2.3 Centrality Measures

We strategically chose to measure in-degree, out-degree, and betweenness centrality for several reasons. Among them, we sought to determine the direction of the flow of currency within the network. A significant indicator of this flow is *Degree Centrality*. Degree centrality measures the number of links between nodes and can identify important hubs within the network. As each edge is a transaction between nodes, degree centrality gives an idea of where Bitcoin is flowing. In the Bitcoin network, in-degree and out-degree centrality give distinct depictions of the network. In-degree distributions signify the amount of Bitcoin a node receives, while out-degree distribution represents the amount of Bitcoin a node sends to others. Nodes with higher in-degree and out-degree centrality are more central in the network and tend to have more influence within the network as many of the transactions involved their participation. To compare the centrality of nodes throughout the various time ranges, we normalized the in-degree and out-degree centrality measures by dividing by the total number of edges, or the total number of transactions that occur within the time frame. Thus, the centrality of a node is given by:

$$x_i = \frac{k_i^{in}}{m} = \frac{k_i^{in}}{\sum_{i=1}^{n} k_i^{in}} \qquad \text{(In-Degree Centrality Normalized)}$$

$$x_i = \frac{k_i^{out}}{m} = \frac{k_i^{out}}{\sum_{i=1}^{n} k_i^{out}} \qquad \text{(Out-Degree Centrality Normalized)}$$

where k_i^{in} is the number of incoming edges to node i, k_i^{out} is the number of outgoing edges from node i, and m is the total number of edges in the network. This in effect gives the proportion of the transactions a node engages in compared to the total number of transactions that occur within the network (Table 1). As a result, we are able to compare how the centrality of a node changes over time or how the number of transactions a node participates in evolves.

However these measures alone are not sufficient to determine the overall influence of a single node. Consequently, we introduce *Betweenness Centrality*. By definition, betweenness centrality is the number of shortest paths between nodes that pass through a given node [2]. The equation for the centrality of node i is given by:

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}} \qquad \text{(Betweenness Centrality)}$$

where n_{st}^i is the number of shortest paths from s to t that pass through node i, and g_{st} is the total number of shortest paths from s to t. Again, normalizing these values by dividing by the total number of transactions in the respective periods, we obtain:

$$x_i = \frac{\sum_{st} \frac{n_{st}^i}{g_{st}}}{m} \qquad \text{(Betweenness Centrality Normalized)}$$

Note that dividing by total number of in-going edges produces the same result as dividing by the total number of outgoing edges, since every in-going edge is also outgoing edge from the perspective of the other node.

This value is important because it is likely able to measure a node's influence within a network due to the control over the flow of information in a network, under the assumption that information primarily flows over the shortest paths between nodes [2]. In the case of the Bitcoin network, this means a node with high betweenness centrality relative to the rest of the network plays a more central role in the flow of transactions between actors. So, nodes with higher centrality act as brokers in the network and have greater influence over the transactions that occur.

Table 1: *In-Degree Centrality for Top 10 Nodes in Period 1*

Node ID	Raw Score	Normalized Score
'5104'	515	0.009934030323
'224319'	511	0.009856872806
'645014'	397	0.007657883569
'4066244'	244	0.004706608541
'4365385'	239	0.004610161645
'1396811'	211	0.004070059026
'1252404'	210	0.004050769646
'1379778'	209	0.004031480267
'741836'	206	0.003973612129
'741282'	206	0.003973612129

Other centrality measures we initally considered calculating included closeness centrality and eigenvector centrality, both of which provide different interpretations of the network dynamics. In the context of the Bitcoin transaction network, closeness centrality considers the distance of a node to all other nodes and gives a measure of how long it takes for information, or in this case Bitcoin, to travel from one node to other nodes in the network. As a result, closeness centrality can give an interpretation of how efficiently Bitcoin flows within the network. As we were more interested in identifying key players who had the most impact on the transactions that occur, we decided against analyzing the network using closeness centrality. Additionally, eigenvector centrality measures the influence of one node on another node, which could be another useful measure to study in future analyses. However, due to its high computing time we forwent examining it.

Table 2: *Change in the Number of Actors & Transactions in the Bitcoin Network*

Date	Nodes	Edges
May 2011	4298	5402
June 2011	67706	119558

3 Results

3.1 Summary of Findings

Through our examination of the Bitcoin network over a relatively short period of time, we were able to observe how the network reacted to a major change. The June 2011 surge in price and popularity of Bitcoin resulted in an upheaval of the network and the introduction of many more users. This surge is demonstrated by the dramatic increase in the number of nodes and edges (Table 2). Identifying how the network changed as a result of this was an important step in drawing conclusions on how much influence individual nodes have on the network.

While it is difficult to draw lasting conclusions about the ability for an entity to gain influence in the Bitcoin network, we produced several thought-provoking results regarding the top most central nodes. By extracting the top ten most central nodes and examining how their centrality changed over time, we saw the tendency for centrality to peak during time periods two and three (Figure 3 & 4), which corresponds to the peak in Bitcoin value. We observed that one node in particular achieved consistently high centrality rankings through all three centrality measures and through every single time period. In fact this node, Node 5104, accumulated more centrality during the 4 periods in comparison to other nodes. As discussed earlier, betweenness centrality provides a good basis for determining the influence of an actor in the Bitcoin network. As such, Node 5104's growth in betweenness centrality which accelerated during the peak in the price of Bitcoin is indicative of an imbalance of influence in the network.

3.2 Out-Degree

Interestingly, for out-degree only two nodes (Node 5104 and Node 645014) appeared in the top ten list for more than one time period. In fact, these two nodes appeared at the very top in all four time periods. Their centralities grew larger, becoming less rivaled by other nodes as the spike occurred in periods two and three, while the values of the other eight nodes became even became smaller during these two periods. As the peak fell back down in period 4, values became closer together among these ten nodes, and the values of the nodes besides Node 5104 and Node 645014 grew larger. The result in periods two and three demonstrates the "rich get richer effect," as described in other currency network analysis papers. Specifically, Kondor et al. found a positive correlation between wealth and degree of a node [3]. Realistically, we would need to track node centrality over a much longer period of time to get an idea of this phenomenon, however even in the short period of time we analyzed we see that nodes with higher degree grow more drastically than nodes with lower degree. Due to the positive correlation mentioned by Konder et al., this corresponds to "rich" nodes getting richer.

3.3 In-Degree

Computing the distribution of in-degree centrality resulted in values that did not vary as significantly as out-degree centrality values, nor did the top few nodes become as further separated from the rest of the top ten nodes. Rather, the distribution of the node centrality increased relatively moderately. One conclusion we could draw is that nodes within the top ten in-degree centrality distributions, aside from Node 5104 and Node 645014, varied heavily from time period to time period. This high variability in the ranking of the most central nodes led us to conclude that there are fewer hubs of actors that are on the receiving end of transactions.

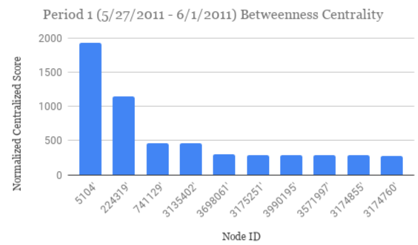


Figure 2: *Betweenness Centrality Top 10 Nodes for Period 1*

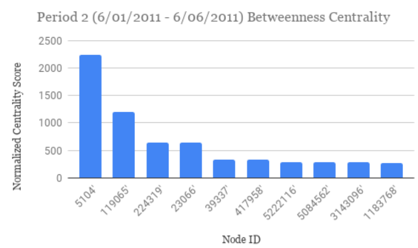


Figure 3: *Betweenness Centrality Top 10 Nodes for Period 2*

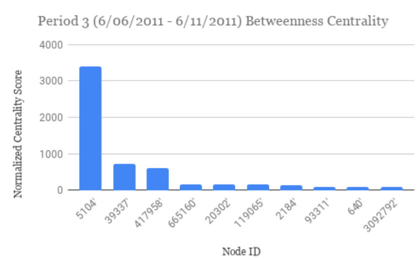


Figure 4: *Betweenness Ranking Top 10 Nodes for Period 3*

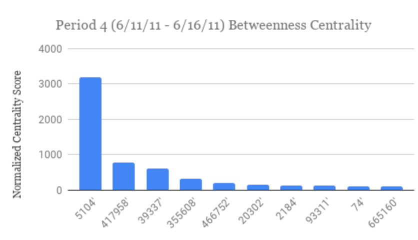


Figure 5: *Betweenness Centrality Top 10 Nodes for Period 4*

3.4 Betweenness

Next, analyzing the change in betweenness centralities, we found Node 5104 to consistenly be the most central node in each time period. It is important to note that this node was also the highest ranking in all other centralities and time periods except for the fourth period for out-degree centrality in which it was the second most central node.

While this node was consistently central, it reached peak betweenness centrality in time period and 3 (Figure 6c). Interestingly, the distribution of betweenness centrality became more polarized over time, meaning that its centrality grew much more in comparison to other nodes in the network (Figure 2, 3, 4, & 4). The greater growth in centrality of Node 5104 compared to other nodes indicates that high flux within the Bitcoin network can result in a greater influence imbalance.

4 Conclusion

Through the course of this project, we examined the dynamics of the Bitcoin Transaction network during a period of high fluctuation. As the first dramatic increase in price since Bitcoin's conception, analyzing how the network responded can give valuable insight into how the network will evolve and develop in the future.

We used degree and betweenness centrality measures to obtain quantitative data to analyze the relationships between users in the network, and determine which nodes gained influence during the spike. As a result, we were able to show it is likely that nodes are able to gain influence when a surge in Bitcoin occurs. We observed larger jumps in value of out-degree and betweenness centralities of the top ten values. In other words, highly central nodes increase in centrality over time while less central nodes decrease in centrality. This analysis serves to confirm the rich get richer phenomenon. The distribution of in-degree centrality demonstrates that the top few actors in the network are the only nodes that remain a hub during such times. When Bitcoin price and number of Bitcoin users rapidly increased, we observed that the number of hubs decreased.

While there are nodes with high centrality in comparison to the vast majority of other nodes, the influence they realistically have over the network is quite small when you consider the network as a whole. Despite seeing Node 5104 harness extreme centrality relative to other nodes, as the network grows there is such a large volume of other nodes with lower centralities that it outweighs the change in centrality of the few top ranked nodes. In the long run, nodes with increasing centrality will not surpass a certain threshold of influence over the network, considering that increases in centrality only occur when more users join the network. We claim that Bitcoin currency can be sustainable as global currency, given that the growth continues in the same manner over the years to come. It is extremely unlikely that any single node will ever be capable of committing a 51% attack given the analysis of our results.

References

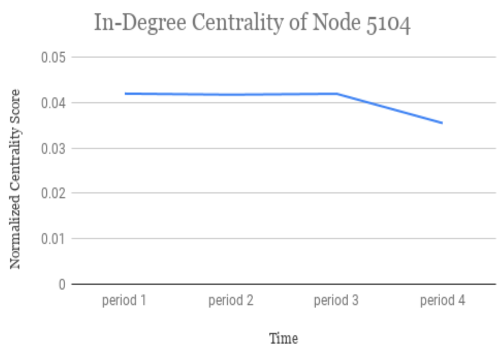
[1] M. Fire and C. Guestrin, "The rise and fall of network stars," *CoRR*, vol. abs/1706.06690, 2017. [Online]. Available: http://arxiv.org/abs/1706.06690

[2] M. Newman, *Networks, Second Edition.* Oxford University Press, 2018.

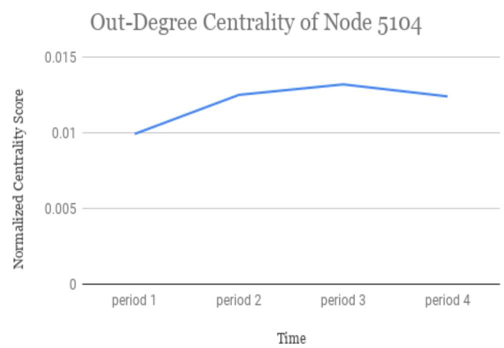
[3] D. Kondor, M. Pósfai, I. Csabai, and G. Vattay, "Do the rich get richer? an empirical analysis of the bitcoin transaction network," *PLOS One*, vol. 9, no. 5, 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0097205

A Appendix

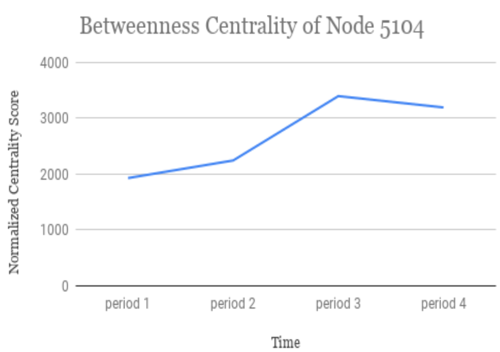
A.1 Additional Graphs



(a)

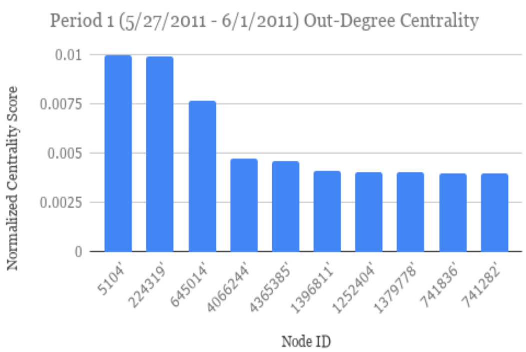


(b)

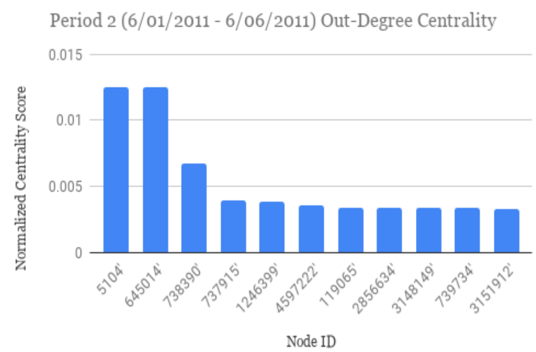


(c)

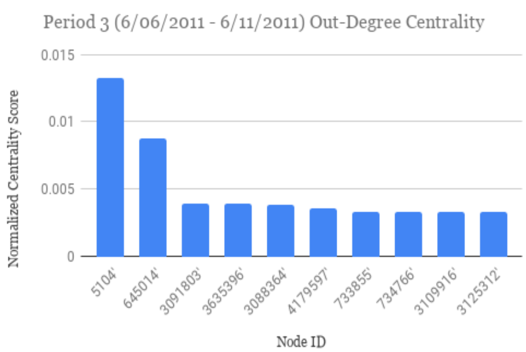
Figure 6: *Centrality of Node 5104 Over Time*



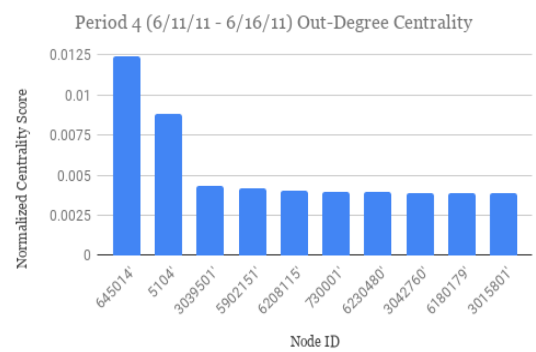
(a)



(b)

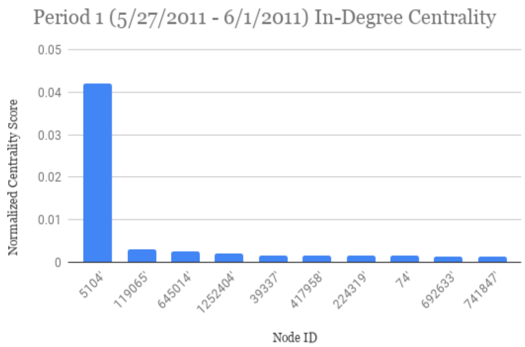


(c)

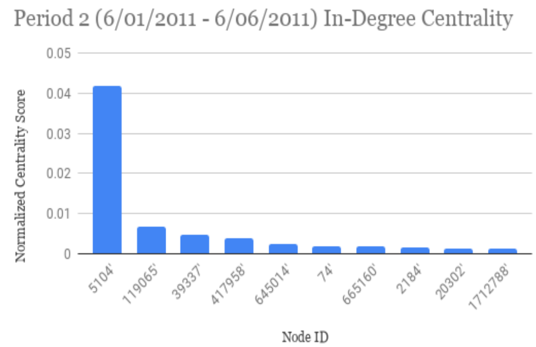


(d)

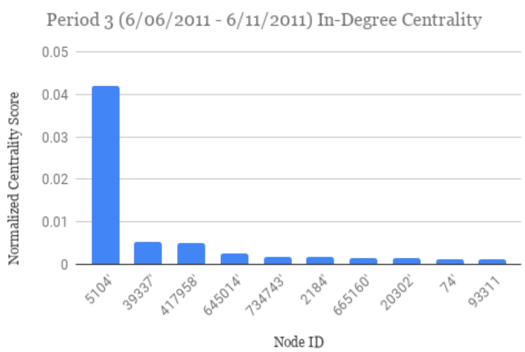
Figure 7: *Out-Degree Centrality of the Top 10 Nodes for Each Time Period*



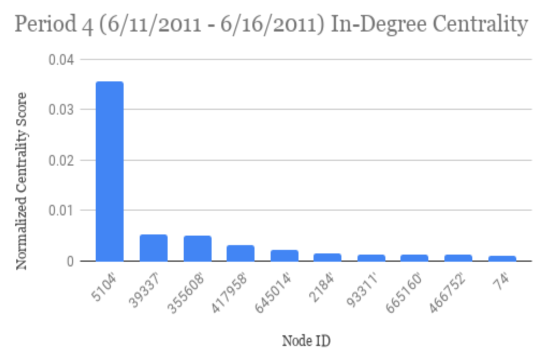
(a)



(b)



(c)



(d)

Figure 8: *In-Degree Centrality of the Top Ten Nodes for Each Time Period*

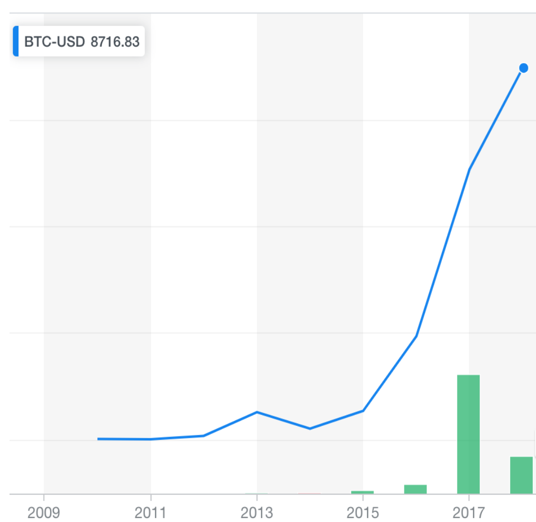


Figure 9: *Value of Bitcoin (blue) and Volume of Transactions (green) Over the Years. Chart from* `www.finance.yahoo.com`