COSC 421

THE LIGHTNING NETWORK

THE POTENTIAL BITCOIN SCALABILITY SOLUTION

Zack Dupont

Vic Dhand

Julius Wu

Susheel Palakurthi

Table of Contents

Introduction

The bitcoin lightning network was proposed in 2015 by Joseph Poon and Thaddeus Dryja as a potential solution to the bitcoin scalability problem that arises due to the limited transaction capacity of the bitcoin blockchain[1]. It is often argued that the lightning network can be completely distributed; however, the network displays typical tendencies of a decentralized network due to the presence of banking hubs[2]. The goal of this project was to analyze the bitcoin lightning network data in order to determine if these banking hubs do in fact exist in the network. To achieve this goal, the data was processed using a range of methods, the most significant being Microsoft Excel, the R programming language, and different R methods available through a variety of packages.

Definitions

Bitcoin: form of electronically stored currency considered to be a major influence in the future of finance.

Block: component of the blockchain. Transactions are recorded in blocks, and subsequently added to the end of the blockchain.

Blockchain: a digital ledger, or simply a database, containing records of all transactions that have occured on the network. It is a secure and public record that cannot be corrupted.

Block height: the number of blocks that currently make up the blockchain.

Hub: central node through which many transaction in the network pass.

Lightning Network (LN): layer created on top of the blockchain enabling fast transactions between users, and reduces the number of transactions written to the blockchain.

Throughput: number of tasks that can be completed in a given time.

Bitcoin, Blockchain, the Bitcoin Scalability Problem and the Lightning Network

Before examining the data collected for this project, it is important to establish a more in-depth understanding of the three major components surrounding the lightning network, as well as the lightning network itself. The reason for this detailed view of components is to explain the relationships between them in order to understand where the data being analyzed originates.

What is Bitcoin?

Bitcoin is a peer-to-peer version of digital currency created in 2009 by an anonymous person, or group, using the pseudonym Satoshi Nakamoto[3-4]. This currency is completely decentralized, meaning that no government or other organization has control over it, which allows direct transactions between users in the network without having to go through a financial institution. On the network, users are completely anonymous and only connected by encryption keys using public key cryptography, making all transactions untrackable despite the fact that they are all permanently written to the blockchain. As it is a digital currency, bitcoin is not manufactured like bills or coins; it is mined by a person, or company, using specialized hardware and software that provide massive amounts of computational power. Essentially, the mining process is as follows:

1. A user on the network sends another user bitcoin,
2. The network records this transaction, as well as all other transactions during a certain time period, in what is called a block,
3. "Miners" print these transactions to a digital ledger known as the blockchain by generating a hash of these blocks and add them to the end of the blockchain,
4. For doing this work, miners are rewarded with bitcoin, which can be sold and purchased on a number of different digital currency exchanges.

Although this is still a high-level overview of bitcoin, and the processes surrounding bitcoin are much more complicated in reality, it provides enough background information for the purpose of this project.

What is Blockchain?

Blockchain is a form of Distributed Ledger Technology used for recording digital transactions in a network securely[5]. Like bitcoin, blockchain technology was also created by Satoshi Nakamoto. Essentially designed as a base technology for bitcoin, blockchain maintains all records regarding transactions within a bitcoin network by writing transactions to blocks, which, once filled, are added to the end of the blockchain and linked together using cryptography. To add another layer to security to the blockchain, all blocks written to the blockchain are immutable, meaning they cannot be altered or deleted and remain permanently on record in order to provide proof of ownership. This security is what makes blockchain technology a suitable substitute for a centralized financial institute; however, there are inherent characteristics that limit the bitcoin blockchain's capacity to scale.

What is the Bitcoin Scalability Problem?

The bitcoin scalability problem arises as a result of 3 key characteristics of the bitcoin blockchain:

1. Blocks in the blockchain are limited to only 1MB in size,
2. Creating a new block after one has been completed takes 10 minutes or more,
3. There is a maximum of 7 transactions per second[6].

Due to these limitations, it is clear that if the number of bitcoin transactions were to increase from under 200 000, which is the number of transactions per day at the time of this report, to millions, or even billions, of transactions per day, the throughput of the bitcoin blockchain network would be hugely impacted[7].

What is the Lightning Network (LN)?

The lightning network was proposed in 2015 as a potential solution to the bitcoin scalability problem. In general, the idea of the lightning network is to create a second layer on top of the bitcoin blockchain that is made up of user-generated channels, allowing instant payments between users without having to broadcast to the entire network. This means that each transaction would only have to be written to the blockchain when the payment channel is opened, and again when the channel between the users is settled, in turn reducing the number of transactions written to the blockchain per second, as well as reducing the number of times a new block would have to be generated.

Centralized, Decentralized, and Distributed Networks

As the goal of this project is to determine if the lightning network is in fact a decentralized network, it is important to clearly define the difference between centralized, decentralized and distributed networks.

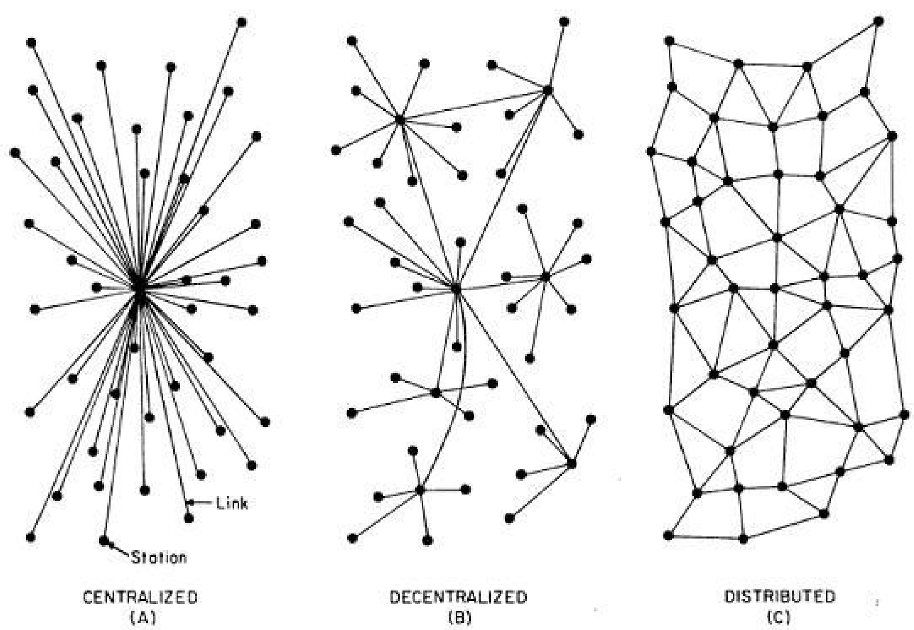

Figure 1. *Graphical Representation of Centralized, Decentralized and Distributed Networks.*
Retrieved From: https://medium.com

Centralized Networks

In a centralized network, there is a single, centralized node that connects to all of the other nodes in the network. As can be seen in figure 1(A) above, the centralized network contains a node in the center of the graph that is a direct neighbour to each of the other nodes. An example of this kind of network is a banking system. The central node is the bank, as each transaction that occurs involving this bank must be approved by the institution.

Decentralized Networks

In a decentralized network, such as the one depicted in figure 1(B), there are more than one central node. Hubs begin to develop, as there are a few major nodes that connect centralized groups of nodes. As will be shown, the lightning network is a decentralized network where central nodes are ones that have many transaction passing through them.

Distributed Networks

Distributed networks are networks in which there are no central nodes, and nodes are only connected to peers. Figure 1(C) illustrates such a network. There has been significant contention that the lightning network could become this kind of network; however, based on analysis of the lightning network data this seems very unlikely. Realistic examples of this kind of network can be seen in peer-to-peer projects, such as blockchain technology.

Data Collection and Cleaning

The data used for this project was collected from the site https://shabang.io, a site dedicated to providing information regarding live lightning network channels every 15 minutes, on March 19th, 2018 at 4:38pm[8]. At the time of collection, the network featured 843 nodes, 4779 channels, and a block height of 514 312. A more detailed breakdown of the following raw data and data cleaning can be found in the file finalScriptR.docx submitted along with this report.

Raw Data

The raw data collected included 3 .json files consisting of lightning network information. The following list contains the name of the .json file, as well as all of the keys belonging to each:

1. Channels.json
    a. Source
    b. Destination
    c. Short_channel_id
    d. Flag
    e. Active
    f. Public
    g. Last_update
    h. Base_fee_millsatoshi
    i. Fee_per_millionth
    j. Delay
2. Nodes.json
    a. Nodeid
    b. Alias
    c. Color
    d. Last_timestamp
    e. Addresses
    f. Port

3. Stats.json
    a. Height
    b. Lightning_node
    c. Lightning_channel
    d. Segwit_input
    e. Total_input
    f. Lightning_channels_funded

The most important files considered for this project were channels.json and nodes.json. Channels.json provided the information necessary to establish links between nodes, and nodes.json supplied information about each node, such as their id and name. Stats.json was not considered as it did not contain information useful for the purposes of this project.

Data Cleaning

The data cleaning process began with converting all of the downloaded .json files into .csv files for convenient reading in Microsoft Excel and R. From there, it was clear that the data would have to be subsetted in some way so that meaningful analysis could be performed. Initially, data was arbitrarily halved, however, it was determined that this method of elimination resulted in the loss of too many connection in the network. To resolve this issue, the original data was subsetted based on whether or not nodes contained valid IP addresses. The resulting data set contained 381 nodes, 462 nodes less than the original data provided. The columns of the new data files were then renamed for readability, and joined in order to facilitate the analysis process.

Analysis

To analyze the data, nodes and channels were plotted to a number of different graphs in order to determine if the lightning network is indeed a decentralized network. From these plots, network measures were applied to verify that these visualization were correct, as well as to provide quantitative results. Code snippets of the measures used and plotting functions, as well as the details of other tools used, can be found in the finalScriptR.docx file and source code submitted along with this report.

Geolocation

Before plotting the cleaned data set in any kind of graph, IP addresses were converted to latitude and longitude coordinates so that the network could be visualized on a world map. The mapping was done using ArcMap, a program of the ArcGIS software.

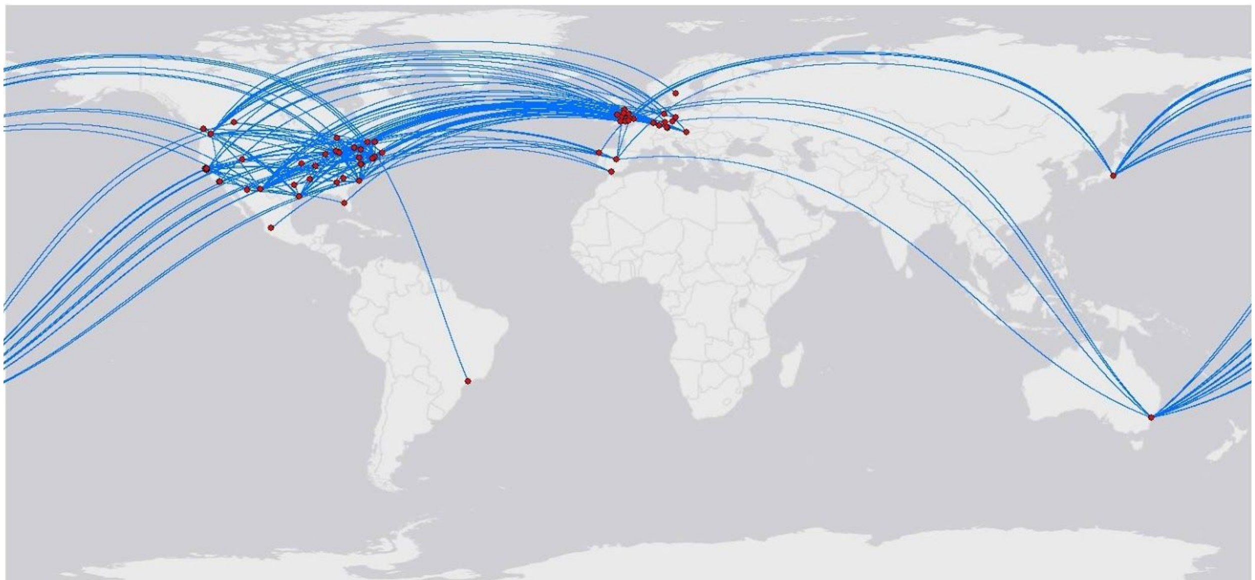

Figure 2. *World Map Containing Geolocated Lightning Network Nodes.*

The map shown in figure 2 shows the 381 nodes and their geographical location. Although the map does not appear to have all 381 many nodes, they are all plotted. The reason for this is appearance is that many of the nodes are clustered so closely together that a group of them on such a large-scale map results in what looks like a single node. The breakdown of nodes by continent is as follows:

- North America: 323 nodes, 307 of which are in the United States.
- Europe: 48 nodes.
- Asia: 9 nodes.
- South America: 1 node.

It was determined that Ann Arbor, Michigan contained the most nodes of any region, with 102. This was the first indication of hubs in the lightning network, which suggested that the network is likely decentralized, as opposed to completely distributed.

Simple Graph

The first graph visualized was done by plotting a simple graph in R. The layout of this graph is random; edges represent channels, and nodes are the users active in the channel. Nodes with no connections in this case represent nodes that were connected to other nodes that were removed during the cleaning process.

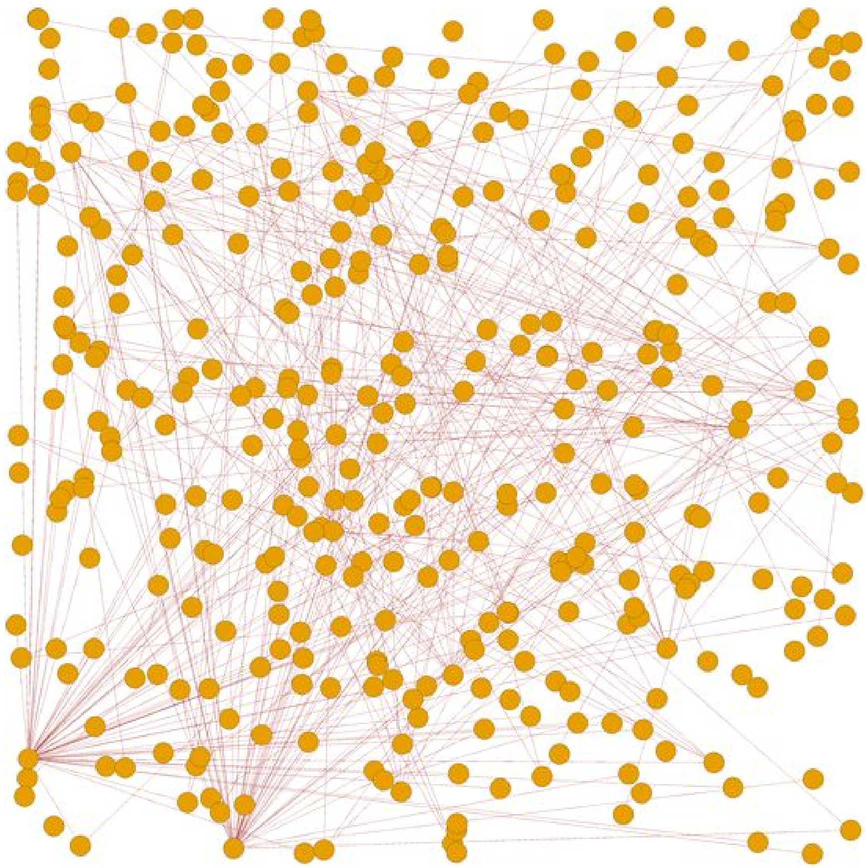

Figure 3. *Simple Graph*

Fruchterman-Reingold Force-Directed Graph

The Fruchterman-Reingold is a force-directed layout algorithm that applies a force to the set of edges and connected nodes based on their relative position. These forces attempt to layout the graph with minimal overlap, unlike the simple graph above.

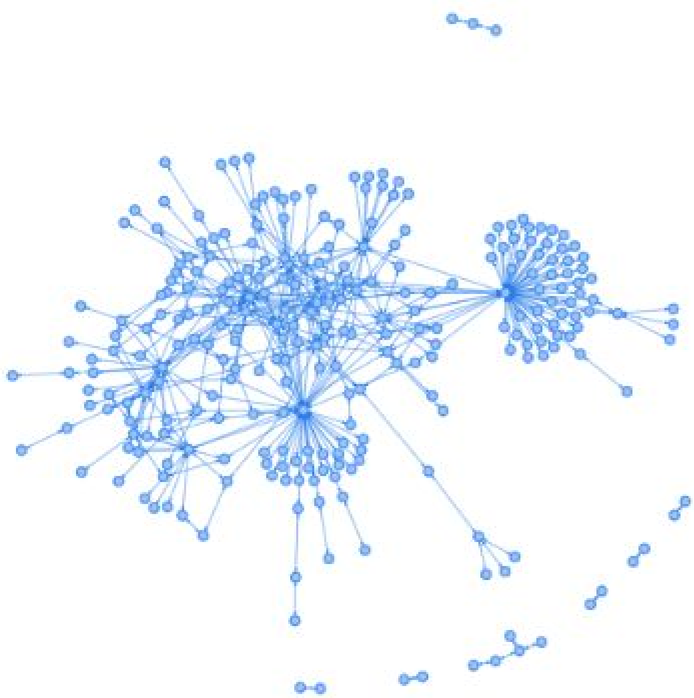

Figure 4. *Fruchterman-Reingold Force-Directed Graph*

As can be seen in figure 4, points that are closest to each other form their own hubs, while nodes that distant from one another extend away from these groups. By examining this graph, it becomes clear that hubs are a major component of the bitcoin lightning network, and that this network is in fact decentralized, rather than distributed.

Circle Graph

By creating a circle graph, a conclusion similar to that of the one determined by the Fruchterman-Reingold graph arises: there are hubs clearly visible in the lightning network.

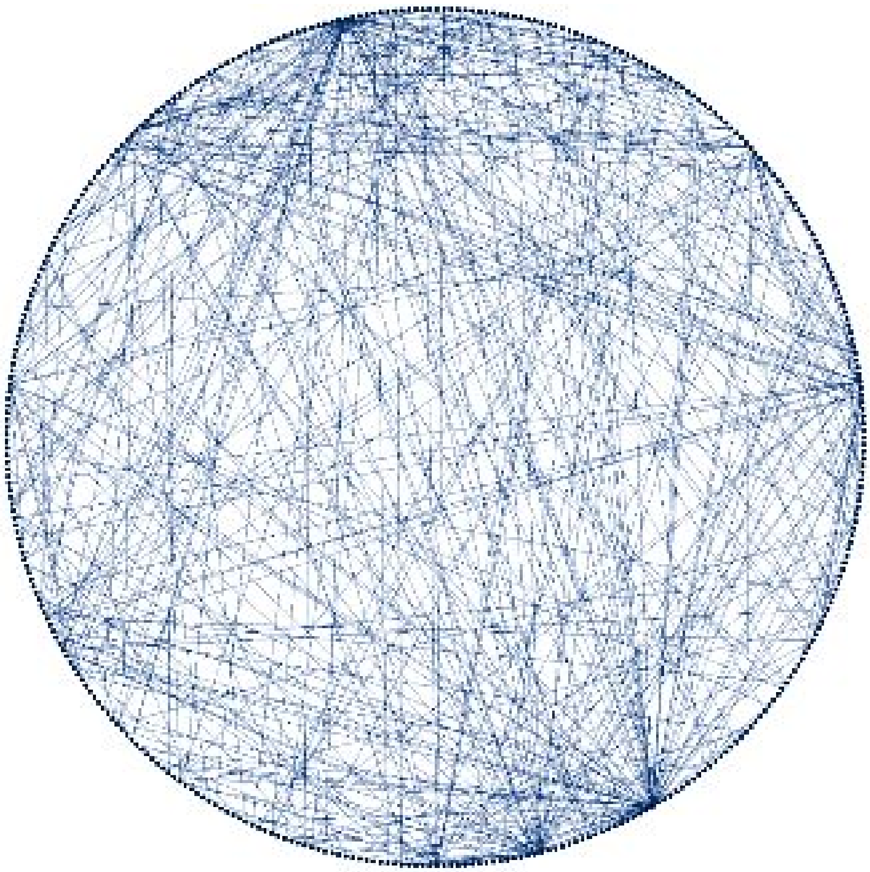

Figure 5. *Circle Graph*

In the top and bottom portions of the circle, a number of different hubs can be seen. Determining exactly which nodes these are in the graph (i.e. retrieving their id) is as simple as adding labels to each node, however, for the sake of readability and clarity of the graph, labels have been removed.

Vertex Degree

The degree of a vertex is defined as the number of edges adjacent to a specific node. By using this measure, the number of transactions passing through a certain node can be determined.

020d3d5995a973c878e3f6e5f59da54078304c537f981d7dcef73367ecbea0e90e	128
02f6725f9c1c40333b67faea92fd211c183050f28df32cac3f9d69685fe9665432	100
021f2cbffc4045ca2d70678ecf8ed75e488290874c9da38074f6d378248337062b	54

Figure 6. *Top 3 Largest Hubs in the Lightning Network Based on Vertex Degree*

This measure was used to determine the largest hubs in the network. Sample output of this calculation done in R can be seen in figure 6. The ID of the node is in the left column, and the degree of the node is in the right column.

Global Clustering Coefficient

The global clustering coefficient, sometimes referred to as transitivity, is the number of closed triplets over the total number of triplets in a network[9]. The aim of this measure is to calculate the probability that the adjacent nodes of specified node in a network are connected. The global clustering coefficient was used to determine that the lightning network is undoubtedly a decentralized network, and not a distributed one.

```
> # Clustering coefficient (transitivity of a graph)
> transitivity(netGraph, type = ( "global"), vids = V(netGraph),
+              weights = E(netGraph)$weight, isolates = c("NaN", "zero"))
[1] 0.02839833
```

Figure 7. *Global Clustering Coefficient of the Lightning Network*

With a global clustering coefficient of 0.02839833, the lightning network is far from a fully connected, distributed network. If the network were to be fully distributed, the transitivity should be significantly higher, as every node would transitively connected as a results of the fully peer-to-peer network.

Betweenness Centrality

Betweenness centrality is the measure of centrality in a graph based on the shortest paths between nodes[10]. For every connected pair in a graph, there must be at least one shortest path between nodes that minimizes either the number of edges passing through the node, or the sum of the weights of the edges[11].

```
03b7d8b6bbaf02239277ed32378d50d29840292b61d516d30057b74044992c93dd 0330c464cb2be97cd4ca5057b192a2be3c775a5f0356aced805769cb8790b879c0
                                                           0.00000                                                            0.00000
0267db6fbe76e1fbae50e27529a330837e9e1f4b9e4c7bbfefd7d6a1b3ffe2b245 026b4f8931fcf87033d0f601ad7e4baa8e93ee74acf313292fd397fd6c27524162
                                                           0.00000                                                            6.50000
03efeea8961f376931a390ed9ae62be116abc1a8abaad6d1998efcc11d63e86526 0207197d1028b8a7edfee28f4e2dc47905333b2c4f9ed40bcd2c7481abe8fb049a
                                                        1311.23177                                                          604.26638
0235447c7485ff2b945bac5fbc366d54a87389bab8cacf1b64b26ec01e96bd165a 0323e18348bb2afc29660da8ba06fb1e91fc37a5752301180b8f3afca7f5c49f01
                                                        8011.68002                                                            0.00000
02c119d2fd2e98a88f50d0d2ee4213255b7b8ec2be3a95f9aabd6afb09dd25b083 03587e75c5928c0bdaae3b100f3edf7211bfd900d08a93f391e7d20fe062eeb37c
                                                        4652.37868                                                            0.00000
0208b6ec8c4c77cabef8351f92327c13e0b7cc2e3724083669de305e2b395341fe 0226b317ebf63a888838c2900b0e77e45b3ad35c59bdbbd6a8eaba0c0362303cee
                                                           0.00000                                                            0.00000
033ac2f9f7ff643c235cc247c521663924aff73b26b38118a6c6821460afcde1b3 02b7060f74b7e04d3d8af97fab20381fcc16f7a33c7e526fa5c9b96afdb288d7d2
                                                         280.40898                                                          112.66667
02a90ee457ad397e5e882780b1fa0d109c3b66479d38e7682f5ee75fe995800f8e 03cea3557a68bb4f2845808b937531e8666b16724e1d5f24215d9234efc5a57e7d
                                                           0.00000                                                            0.00000
0355cf9ce813a343313e1a4844f82c34810619943cb193cbfbead08da15d80fb98 0392e78c508987c97de70b2474493e9b0dbbd0dcad69dec2648d686701da1df44a
                                                           0.00000                                                            0.00000
0325fd957aeaead3635d7593b56c717090f95e7808290216b452e23fb2c0d77d11 02a0f1e7defe594d4f89e056ac45dbc3c2bdb29480270ed11c11040128a4ca6285
                                                           0.00000                                                           91.66667
0388157a19ed80015b64bf9cd3cbc3c3c4492822341ab7813c7f9c0dbccc08fb26 035ec1f8ea7d376385dfa7c92bd9a0c43e1c88a4058179c3a80ac68abc27dd2389
                                                           0.00000                                                            0.00000
```

Figure 8. *Betweenness Centrality of First 22 Nodes in the Lightning Network*

The sample output shown in figure 8 reaffirms the claim that the lightning network is decentralized, as it shows that there are few nodes with high betweenness centrality values. In this case, the nodes with high values indicate that they are a hub in the network.

Edge Betweenness Centrality

The edge betweenness centrality measure is used to determine the important edges in the network. It is defined as the fraction of shortest paths between all pairs of vertices passing through a particular edge[12].

$$c_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}$$

Figure 9. *Mathematical Definition of Edge Betweenness Centrality*

From this information, important edges in the network could be identified in order to determine which ones were critical to information flow. Figure 10 shows sample output of the edge betweenness centrality associated with the first 110 edges in the network.

```
  [1]  255.00000  244.50000   17.00000  697.37519  295.68911  478.69884   94.46862  115.01568  355.10399  191.71221  197.43450
 [12]  255.00000  237.62882  209.47108  693.75833  195.65514  104.85244  391.50878  341.19506  285.45280  665.60403  743.75000
 [23] 3769.18569  373.61784  134.53611  240.88957  260.01449   86.17338  227.20390  169.88922  481.30000  212.56739  255.00000
 [34]  118.10609  653.54573  255.00000  178.01268  255.00000  255.00000  441.96139  197.12019  231.05855  255.00000  255.00000
 [45]   86.02857  165.82118  110.30992  173.24931   14.50000  171.00000   70.00000   74.00000   38.16667  255.00000  255.00000
 [56]    1.00000  259.63810   87.02857  255.00000  173.69299  219.76997    1.00000  255.00000  255.00000 1025.82372 1052.50058
 [67]   91.96941  714.40924  255.00000  108.64591  255.00000  255.00000  319.53492   66.66190  255.00000  255.00000    1.00000
 [78]  255.00000 1040.17750   24.44136  145.10837 1223.13495  101.88501  255.00000  255.00000  277.00000  183.00000  616.82497
 [89]  147.24525  255.00000  255.00000   35.53096  255.00000   56.00000   68.67621   63.53096  115.76367   25.90000    7.00000
[100]  255.00000    7.00000  543.69701  255.00000   55.83096  255.00000  255.00000  255.00000  315.78333   80.18333  171.03889
```

Figure 10. *Edge Betweenness Centrality of First 110 Edges in the Lightning Network*

Closeness Centrality

The closeness centrality of a node in the lightning network describes its distance to all other nodes in the network. A node with a large closeness centrality value indicates that this node is more central to the network, and is able to transfer data between many other nodes. Similar to betweenness centrality, this measure is used to determine the which nodes in the network are hubs, since hubs will generally be closer to many other nodes.

0330c464cb2be97cd4ca5057b192a2be3c775a5f0356aced805769cb8790b879c0	6.91E-06
0267db6fbe76e1fbae50e27529a330837e9e1f4b9e4c7bbfefd7d6a1b3ffe2b245	6.91E-06
0323e18348bb2afc29660da8ba06fb1e91fc37a5752301180b8f3afca7f5c49f01	6.91E-06

Figure 11. *Node ID of Three Highest Closeness Centrality Values in the Lightning Network*

Of course, this data also provides proof that the lightning network is decentralized, as there are a number of nodes that are close to many other nodes in the network. The three highest rated closeness centrality nodes can be seen in figure 11.

Conclusion

The limited transaction capacity of the bitcoin blockchain network poses a significant risk for the future of bitcoin. As the bitcoin network scales in size, it becomes increasingly difficult to maintain an efficient and secure system using the blockchain alone. The lightning network emerges as a possible resolution to this scalability issue, as the number of transactions written to the blockchain are greatly reduced by introducing off-chain payment channels designed to write to the blockchain only twice. Although the lightning network was intended to be a decentralized network containing multiple hubs through which transactions would occur, there are many arguments claiming it can be a fully distributed peer-to-peer network. The goal of this project was to analyze the lightning network data available in order to determine if the lightning network is in fact a decentralized network. By plotting the data and applying various network measures using a range of software, including Microsoft Excel and R, it was determined that the lightning network is undoubtedly a decentralized network due to the abundance to hubs through which transaction travel.

References

1. "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments." Joseph Poon and Thaddeus Dryja. Accessed April 2, 2018. https://www.weusecoins.com/assets/pdf/library/Lightning%20Network%20Whitepaper.pdf
2. "Mathematical Proof That the Lightning Network Cannot Be a Decentralized Bitcoin Scaling Solution." Jonyald Fyookball. Last modified June 26, 2017. Accessed April 2, 2018. https://medium.com/@jonaldfyookball/mathematical-proof-that-the-lightning-network-cannot-be-a-decentralized-bitcoin-scaling-solution-1b8147650800
3. "Bitcoin: A Peer-to-Peer Electronic Cash System." Satoshi Nakamoto. Accessed April 2, 2018. https://bitcoin.org/bitcoin.pdf
4. "What is Bitcoin?" Tal Yellin, Dominic Aratari and Jose Pagliery. Accessed April 2, 2018. http://money.cnn.com/infographic/technology/what-is-bitcoin/
5. "Blockchain, Digital Trust and Distributed Ledger Technology." i-Scoop.eu. Accessed April 2, 2018. https://www.i-scoop.eu/blockchain-distributed-ledger-technology/
6. "On Scaling Decentralized Blockchains." Kyle Croman, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, Emin Gün Sirer, Dawn Song and Roger Wattenhofer. Accessed April 2, 2018. http://www.comp.nus.edu.sg/~prateeks/papers/Bitcoin-scaling.pdf
7. "Blockchain Charts." Blockchain Luxembourg S.A. Accessed April 2, 2018. https://blockchain.info/charts
8. "Network Data Collection." shabang.io. Accessed March 16, 2018. https://shabang.io/
9. "Clustering Coefficient in Graph Theory." GeeksforGeeks. Accessed April 6, 2018. https://www.geeksforgeeks.org/clustering-coefficient-graph-theory/

10. “Betweenness Centrality (Centrality Measure).” GeeksforGeeks. Accessed April 6, 2018. https://www.geeksforgeeks.org/betweenness-centrality-centrality-measure/

11. “Betweenness Centrality (Centrality Measure).” GeeksforGeeks. Accessed April 6, 2018. https://www.geeksforgeeks.org/betweenness-centrality-centrality-measure/

12. “Uniform Edge Betweenness Centrality.” Heather A. Newman, Hector Miranda, Rigoberto Florez and Darren A. Narayan. Last modified September 14, 2017. Accessed April 6, 2018. https://arxiv.org/abs/1709.04773

Supplemental Materials

1. https://explorer.acinq.co - visualization tool shown during in-class presentation.