

The origin of the bitcoin:

A network analysis

By:

Dmitrii Brystov

Joran Vergauwen

-

In the framework of the course:

Network modelling and design

-

Given by:

M. Pickavet, P. Audenaert

T. Walcarius, P. Stroobant



Contents

The roles.....	3
The setup	5
The algorithm.....	6
Basics.....	6
Computational shortcuts	9
The analysis.....	10
Networks after 10.000 transactions (3 months).....	10
Analysis: network originated from the most important node.....	11
Analysis: network originated from the 7th most important node	14
Analysis: network originated from the 8th most important node	16
Analysis: network originated from the 10th most important node	18
Networks after 50.000 transactions (1 year and 2 months).....	19
Analysis: network originated from the 4th most important node	20
Analysis: network originated from the 6th most important node	23
Analysis: network originated from the 7th most important node	24
Analysis: network originated from the 9th most important node	27
Networks after 70.000 transactions (1 year and 6 months).....	29
Analysis: network originated from the most popular node.....	32
Network after 100.000 transactions (1 year and 8 months)	37
Analysis: network originated from the most popular node.....	37
Networks after 200.000 transactions (2 years)	48
Mining behaviour	49
Network after 500.000 transactions (2 years and 5 months).....	52
Slush Pool.....	52
Conclusions	57
Limitations and going forward.....	58
REFERENCES.....	60
ADDENDUM	61
Input file and output file analysis	61

The roles

Bitcoin was released in 2009 with the purpose of creating a decentralized digital currency in which the transparency of the transactions would be one of the strongest assets. The general idea that everyone could prove the legitimacy of the transaction gathered a lot of faith in the minds of the crypto enthusiasts. This report sheds light on the network that was created by the transaction behaviour of the BTC users and shows that the intended function as a payment method was not fulfilled.

For a better understanding of the future analysis of the BTC network, it is important to note some important roles a node in the network can fulfil. Therefore, a short introduction towards the behaviour of the different roles of the main bitcoin users is given below.

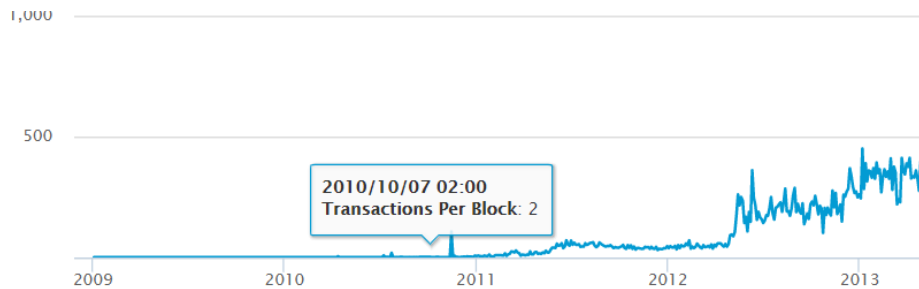
The bitcoin exchanges

Originated on 17 March 2010, the first bitcoin exchange started to operate as a central market where one could buy and sell bitcoins. For this reason, the first bitcoin exchange was identified as the node with the biggest degree, i.e. the summation of the indegree and outdegree. For the time periods after 17 March 2010, this bitcoin exchange was used as the first node to grow a network from. Important to note though is that over the entire existence of bitcoin many bitcoin exchanges were set up. Unfortunately, about 50% of them had to be closed again, often due to a security breach. For this reason, it might be a bad idea to keep using the 'first bitcoin exchange' as the first node when going further in time. Identifying other exchanges and using them as the starting point might be a good idea. Furthermore, it is also important to keep track of the degree of the particular exchange node that is chosen to grow the network from. Namely, when the degree of the exchange node does not grow any further while increasing the amount of analysed transactions, the exchange was probably shut down. For this reason, it might be interesting to grow a network starting from a different exchange node. Since bigger datasets imply a lot of computational power, our analysis did not go this far in time so we did not have to deal with this particular problem.

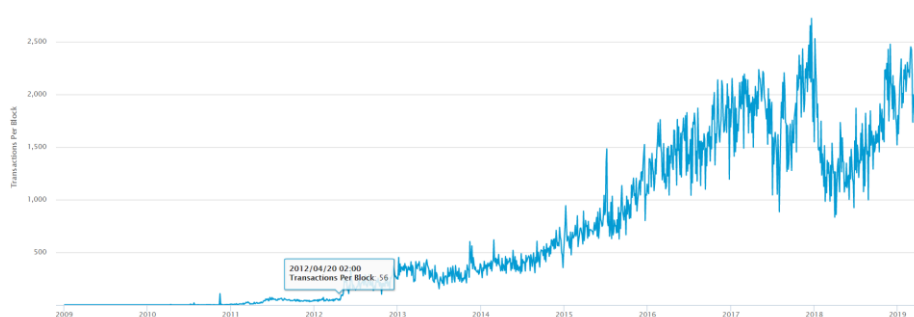
The miners

Bitcoin transactions are stored in blocks with a certain block size. After a block is filled, it has to be approved by a party from the network, who is called the miner. To compensate for the work associated with verifying the block, the miner gets a reward in the form of fresh bitcoins that enter the system. The approved blocks are then sequenced after each other to create the actual 'chain'. For example, when 30 transactions are executed and the block size is 10, 3 blocks containing 10 transactions are filled and sequenced after each other. Further details about how this chaining exactly happens will not be discussed here. Although, the size of the blocks is an important factor to take into consideration. Since, if the block sizes are relatively high to the transaction volume, the amount of blocks that would have to be approved would be rather small. For example, in 2019 block sizes containing between 2.000 and 2.500 transactions are quite common. Since our goal is to analyse how the bitcoin started, it is important to know how big the blocks were at the time to come up with a proper approximation of how many blocks needed to be approved. And, given the fact that the mining fees generate a bitcoin inflow for the miners, the number of blocks then gives a first approximation towards how many indegrees we could expect resulting from the blocks that are mined. So, considering those block sizes, graph 1 and graph 2 show the evolution of the block sizes through time. Given the fact that this report only studies the first 2 years of the bitcoin existence, the assumption that the number of blocks is almost equal to the amount of transactions can be made.

As a last mention, it is very important to know where these mining rewards actually come from. The rewards that are associated with the mining of the blocks are fresh bitcoins that get injected into the network. So far, the amount of bitcoins in rotation is roughly 16.7 million coins, with 12.5 new ones released approximately every 10 minutes via the mining process. Bitcoin plans to do this until the supply limit of 21 million bitcoins is reached, which is expected to happen around the year 2140. So, until then these miners will be competing with each other to receive the fresh coins that are released in the system.



Graph 1: Size of the bitcoin blocks – relevant period



Graph 2: Size of the bitcoin blocks – full history

The mining pools

However, attracted by the mining rewards some parties in the bitcoin network decided to build a business model around the mining of the blocks. Facilities with a lot of computational power were set up to increase the odds of approving the block first and claiming the reward, heating up the competition of the mining of the blocks. I.e. whoever mines the block first, claims the reward. Although this behaviour had a pure financial motivation, it also had other implications for the network. Namely, these so called ‘mining pools’ started to dominate the approving process of the blocks, making the ‘decentralised’ property a subject of heavy discussions.

The investors (in the currency)

Another part of the population just bought bitcoins as an investment at the time. These people did not really use the currency for its intended purpose which was performing transactions. The transaction behaviour of these investors is generally characterised with a very low outdegree. In the report, outdegrees of 0 and 1 will seem to occur very frequently, meaning that a lot of people bought some bitcoins to leave them in their wallet as an investment.

The traders

These are the people who actually use the BTC as a currency. They buy and sell goods and currencies for other currencies. The analysis will show that there were not a lot of people using the currency in the way it was intended to be used.

The setup

Going through time, some specific points were chosen to be able to make conclusions about the impact of certain events on the network. Due to the exponential behaviour of the computational time of the algorithm related to the amount of observed transactions, the choice was made to keep the amount of transactions lower than 500.000. This results in an analysis from 3/01/2009 until 17/05/2011. Due to the fact that the amount of BTC transactions per day also grew exponentially after this period, 1.000.000 transactions were already reached 2 months later. Given the fact that analysing 1.000.000 transactions takes about 60 hours of computational time, the extra information did not outweigh the additional computational time needed. For this reason, the following paragraphs discuss the timestamps which were considered as interesting moments to observe.

The first period is at about 10.000 transactions; the bitcoin exists 3 months now. Given the fact that the BTC is really in its infancy here, it is interesting to look for some first patterns. Analysing this period results in a difficulty to grow graphs from the nodes. This could be surprising since the 10 most popular nodes were chosen to grow a network from. Furthermore, the originated graphs are quite small considering the fact that 10.000 transactions were observed. For some of the 'popular nodes', the algorithm couldn't even grow a network at all. One could interpret these tiny networks as networks that are scattered without being linked to each other. In other words, the most important nodes exist in their own separate network within the bitcoin environment. Though, this conclusion would ignore the characteristics of the given datasets and would not explain why the algorithm fails to grow a network in some cases. Namely, to understand these phenomena one should know about the asymmetry between the input and the output files, which were taken from the MIT bitcoin network dataset. The link towards this dataset can be found in the references of this report. While the output file nicely starts counting the transaction ID's in a sequence of 1 ID per step starting from the first transaction, the input file shows a different behaviour. More specifically, the input file has the tendency to skip some of the transaction ID's, so we cannot track down who was putting in the money here. In other words, the algorithm struggles to grow a network from certain nodes which are popular in the output file but for which the corresponding transaction ID is missing in the input file. Due to this mismatch, no clear-cut decisions can be made considering the existence of a giant component at this stage of the bitcoin network. However, we believe that this mismatch can be explained. One must know that, even 10 years after bitcoin's invention, the true identity of the founder of bitcoin remains a mystery. The official founder, Satoshi Nakamoto, is a pseudonym for a person or group of people whose identity is still unknown. Since the founder obviously gave the input of many of the first bitcoin transactions, he probably wiped himself from the input file to protect the mystery around his identity. If this theory is true, a big point of critique towards the bitcoin can be found here. First of all, it is very strange that the identity of the currency's founder is still unknown. Even more striking is the fact that this person could hide his own input in a transaction system that would later pretend to be '100% transparent'. To make it even more complex Craig Wright, one of the people who are linked to this Satoshi persona, was accused in 2013 of having stolen over \$10 billion worth in bitcoins from his former partner Dave Kleiman. All those events make the early existence of bitcoin very mysterious and not transparent at all. Therefore, this report attempts to make the proper assumptions when analysing the networks that are formed after different time periods.

The second period was chosen a bit before the introduction of the first bitcoin exchange. More precisely, the choice was made to observe the network until the beginning of March 2010. Bitcoin is now 50.000 transactions in, the amount of transactions per day is about 200. Analysing the obtained graph, the conclusion can be made that the input-output mismatch still affects the ability to grow networks from certain nodes. For the other nodes, one can conclude that the networks tend to

become a bit bigger, especially for the 'fourth node'. (The node that was indicated being involved with the fourth most transactions in the observed dataset)

To grasp the impact of the bitcoin exchange as accurate as possible, this time frame was initially chosen along with 2 other periods. A first observation considers the short-term impact of the exchange node, observing the exchange node already 3 months after its introduction. I.e. when the total amount of transactions is 70.000. These observations will be discussed in the next paragraph. A second observation attempts to grasp some medium-term effects of the BTC exchange. Therefore, the network was observed almost 6 months after the introduction of the exchange, at 100.000 transactions.

However, after 200.000 transactions we encountered a network that seemed to have 2 important miners and one important exchange node. The fact that the mining rewards are new bitcoins injected into the system explains the huge indegrees of these nodes. Probably, the bitcoin developers bought some servers to start up the network as efficiently as possible by mining a lot of the blocks themselves. Note that this does not suggest any form of backdooring or illegal mining of the blocks here. The developers probably set up their servers and let them compete with the other miners in a fair way. The reason why they had these big indegrees is then only due to the better processing power of these servers compared to the processing powers of the individual miners.

Going further on this approach, the ambition was to go a bit further in time to shed some light on the first mining pool, Slush pool. For this reason 500.000 transactions were analysed, thereby investigating the bitcoin network until around 05/2011.

The algorithm

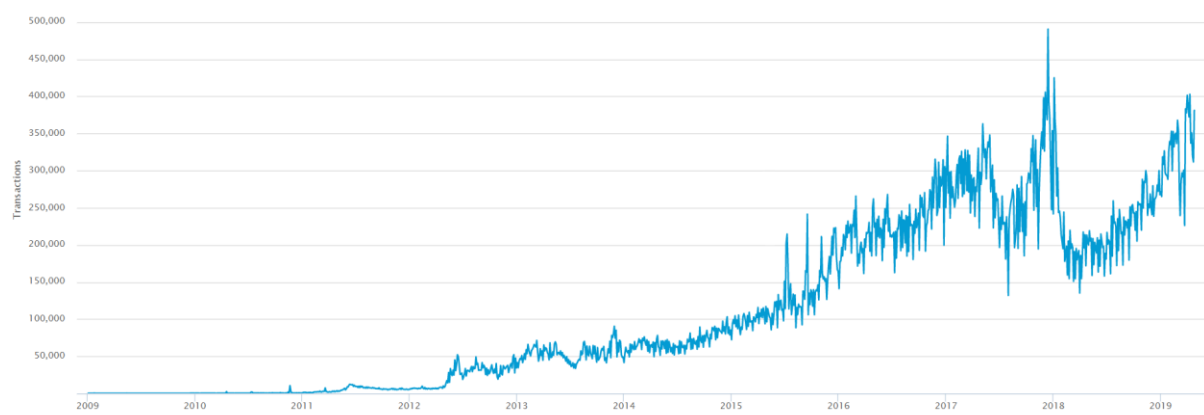
Basics

Considering the size of the dataset, the first objective of the algorithm is to extract a smaller subset from the input and output files. In order to make a right estimate of the amount of transactions associated with a certain period of time, a file describing the amount of transactions per day was extracted from the Blockchain Explorer (www.blockchain.com). Unfortunately, exporting this file from the Blockchain Explorer results in a file that reports the number of transactions per day only for every other 3th day. For this reason, the choice was made to multiply the amount of transactions on a certain day times 3 to approximate the amount of transactions over 3 days. Table 1 provides an example of this approximation method. Though, when performing those kinds of approximations, one must always be aware of the possibility of underestimating/ overestimating the real values. Graph 3, extracted from the Bitcoin Explorer, hints that the number of transactions doesn't really grow exponentially until 2012. Therefore, one would conclude that the approximated cumulative number of transactions will be very close to the actual number of transactions. However, zooming in on the transactions between 03/01/2009 and 01/01/2011, some outliers are reported on graph 4. While normal values are within the [0, 250] range, one of these outliers even has a value of 10.378. Although, multiplying the observed outliers assumes that 3 days have these high transaction volumes, therefore potentially making a big error. When expressing the error in an amount of days, it is possible that the obtained cumulative transaction volume was only reached about 80 days later. (2 days in which the overestimation is 10.378 transactions / 250 transactions per day = 41 days) This is a very significant difference and should be considered when trying to investigate certain events like the introduction of the first bitcoin exchange. The way we accounted for this problem was by making sure that the time window was big enough to see results. If no differences could be observed, this could be due to a bad estimation of the time window. Making the time window wider could bring a solution here. Fortunately, the errors seem to be reported more towards the end of the observed period. We believe

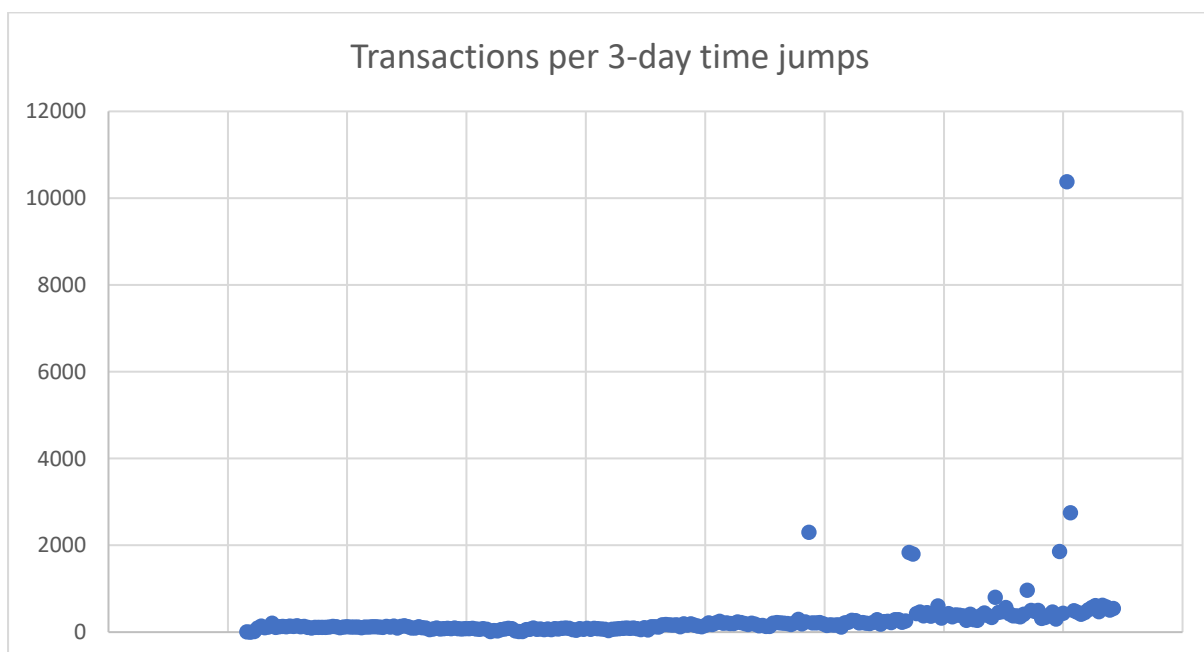
this is one of the reasons why we did not encounter this particular problem when investigating the introduction of the first Bitcoin Exchange. Another reason could be the fact that similar outliers were possibly left out in the 3-day dataset, thereby cancelling out the error made by multiplying the observed outliers.

Day	tx per Day	tx per 3 days	Cumulative tx
03/01/2009 00:00	1	3	1
06/01/2009 00:00	0	0	1
09/01/2009 00:00	14	42	43
12/01/2009 00:00	95	285	328
15/01/2009 00:00	136	408	736
18/01/2009 00:00	101	303	1039

Table 1: Approximation of the amount of transactions through time



Graph 3: Amount of transactions through time



Graph 4: Amount of transactions per 3-day time jumps until 01/01/2011

Once the estimation of the number of transactions associated with a certain time frame is made, the next problem already occurs: there is an asymmetry between the input file and the output file. While the output file seems to be a solid dataset, beginning with the first transaction and reporting every single transaction ID, the input file has a rather odd behaviour. The input file starts at transaction ID 170 and has the characteristic to jump from one transaction ID to the other, skipping the ID's that are in between. For this reason, transaction ID's that are found in the output file might not be matched with their corresponding transaction ID in the input file. Trying to account for this problem, the multiple.dat file was analysed, but this file does only contain transactions with a very high transaction ID, thereby providing no solution for the mismatch. Since these 3 files are the only files containing addresses, no solution could be found for the mismatch between the input and output file. In other words, when looking for popular nodes in the input and output files, one should always be aware of the fact of this mismatch. This becomes even more important when trying to grow a graph around a certain node. When a certain address is very popular in the output file, it could be possible that the transactions in which this address was involved could not be tracked down in the input file. For this reason, it might be very hard to grow networks around graphs that are popular in the output file.

Considering the definition of a 'popular node', the algorithm runs through the imported subsets of the input and output files and calculates how many times a certain address occurs in those files. Adding the popularity in the input file to the popularity in the output file results in a general popularity. Finally, when the popularity of every address is calculated, a ranking of the 10 most popular nodes can quickly be generated. Due to the mismatch between the input and output file, one could propose to only consider the input file when searching for the most popular nodes to improve the probability that a network can be formed in a later phase. However, we believe that this biases the result by not giving the real popularity of the nodes.

After this part follows the core of the algorithm: the growth of the network. The algorithm searches for the most popular address in both the input and output file and uses it as the initial node for the network. Associated transaction ID's are then tracked in the output and input file respectively. If the obtained combination of address ID's is not yet added to the graph, the graph is extended by the combination resulting in an edge between the two nodes. One could extend the graph by adding weights to the edges whenever duplicates are found. However, due to the fact that these weights would not really add value in our analysis, the decision was made to not include the weights. A second important note is the importance of the direction of the relationship. The address from the output file will always be the destination of the transaction while the address from the input file will always be the source of the transaction. In other words, when the link between 2 nodes is found in the 2 directions, both edges must be added to the graph.

In the next step, the algorithm extends the network by considering the nodes that are linked to the initial node. The same sequence of searching through the input and output files is repeated, while making sure that no duplicate edges are added to the algorithm. This step is represented by 2 big loops in the algorithm, which essentially do the same thing. The first loop performs one extra iteration, while the second loop allows the user to choose how many iterations to perform. Obviously, the amount of iterations stands for the amount of times we consider the newest nodes in the graph to continue the growth the network.

After this phase, the data gets written to a .csv file to make the analysis of the network in Gephi and excel possible.

Computational shortcuts

Honestly, most of the implemented shortcuts are actually rather logical steps to prevent unnecessary operations than using brilliant insights to make the algorithm shorter. Anyway, some of the implemented shortcuts are mentioned below.

A first measure was taken since the algorithm was prone to investigate popular nodes multiple times. For example, when multiple nodes refer to the same node, this node will be investigated several times in the next iteration of the algorithm. Of course, this brings zero added value to the table while the computational effort goes through the roof. The computational effort could be trimmed a lot by simply putting the investigated addresses in an arrayList and checking whether this arrayList contained the addresses which were encountered later.

A second measure ensured that only the latest nodes were assessed when growing the graph. In order to do this, a counter was added which kept track of the amount of links that were added in every iteration of the growth process. Consequently, the next iteration only assessed the nodes that were freshly added to the graph, ignoring the nodes which were already assessed by the algorithm. Note that this 'shortcut' is nothing more than the correct implementation of the growing process as was described earlier. Though, it is noted as a shortcut since the initial algorithm did not implement this correctly.

As a last mention, the speed of the algorithm relies quite a lot on our personal programming style. Since the use of arrays is more our 'jam', we had to be aware of the fact that arrays might take more processing power when used incorrectly. Due to the fact that arrays need to be initialised with predetermined dimensions, it is easy to lose processing speed when running through the entire array every time. Therefore, we made sure to cut the loops immediately when a row with a null value was encountered while iterating through the graph.

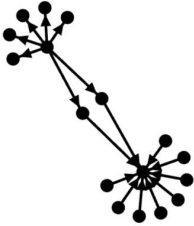

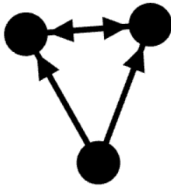
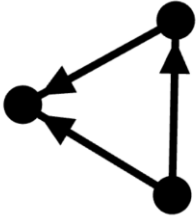
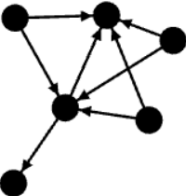
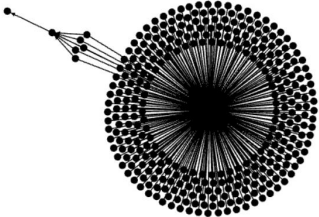
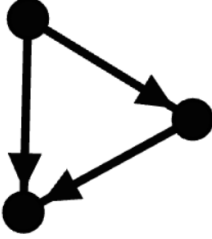
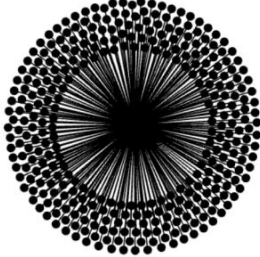
The analysis

As introduced above, an analysis of the networks is performed for different stages at the early phase of the bitcoin network. The following part summarises the most important results and interpretations that could be derived when analysing these networks.

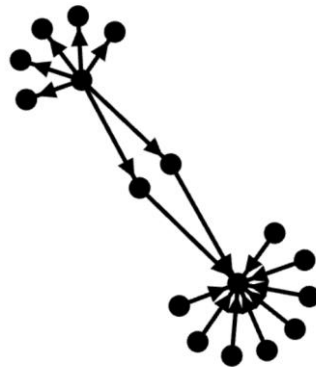
Networks after 10.000 transactions (3 months)

The following table illustrates the networks that could be formed by starting from certain popular nodes in the input and output files. Due to the mismatch in these files, some of the obtained networks are quite small and therefore considered as less interesting to analyse. These networks are the networks originated from the 5th, 6th and 7th most important node. Other nodes suffered even more from this mismatch and couldn't even grow a network at all. Due to the fact that the networks originated from the most popular node and the 4th most popular node are quite similar, the decision was made to analyse the networks resulting from the 1st, 7th, 8th and 10th most important node.

Another observation considers the fact that we seem to find different networks that are scattered within the bitcoin environment, which are not linked to each other. This is rather odd, since one would expect a network of a new currency to grow around the addresses of the developers and not around different random addresses. The mystery about the Satoshi persona could bring an explanation here. This individual was most likely the person who linked all these networks together. However, the bitcoin founder most likely hid his input from the rest of the world to keep his address ID as a secret. For this reason, one could assume that the networks below should all be linked to each other through this unknown node.

1st node	4th node	5th node
		
6th node	7th node	8th node
		
9th node	10th node	
		

Analysis: network originated from the most important node



General information

• Estimated time period	= 01/2009 – 04/2009
• Amount of transactions	= 10.000
• Iterations	= 4
• Amount of nodes	= 17
• Amount of edges	= 17
• Lines from txin file	= 620
• Lines from txout file	= 10.100

Description

The network shows that on the one hand there is a node having a large outdegree. On the other hand, there is also another node having a relatively large indegree. The node with the high outdegree could probably become a trader as the network grows through time. Though, note that the first purchase of a real 'product' with bitcoins was on March 22, 2010, way later than the observed ending date of this period. Adding onto this, the bitcoin network only existed for 3 months here so probably, a lot of transactions were made to try out the bitcoin system and get it up and running. For these reasons, it is hard to explain the actual meaning of high outdegrees in terms of transaction behaviour. The high indegree could probably mean that this node already hashed quite some blocks. Since at this time the network complexity was way smaller and the hashes were shorter, it took less computational time to mine a block. Therefore, some addresses could have been set up with the purpose of mining blocks to get the bitcoin network started more quickly.

Another interesting remark is that this network stopped growing after 4 iterations. The reason for this could be twofold. On the one hand, it could be possible that the bitcoin network wasn't entirely connected at this point and that this graph represents an isolated network in the bitcoin network (in terms of transactions). Although, the reader should constantly be aware of the mismatch between the input file (txin) and the output file (txout) which was already mentioned earlier. It seems like the input file skips a lot of transactions, which might blur the image we get from the bitcoin network here. This asymmetry is most likely the cause of the fact that the algorithm couldn't match any more txID's in both files. Finally, the generation of the source node is also counted as an iteration, so starting from the source node, one could easily reconstruct the 3 hops that were taken by the algorithm to form this network.

Vertex degrees

- **Degree distribution**

Figure 1 shows the indegree distribution of this network. This graph shows once again that there is 1 node with a fairly high indegree compared to the others. The outdegree distribution is plotted on figure 2. The node with the significantly higher outdegree is the most remarkable point on this figure.

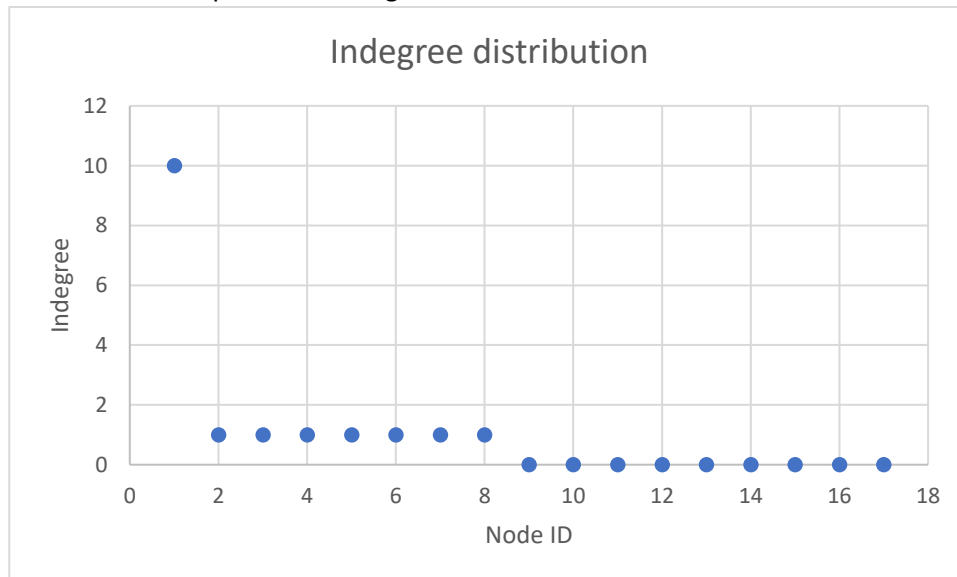


Figure 1: Indegree distribution

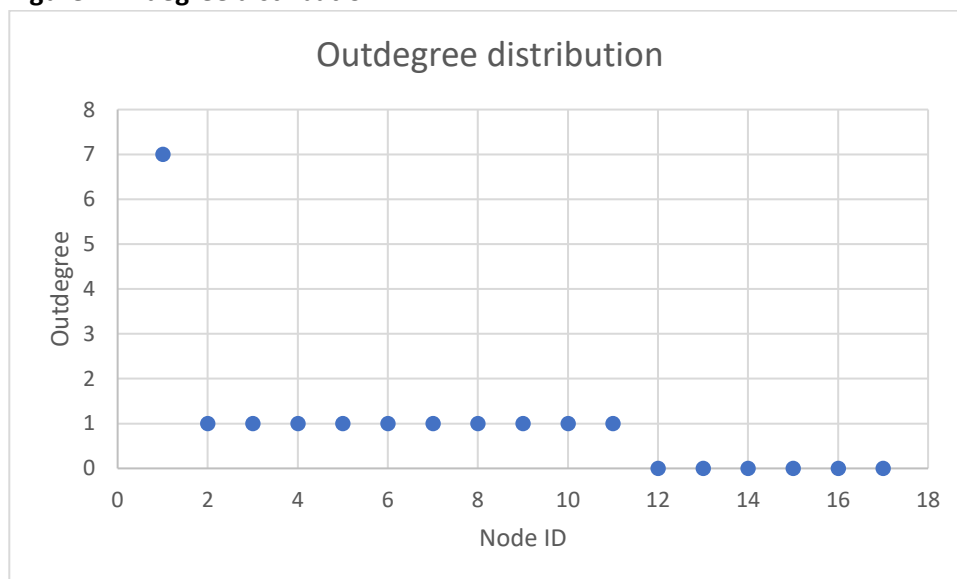


Figure 2: Outdegree distribution

- **Correlation(indegree, outdegree)**

The correlation between the indegree and the outdegree is -0.24398. For this network, nodes with a higher outdegree tend to have a lower indegree.

- **Degree minima and maxima**

Maximum indegree	= 10
Minimum indegree	= 0
Maximum outdegree	= 7
Minimum outdegree	= 0

The same conclusions as for the vertex degree distribution plots apply here.

Distance statistics

- **Average path length**
The average path length is defined for the directed graph, meaning that it is only calculated for the nodes that have a path between them. This average path is 1.056.

Clustering

- **Clustering coefficient**
A quick observation that the networks contains only triples and no triangles can be made. For this reason neighbours from a node will never be each others' neighbours, resulting in a clustering coefficient of 0.
- **Network density**
The network density measures the degree to which a graph is a complete graph. The reported density of 0.125 means that this network is far from complete.

Centrality

- **Betweenness centrality**
Figure 3 plots the betweenness centrality distribution of this network. 2 nodes happen to appear on the biggest number of shortest paths in the graph. Important to note here though is that the betweenness centrality was computed for the directed graph. The visual representation of the graph shows that the 2 nodes in the middle of the graph are the nodes on the shortest path from the node with the high outdegree to the node with the high indegree. Since 2 shortest paths can be constructed between these nodes, the betweenness centrality of each of the 2 nodes on the shortest paths is 0.5.

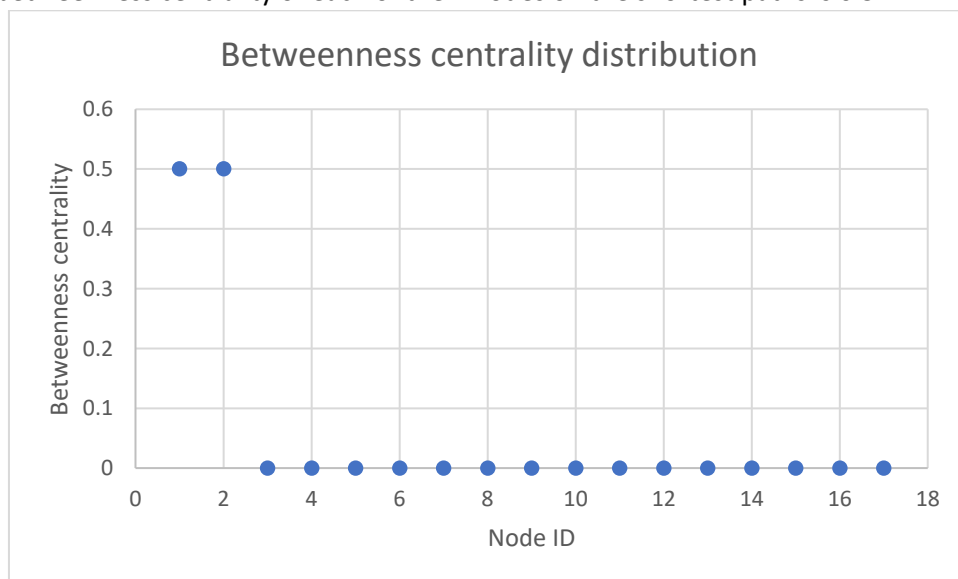
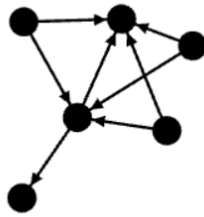


Figure 3: betweenness centrality distribution

Analysis: network originated from the 7th most important node



General information

- Estimated time period = 01/2009 – 04/2009
- Amount of transactions = 10.000
- Iterations = 10
- Amount of nodes = 6
- Amount of edges = 8
- Lines from txin file = 620
- Lines from txout file = 10.100

Description

This network is clearly different from the previous network. In this network, 2 nodes with a relatively high indegree can be observed.

Vertex degrees

- **Degree distribution**

Figure 4 and 5 show the indegree and outdegree distribution. The main take-away for this small graph is the asymmetry in both the indegrees and the outdegrees. Some nodes tend to have higher degrees while others have higher outdegrees.

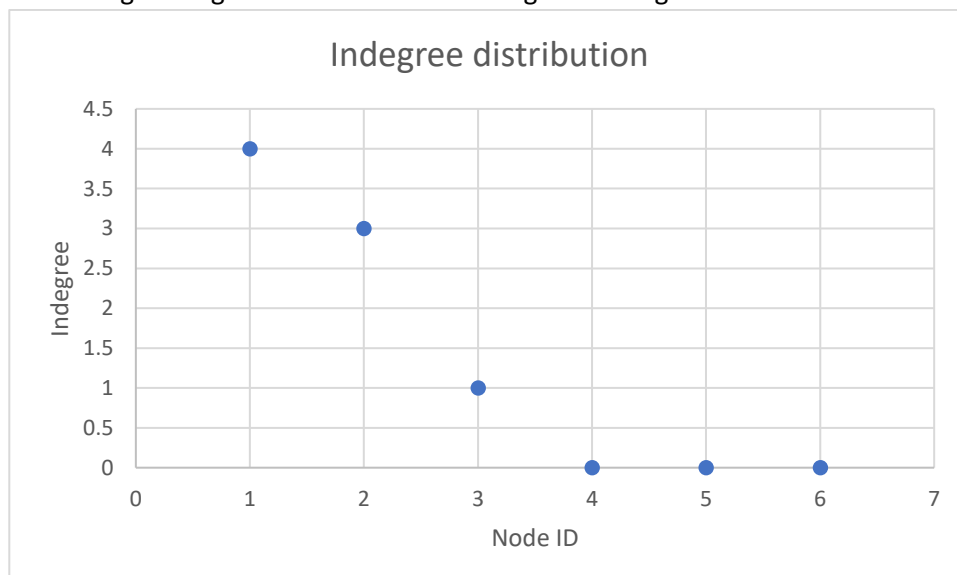


Figure 4: Indegree distribution

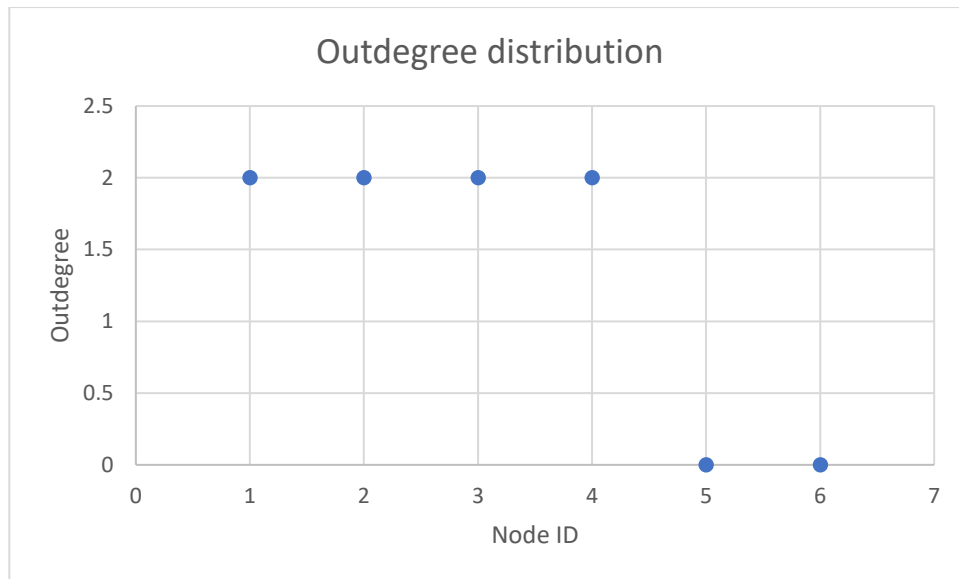


Figure 5: Outdegree distribution

- **Correlation(indegree, outdegree)**

The correlation between the indegree and the outdegree of the nodes is equal to -0.51605. This means that nodes in the network either tend to have a higher indegree or a higher outdegree.

- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 4
Minimum outdegree	= 0
Maximum outdegree	= 2

Distance statistics

- **Average path length**

Note that 3 out of the 6 nodes in this graph cannot be reached from any other node in the graph. For the nodes that do have a path between them, the average path length is equal to 1.273.

Clustering

- **Clustering coefficient**

The clustering coefficient of this network is 0.317. This is a relatively high clustering coefficient in comparison with the other networks that will be discussed in this report.

- **Network density**

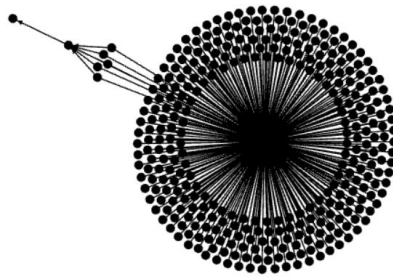
The network density of this graph is equal to 0.533. Thereby, this is the most 'complete' network of the discussed networks in this report.

Centrality

- **Betweenness centrality**

The node in the middle of the network is the only node that creates paths between different other nodes. Therefore, this is the only node with a betweenness centrality larger than 0. The node reports a betweenness centrality of 3 because it is on the unique shortest path between 3 different pairs of nodes.

Analysis: network originated from the 8th most important node



General information

- Estimated time period = 01/2009 – 04/2009
- Amount of transactions = 10.000
- Iterations = 10
- Amount of nodes = 337
- Amount of edges = 341
- Lines from txin file = 620
- Lines from txout file = 10.100

Description

Immediately, one node with a huge indegree catches the attention. This node is probably an address which was set up to mine the blocks quickly in order to set up the bitcoin. At this point in time, the hashes were quite short. It was common that only the first 8 characters of the hash needed to be solved into zeroes, resulting in only little combinations that had to be checked. As the bitcoin became a more mature currency though, these lengths increased to about 18 characters. This way, it became much harder to solve the hashes.

Vertex degrees

- **Degree distribution**
Considering the indegrees of the nodes, there is 1 node with a huge indegree of 335. Almost all other nodes report an indegree of 0. Considering the outdegree, almost every node in the network has an outdegree of 1. These nodes refer to the node with the high indegree.
- **Correlation(indegree, outdegree)**
The correlation between the indegrees and the outdegrees is -0.35516. This negative value means that nodes with a higher outdegrees tend to have lower indegrees. This conclusion was expected when looking at this graph.
- **Degree minima and maxima**
 - Minimum indegree = 0
 - Maximum indegree = 335
 - Minimum outdegree = 0
 - Maximum outdegree = 2

Distance statistics

- **Average path length**

Once again, most of the nodes cannot be reached by any of the other nodes. Therefore the average path length is computed for the nodes that can reach each other. This value is equal to 1.014.

Clustering

- **Clustering coefficient**

The clustering coefficient is 0.008, meaning that the neighbours of a certain node are not each others' neighbours.

- **Network density**

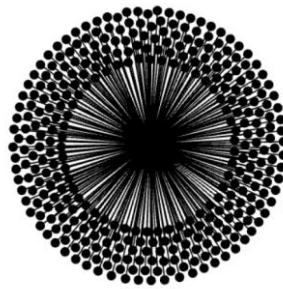
Given the low clustering coefficient, a low network density is expected as well. Due to the fact that there are almost no triangles, the network density is equal to 0.006.

Centrality

- **Betweenness centrality**

Since there are almost no paths between non-neighbouring nodes in this network, the betweenness centrality was not reported for this graph.

Analysis: network originated from the 10th most important node



General information

- Estimated time period = 01/2009 – 04/2009
- Amount of transactions = 10.000
- Iterations = 4
- Amount of nodes = 335
- Amount of edges = 334
- Lines from txin file = 620
- Lines from txout file = 10.100

Description

Just like in the previous network, the central node is probably an address that was set up to focus on mining the blocks. This way, the mining of the blocks could happen quickly in the early phases of bitcoin.

Vertex degrees

- **Degree distribution**
There is one node with a very big indegree, the other nodes all have an outdegree of 1.
- **Correlation(indegree, outdegree)**
The correlation between the indegrees and the outdegrees of the nodes is -1, meaning that a node either has an indegree or an outdegree.

Distance statistics

- **Average path length**
This metric doesn't really add value to the analysis of this graph.

Clustering

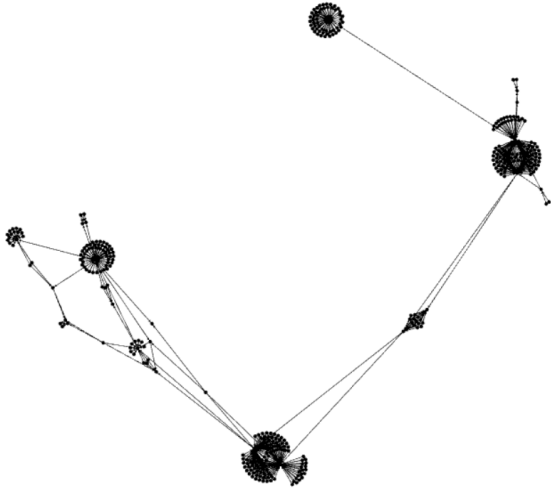
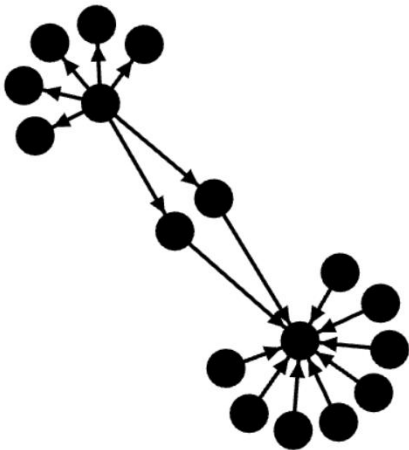
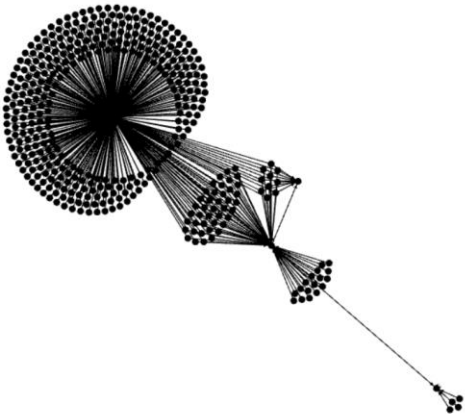

- **Clustering coefficient**
The clustering coefficient is 0. All the nodes that are referring to this particular node in the centre are not interlinked with each other.
- **Network density**
The network density is equal to 0.005, meaning that this is far from a complete graph.

Centrality

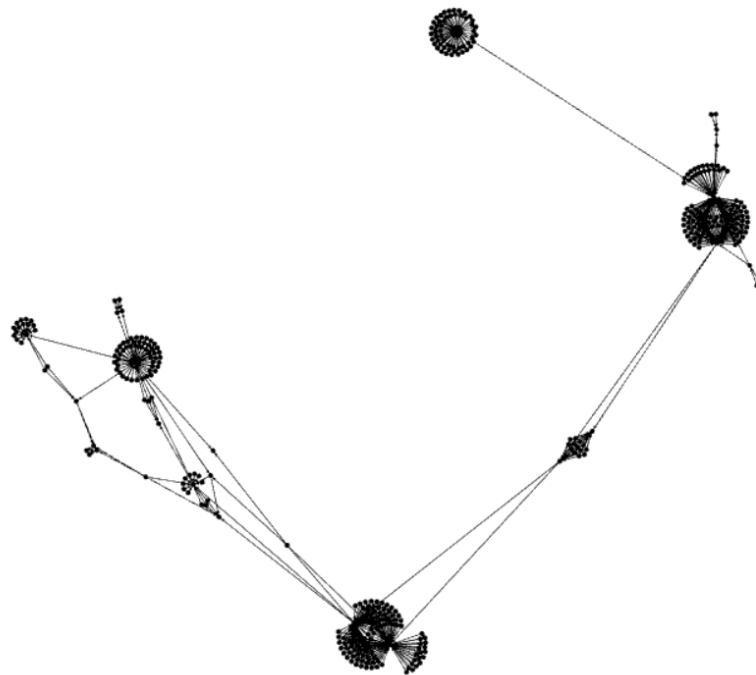
- **Betweenness centrality**
Since there are almost no paths between non-neighbouring nodes in this network, the betweenness centrality was not reported for this graph.

Networks after 50.000 transactions (1 year and 2 months)

The following table illustrates the networks that are originated after 1 year and 2 months. Important to note here is that these networks apparently still struggle with the mismatch between the input and output file: the algorithm can only grow networks around 5 of the 10 most popular nodes. The network that is grown around the 6th most popular node is the same as the network that was grown around the most popular node after 10.000 transactions. For this reason, this section only discusses the 3 other graphs that were originated in this phase.

4th & 5th node (same network)	6th node
	
7th node	9th node
	

Analysis: network originated from the 4th most important node



General information

- | | |
|--------------------------|---------------------|
| • Estimated time period | = 01/2009 – 03/2010 |
| • Amount of transactions | = 10.000 |
| • Iterations | = 4 |
| • Amount of nodes | = 368 |
| • Amount of edges | = 555 |
| • Lines from txin file | = 8.950 |
| • Lines from txout file | = 50.200 |

Description

At a first glance, the networks seem to grow a bit in comparison with the graphs that were grown after 10.000 transactions. Although, after 50.000 transactions one would expect to grow even bigger networks than the networks that are reported in the table above. The mismatch between the input and the output file could maybe explain this phenomenon. The input file reaches the 50.000th transaction already after reading 8.950 lines. The output file seems to report every transaction since extracting 50.000 transactions boils down to reading in approximately 50.000 lines.

Vertex degrees

- **Degree distribution**

The indegree distribution is shown in figure 6. This graph would suggest that the network is transitioning into a scale free network. For this reason, the indegree distribution was also plotted on a log scale in figure 7. Since the analysis of the network in later phases will result in a scale free network, this period was probably the phase where the network started to become scale free. The outdegree distribution is drawn in figure 8. Apparently, a lot of nodes have a very low outdegree in this graph.

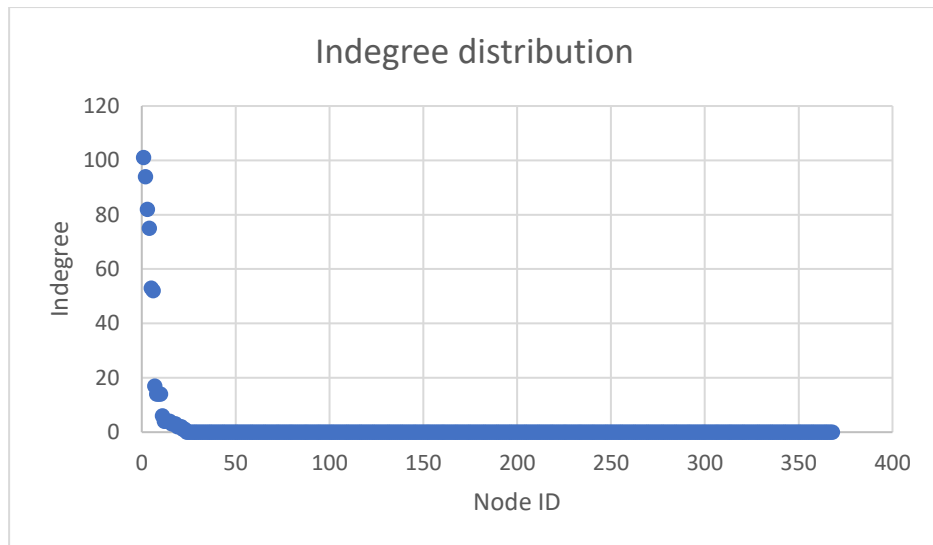
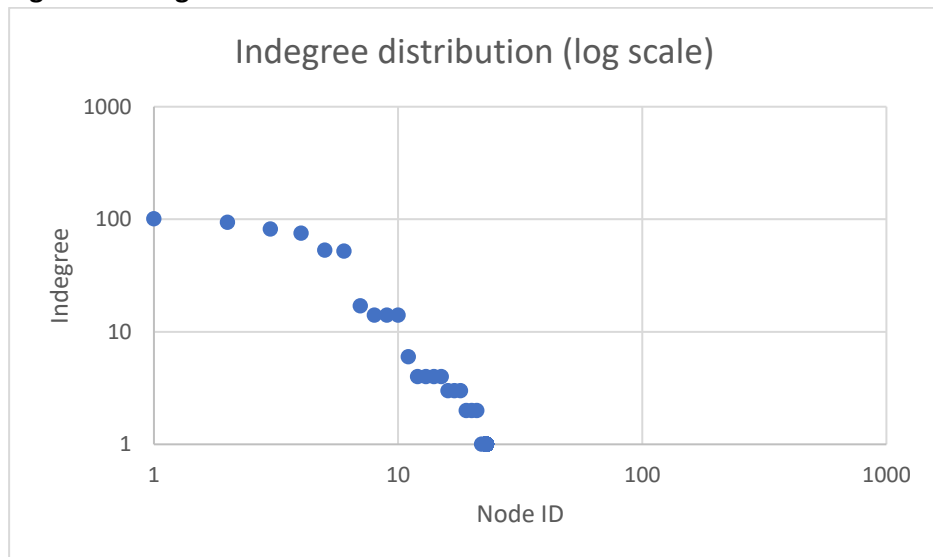


Figure 6: Indegree distribution



- **Correlation(indegree, outdegree)**

The correlation between the indegrees and the outdegrees is 0.098123. Remind that this correlation was negative when the smaller networks were considered. As it turns out, nodes that have a higher indegree also tend to have a higher outdegree at this point. Probably, the network will evolve into a network where nodes with a higher indegree also have a higher outdegree. This is normal, cause we expect active participants in the network to both send and receive transactions.

- **Average vertex degree**

The average indegree and average outdegree is equal to 1.508152. This means that on average a participant in the blockchain isn't really involved in many transactions.

- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 101
Minimum outdegree	= 0
Maximum outdegree	= 6

The node with the indegree of 101 has an outdegree of 0. For this reason, this node is probably a miner who already approved some blocks and was involved in no further transactions. Another possibility could be that a lot of parties just paid some coins to this node for an unknown reason. Further investigation into the exact transaction volume could bring more clarity here.

Distance statistics

- **Average path length**

For the nodes that are connected by a path, the average path length is 3.366.

Clustering

- **Clustering coefficient**

The clustering coefficient is equal to 0.002. This is a very low clustering coefficient. Looking at the network, one could see the nodes having a high outdegree as hubs. These hubs don't really create highly interconnected clusters around them since the parties they are connected with are typically not interconnected.

- **Network density**

The connectivity of this graph is equal to 0.00821. This is also very low and gives an indication of how complete the graph is, or how incomplete the graph is in this case

Centrality

- **Betweenness centrality**

Considering the nodes that can be reached by each other, some of the nodes appear more frequently on the shortest path between 2 other nodes. Figure 9 illustrates the betweenness centrality distribution of this graph.

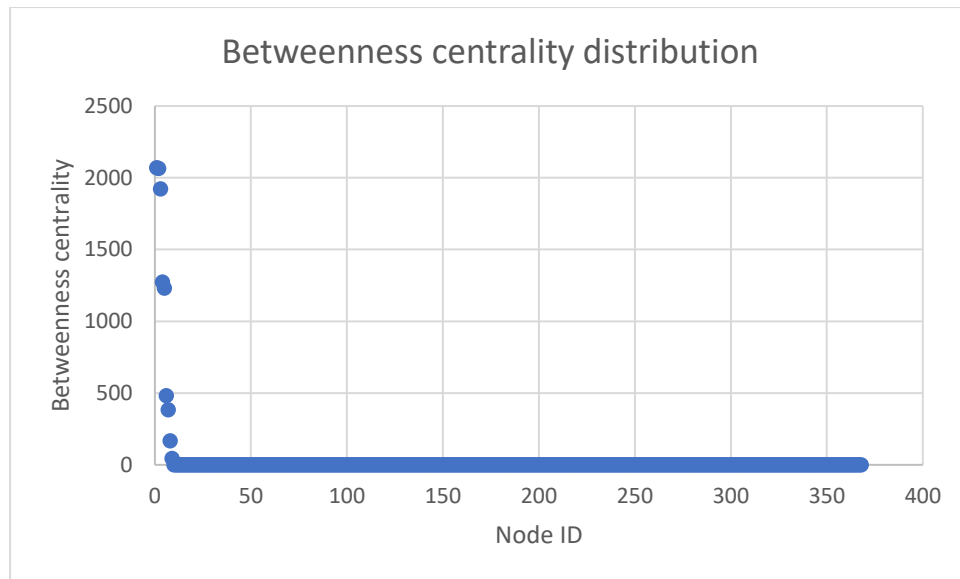
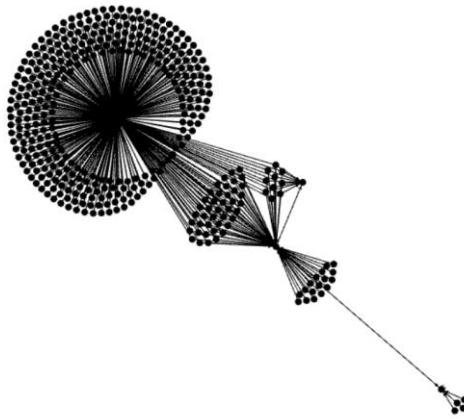


Figure 9: Betweenness centrality distribution

Analysis: network originated from the 6th most important node

This network is the same as the network that was generated by the most popular node in the previous section.

Analysis: network originated from the 7th most important node



General information

• Estimated time period	= 01/2009 – 03/2010
• Amount of transactions	= 10.000
• Iterations	= 4
• Amount of nodes	= 400
• Amount of edges	= 463
• Lines from txin file	= 8.950
• Lines from txout file	= 50.200

Description

This network gives the same ‘feel’ as the network that was originated from the 8th most popular node after 10.000 transactions. Once again, it seems like the algorithm found an important miner here since the most popular node in this graph has a huge indegree. Furthermore, the algorithm performed 10 iterations, yet only 4 hops can be made in this network. This means that it doesn’t make sense trying to make this network bigger by doing more than 4 iterations.

Vertex degrees

- **Degree distribution**
The indegree distribution (figure 10) shows a big outlier, this is probably a miner that mined already quite some blocks. In the outdegree distribution (figure 11) the nodes that refer to this miner are represented by the big number of nodes that only have an outdegree of 1.

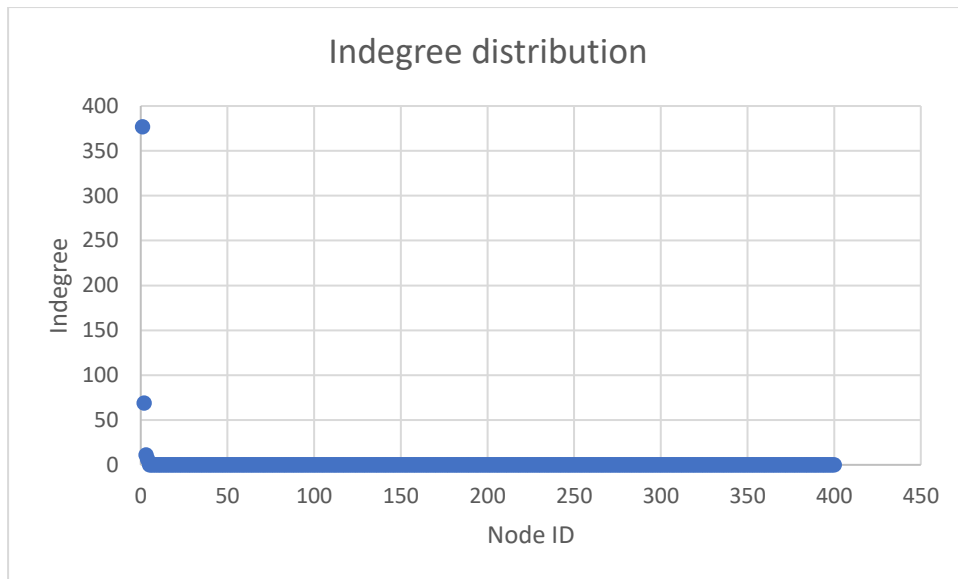


Figure 10: Indegree distribution

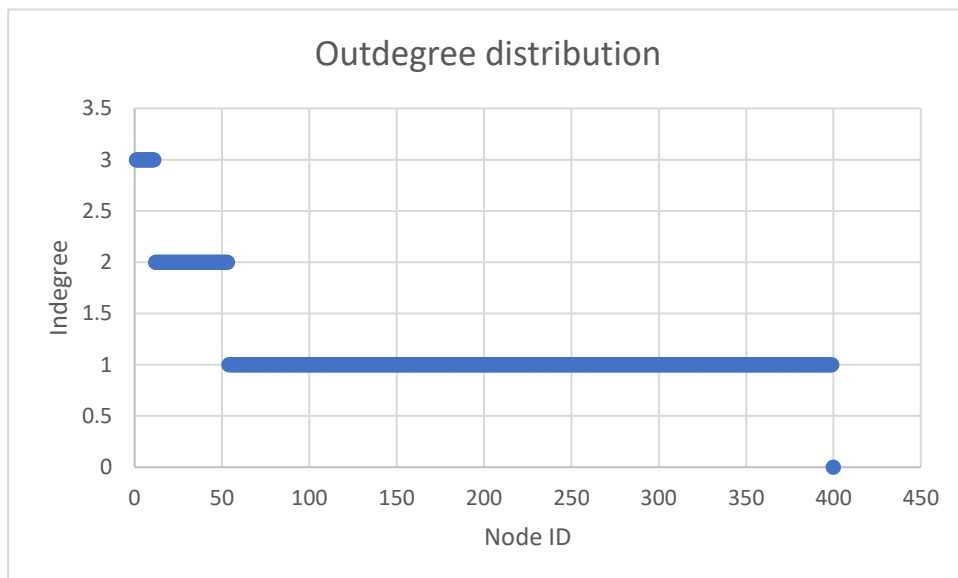


Figure 11: Outdegree distribution

- Correlation(indegree, outdegree)**
 The correlation between the indegrees and the outdegrees of the nodes is -0.092821. This means that, for the sample sets we are examining, we cannot make a clear-cut decision yet whether participants sending more BTC transactions do also receive more transactions and vice versa.

- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 377
Minimum outdegree	= 0
Maximum outdegree	= 3

As discussed above, the algorithm probably found a miner here since a lot of transactions were transferred to one particular node. This could also explain why the network doesn't really propagate as far as the network that was grown around the fourth most popular node since miners can perfectly approve their blocks without necessarily interacting with the community, whereas traders actively send transactions to other parties in the network, making the possibility of spreading the network bigger.

Distance statistics

- **Average path length**
Considering the nodes that can reach each other, the average path length is 1.239.

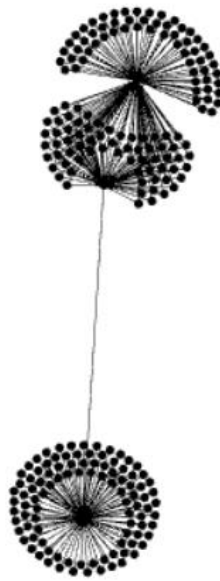
Clustering

- **Clustering coefficient**
The clustering coefficient is 0.066. As for every other inspected graph in this network, this is quite low.
- **Network density**
The network density is 0.005802, conform with values found for previous networks.

Centrality

- **Betweenness centrality**
Plotting the betweenness centrality doesn't really bring a lot of added value in this case. There is 1 node that occurs in the most number shortest paths. These paths are all leading to the most popular node in this network.

Analysis: network originated from the 9th most important node



General information

- Estimated time period = 01/2009 – 03/2010
- Amount of transactions = 10.000
- Iterations = 4
- Amount of nodes = 231
- Amount of edges = 297
- Lines from txin file = 8.950
- Lines from txout file = 50.200

Vertex degrees

- **Degree distribution**
There are only 3 nodes with an indegree. Their indegrees are respectively 128, 101 and 68. These nodes have a very low outdegree. The other nodes have an outdegree of either 1 or 2.
- **Correlation(indegree, outdegree)**
Obviously, the correlation between the indegrees and the out degrees is negative. The actual value is -0.151. We expected this value to be closer to -1 since the nodes with the high indegrees only have an outdegree of 0 or 1, while the other nodes all have an indegree of 0.
- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 128
Minimum outdegree	= 2
Maximum outdegree	= 0

Distance statistics

- **Average path length**

Considering the nodes that can reach each other, the average path length is 1.515.

Clustering

- **Clustering coefficient**

The clustering coefficient is equal to 0.145.

- **Network density**

The network density amounts 0.006.

Centrality

- **Betweenness centrality**

Two nodes are on the shortest paths between some of the nodes that can be reached by each other. Therefore, they have a relatively high betweenness centrality. The other nodes all have a betweenness centrality of 0, simply due to the fact that there are not many paths to be drawn between different nodes.

Networks after 70.000 transactions (1 year and 6 months)

At this point, the exchange node exists for about 3 months. For this reason, this part of the analysis tried to grasp the short-term effects of the bitcoin exchange on the network. The following table provides a graphical representation of the network that was grown at different points in time going forward from this period. However, we believed that the node we were growing our network around (node 50215) was the exchange node, but later analysis indicated otherwise. This node is represented by the 'big black dot' that dominates the image of the graphs. Although, showing these networks is relevant since they represent the process of how the actual bitcoin exchange was found. Based on the graphical representations, one can already conclude that the indegree of the observed node grows very fast, even to a certain point where the area around the node becomes one big black dot. However, the observed node doesn't really connect with many of the other nodes in the network. Both phenomena can be explained.

The fast growing indegree hints that this node was probably not the exchange node but an early miner. This miner could have been set up by the BTC developers to speed up the process of getting bitcoin started. Another explanation could be that someone heard about bitcoin and decided to dedicate a server to the mining of the blocks. This means that this miner isn't necessarily owned by the bitcoin developers.

If the miner was owned by the BTC developers, the decision was probably made to choose for speed over decentralisation in the mining process, since the bitcoin wasn't really used by the masses at this point. This way, no time would be lost while hashing the blocks so the network could grow faster. Analysing the bh.dat file also backs this claim. This file contains more data about the blocks and reports their hashes. Apparently, in the beginning of the bitcoin the lengths of the hashes were way shorter than they are nowadays. In the beginning, only the first 8 tokens of the hash had to be solved into zeroes, resulting in little combinations that had to be checked. Nowadays, lengths of 18 tokens are quite common, resulting in way more combinations. This evolution of the hash lengths also indicates that the bitcoin developers probably wanted to start up the bitcoin network as efficiently as possible. Therefore, hash lengths were shortened, i.e. easier to solve, and possibly addresses which were dedicated to solving the hashes of the blocks were also installed. As time passed by, it became more and more important to have a system that is trustworthy. Therefore, the bitcoin developers lengthened the hashes, making them harder to solve.

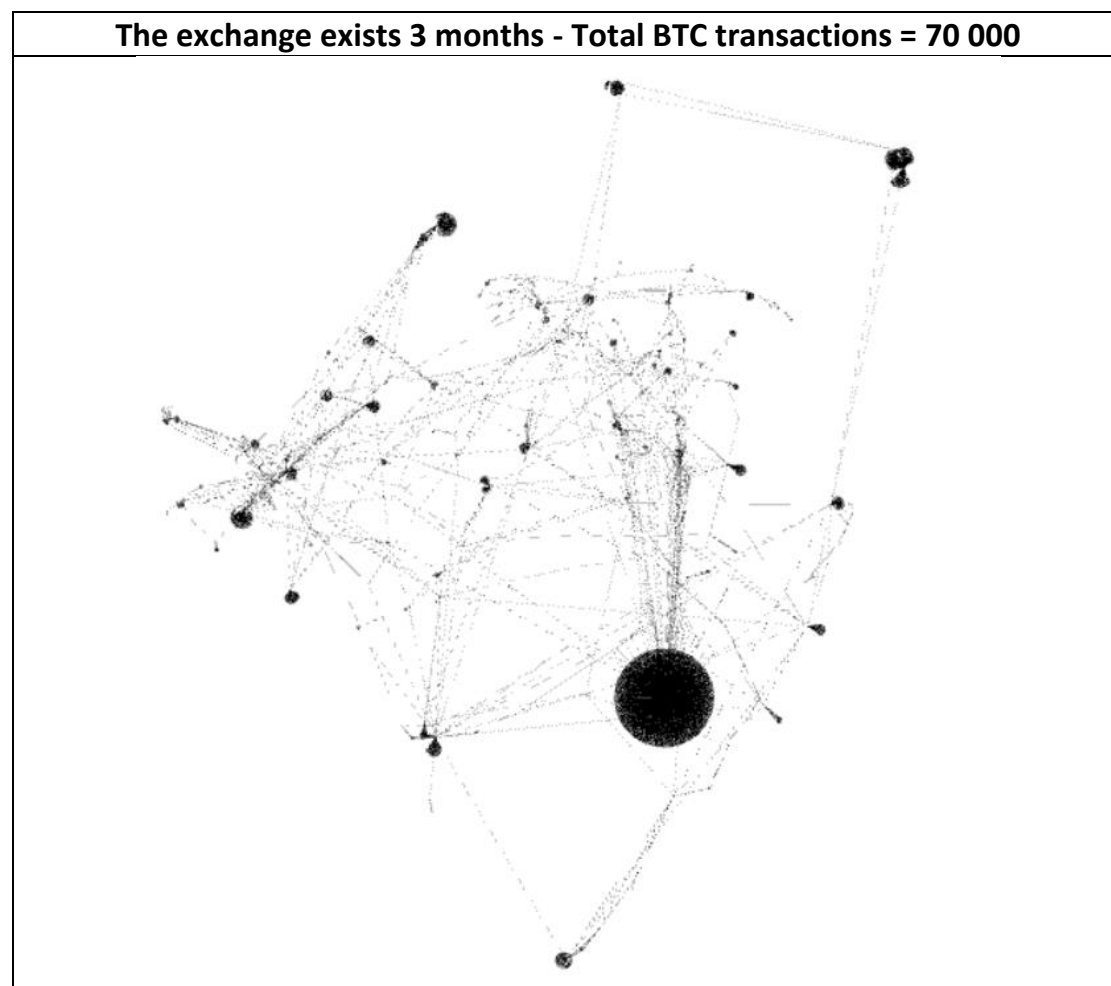
Though, if the miner was not owned by the BTC developers, it is possible that someone heard about bitcoin and decided to dedicate a (idle) server to the mining of the blocks. At this point in time, the bitcoin was nowhere close to the hype it is nowadays, so this person could have been unaware of the fact that this particular blockchain would later become very popular.

In addition to this, it appears that some of these mining addresses are quite successful. Later analysis will show that after 200.000 iterations, 2 mining addresses can be identified as 'black dots'. For this reason, we believe that these mining addresses actually founded the idea of creating the so-called 'mining pools'. Recognising the fact that the hashes would become more difficult to solve, some private parties bundled their forces to mine the blocks together. Furthermore, they invested a lot of money in facilities with a lot of computational power while investigating algorithms that could solve the hashes more efficiently. Therefore, inspired by the initial mining addresses that were quite successful, they developed mining strategies that outperformed the others. Because of this, the mining of the bitcoin blocks became a 'centralised' process after all. Although, the interpretation of a decentralised mining process needs to be refined here. Since, in a perfect environment every miner is mining for themselves, creating a global network of miners who are competing for the rewards.

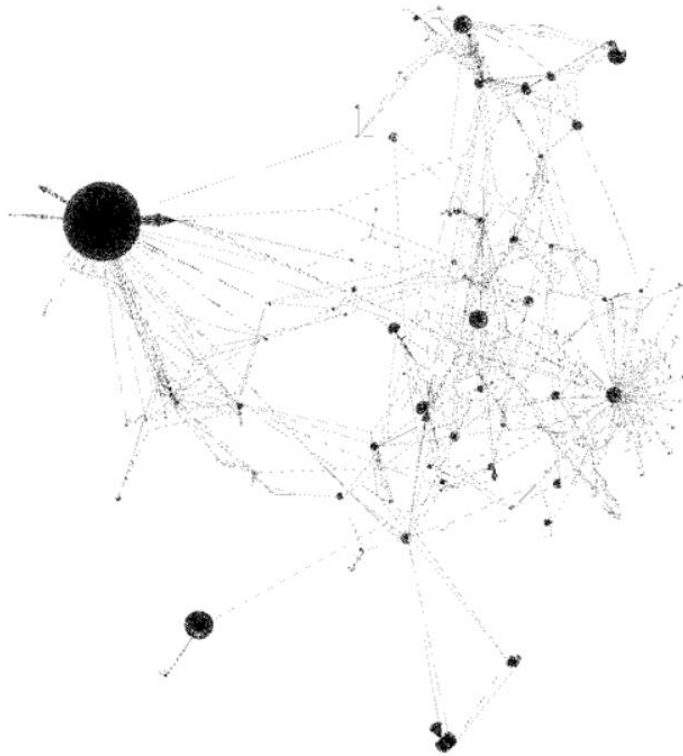
However, the so-called mining pools bundled their forces to mine the blocks together. And, if a mining pool happens to control more than 50% of the network hash power, it could theoretically take over the blockchain, thereby defying the property of decentralisation the bitcoin stands for.

To continue the network analysis aspects, the low outdegree of node 50215 also indicates that this node was probably not the exchange node. Nevertheless, the actual exchange node can be extracted when analysing the degree distributions of the nodes after 100.000 and 200.000 transactions. Namely, the exchange node can be identified as the node with the highest outdegree (node 65393). More importantly, this node reports a similarly high indegree. Because of this, the node appears in the centre of the graph that was actually created around node 50215, the perceived miner. The miner itself is one step away from the centre of this graph. We believe that it is very special to find a certain node in the centre of a graph that was actually built around a different node. Combining these observations, the conclusion was made that node 50215 is a miner while node 65393 is the first bitcoin exchange.

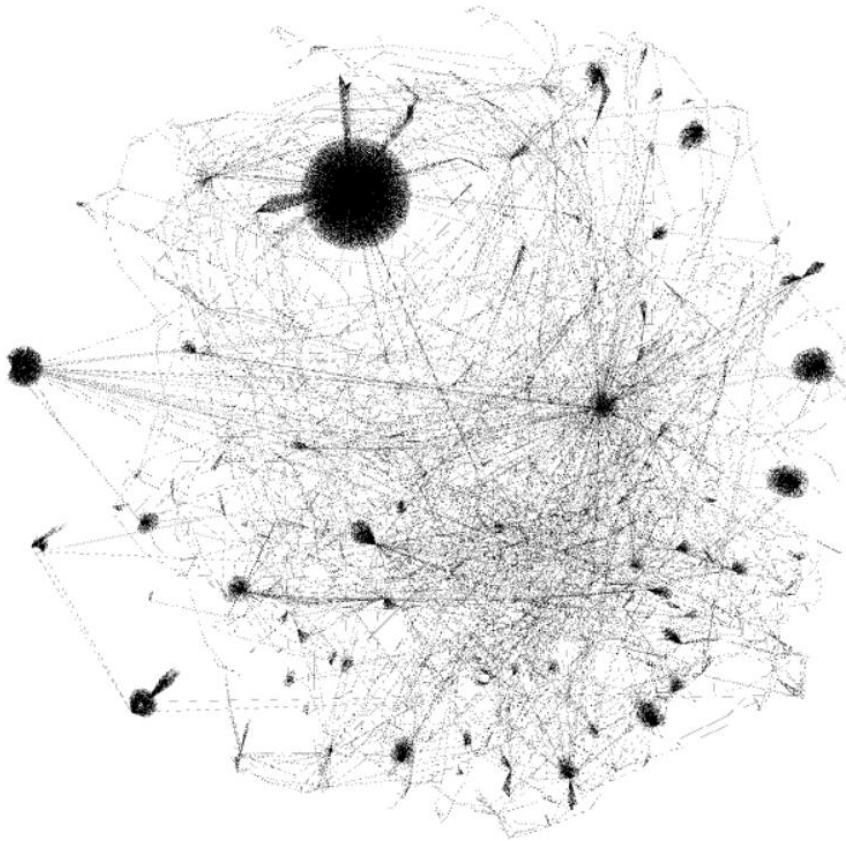
For this reason, the following part offers a short analysis of the network that is originated at 70.000 transactions. The bitcoin exchange was still in its infancy here, so the analysis after some more time might be more interesting. Therefore, the next sections analyse both the exchange node and the miner more intensively.



The exchange exists 4 months - Total BTC transactions = 80 000



The exchange exists 5 months - Total BTC transactions = 100 000



Analysis: network originated from the most popular node

General information

- Time period = 01/2009 – 06/2010
- Amount of transactions = 70.000
- Iterations = 10
- Amount of nodes = 7324
- Amount of edges = 8266
- Lines from txin file = 25400
- Lines from txout file = 74000

Description

The big black dot on the graph is a probably a miner. As discussed in the introduction to this part, this miner could either be an initiative of the BTC developers to speed up the early phase of the network or be a mining server that was ran by a particular person. The analysis of the network after 70.000 transactions is rather concise. The network after 100.000 transactions will be analysed more in detail.

Vertex degrees

- **Degree distribution**

Figure 12 shows the indegree distribution immediately on a logarithmic scale to show the scale free behaviour. The outlier is obviously the miner here. Figure 13 also proves the scale free behaviour for the outdegree. Note that there is no real outlier for the outdegree distribution. Furthermore, a lot of nodes report outdegrees of 0 and 1. A point of critique on bitcoin can be found here: many blockchain participants just bought bitcoins as an investment without actually using it as a payment method. This property will be re-encountered for later periods.

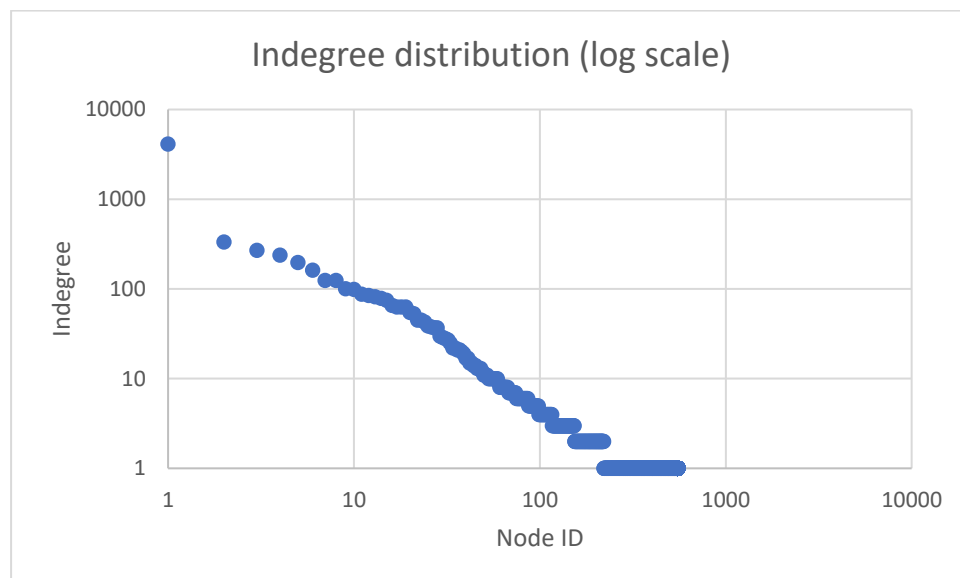


Figure 12: Indegree distribution (log scale)

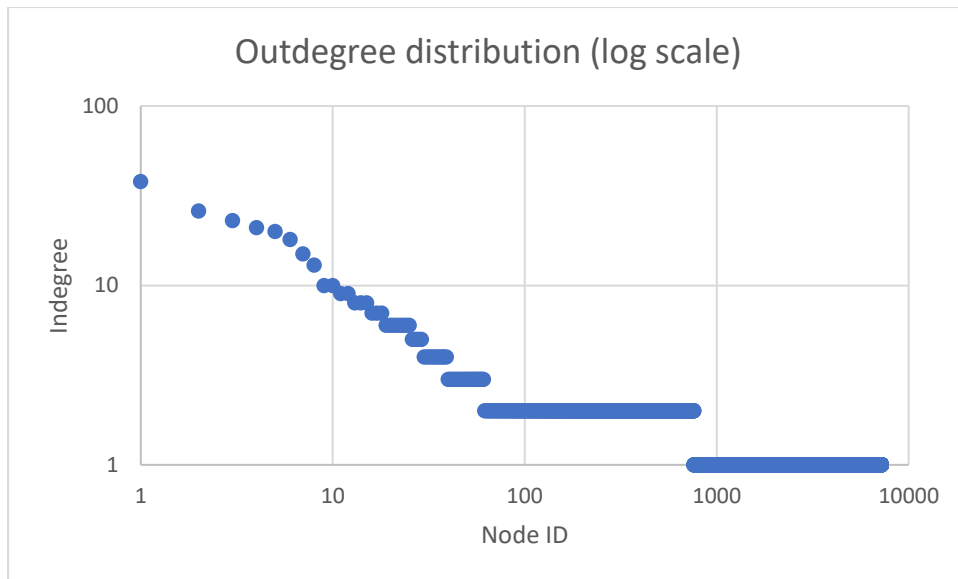


Figure 13: Outdegree distribution (log scale)

- **Correlation(indegree, outdegree)**

The correlation between the indegrees and the outdegrees of the nodes is 0.146406. This means that nodes with high indegrees also tend to have higher outdegrees and vice versa.

- **Average vertex degree**

The average vertex degree is 1.129, meaning that on average, a participant in the bitcoin network isn't involved in transactions with a lot of different other participants.

- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 4151
Minimum outdegree	= 0
Maximum outdegree	= 38

The maximum indegree can be associated with the miner.

Distance statics

- **Average path length (directed)**

Considering the nodes that can be reached by each other, this network has an average path length of 7.776.

- **Average path length (undirected)**

When making abstraction of the directions of the links, a more general average path length can be calculated. The average path length is 4.54 in this case. Note that this does not mean that every node can be reached from every other node in 4.5 steps on average. This is because many nodes cannot be reached at all due to the directions of the links. Interpreting this value of 4.5 boils down to comparing it with the amount of hops the algorithm takes from the exchange node. Given the 10 iterations of the algorithm, 9 hops from the exchange node were made. (The algorithm starts counting from the initialisation of the first node) This results in the fact that the network succeeds to interconnect some important nodes to make the shortest path significantly smaller than 9.

- **Radius (undirected)**

The smallest eccentricity in the network is 7. Note that this eccentricity metric was calculated for the undirected variant of the actual graph. In other words, this does not mean that the nodes in the centre can reach every other node in the network within 7 steps. In terms of transactions, this radius does not mean that an amount of bitcoins can be transferred from the core to every other node in the network in a maximum of 7 steps.

- **Diameter (undirected)**

The diameter is equal to 13. Again, the diameter was calculated for the undirected variant, so the same conclusions as for the radius apply.

Clustering

- **Clustering coefficient**

The average clustering coefficient is 0.011. This is in line with previously encountered clustering coefficients.

- **Network density**

The network density is equal to 0.000308, meaning that this network is represented by a graph that is far from complete.

Centrality

- **Betweenness centrality**

Figure 14 displays the betweenness centrality. Apparently, some of the nodes have a very high betweenness centrality, while most other nodes report a betweenness centrality of 0. Figure 15 provides the betweenness centrality on a logarithmic scale to get some more insight in the specific behaviour. Apparently, only about 400 nodes report a betweenness centrality that is bigger than 0. The miner reports the biggest betweenness centrality on these plots. This can be explained by the high indegree of the node and the low clustering of the network. The low clustering ensures the fact that the miner is in the shortest path from practically every node that refers to the miner to the nodes that are targets of the miner. Given the high indegree of the miner, the amount of shortest paths this node is involved in is very high. The fact that there are also other nodes with similarly high betweenness centralities can be explained by a simple pseudo-algorithm. Namely, the miner refers to some other nodes since he has an outdegree of 8. If these nodes also have an outdegree different from zero, they appear on the shortest paths from the nodes that are attached to the miner to the nodes they refer to due to their outdegree. If those latter nodes also have an outdegree, the path extends further and new nodes with relatively high betweenness centralities are found. Probably, the ability to expand these paths stopped after about 100 nodes were investigated via this pseudo-algorithm. The sudden drop in the betweenness centrality can then be explained by paths that existed in a similar way from smaller clusters in the network. (But the fact that the miner has an indegree of 4159 skyrockets the betweenness centralities of all the nodes that are found from the miner) As a last note here, we want to stress that this pseudo-algorithm does not hold for any network. In networks with a high clustering degree, the probability of omitting a different path between the two observed nodes that might be shorter is very big.

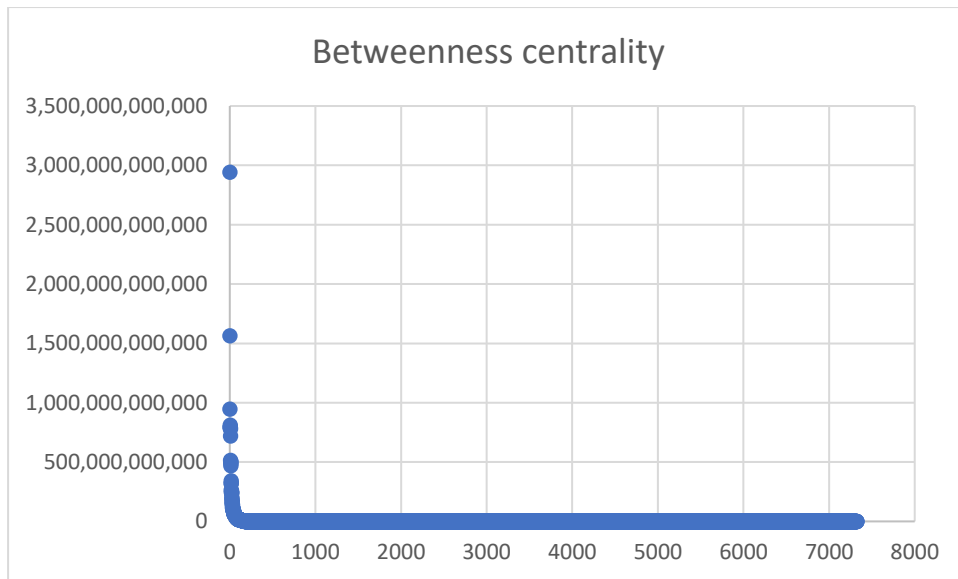


Figure 14: Betweenness centrality plot

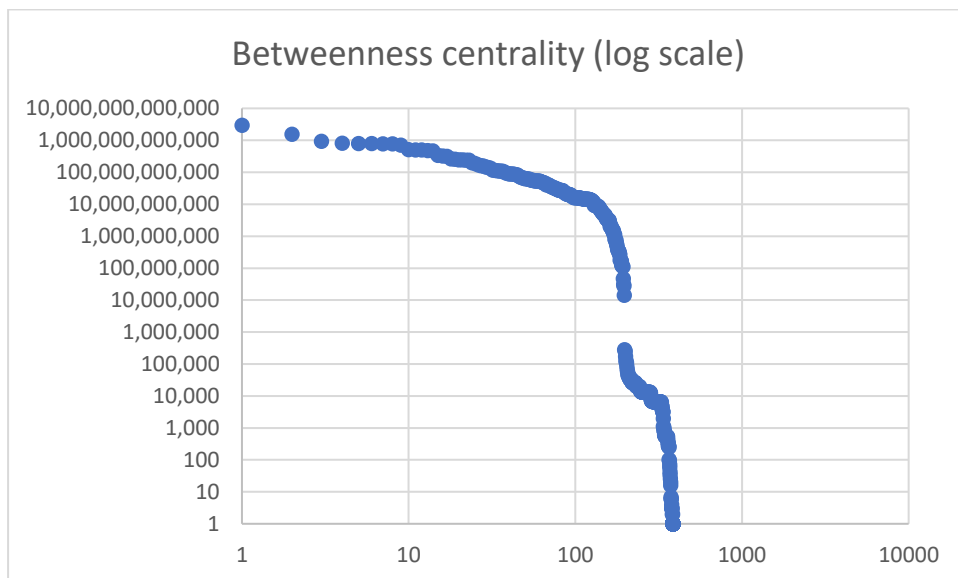


Figure 15: Betweenness centrality plot (log scale)

Page rank

- **Page rank distribution**

Figure 16 displays the page rank distribution of this network. Since, one could see every node as a 'page' and every directed edge as a 'link'. Therefore, the page rank gives an indication towards the importance of the nodes in the network. For the page ranks in this report, the standard settings proposed by the network analysis tool Gephi were used ($p=0.85$, lowest = 0.001). No further optimization about these settings were considered. Considering the page rank distribution, the miner shows up as an outlier with a remarkably high page rank. This page rank is due to the high indegree of the node from mining a lot of blocks. On this graph, the first BTC exchange doesn't seem to be very important yet. However, when analysing bigger datasets, the page rank of the exchange node will 'climb' in the rankings.

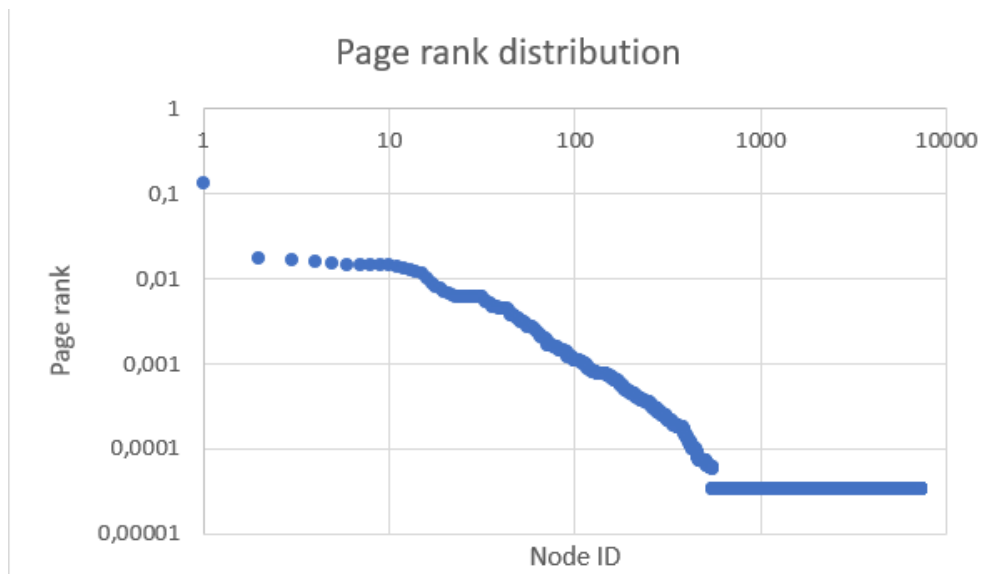
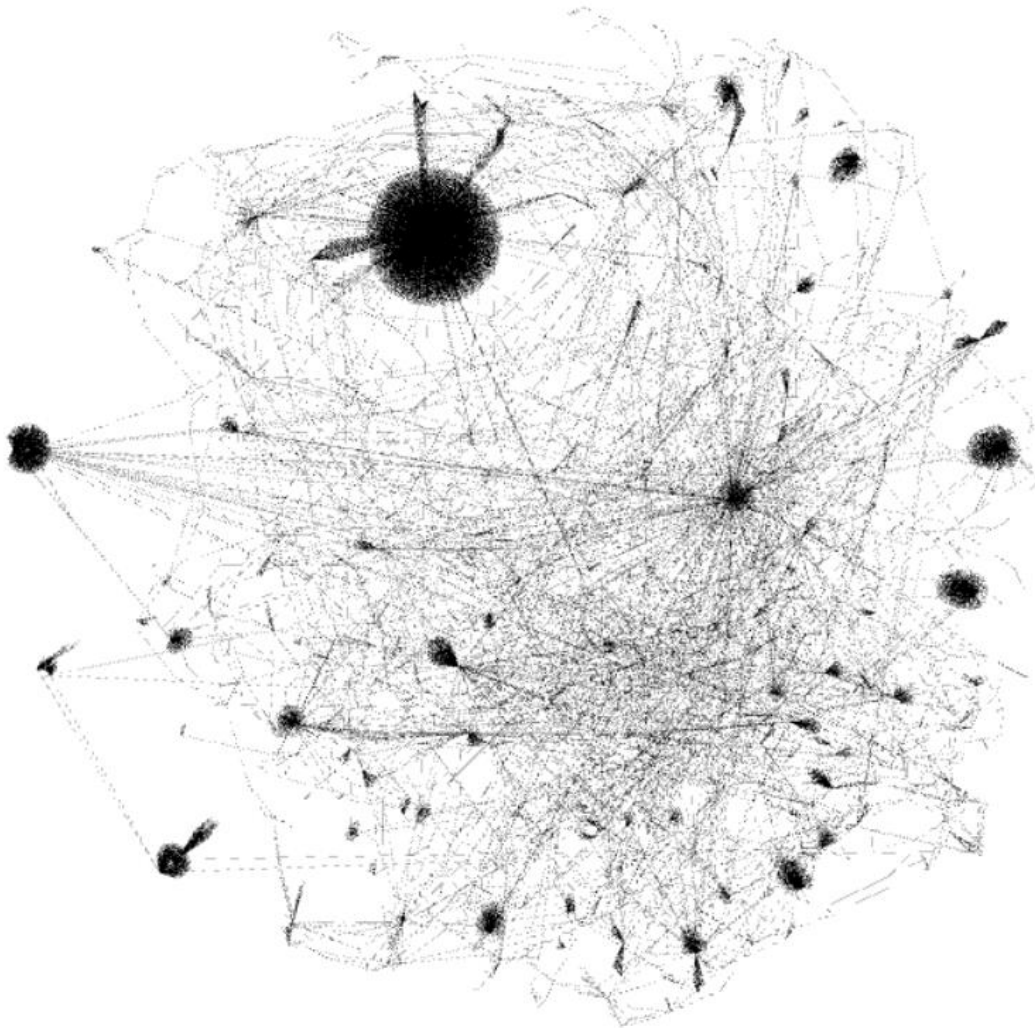


Figure 16: Page rank distribution

Network after 100.000 transactions (1 year and 8 months)

The most interesting part of the analysis is the part where the network starts to get its actual form. This analysis discusses the network after 10.000 transactions in detail.



Analysis: network originated from the most popular node

General information

- | | |
|--------------------------|---------------------|
| • Time period | = 01/2009 – 08/2010 |
| • Amount of transactions | = 100.000 |
| • Iterations | = 10 |
| • Amount of nodes | = 14.469 |
| • Amount of edges | = 18.243 |
| • Lines from txin file | = 58.000 |
| • Lines from txout file | = 117.000 |

Description

In this network, a huge black dot is again reported when displaying the graph. Apparently, this node was the only node having about 4700 transactions in this network. This node was initially perceived to be the first bitcoin exchange, which was started in March 2010. The bitcoin exchange was meant to operate as a central market where one could buy and sell bitcoins. However, degree distributions, behaviour of other nodes in the network and analysis of bigger datasets were contradictory to this assumption. More specifically, the most popular node in this network is probably an address that was set up to mine a lot of the initial blocks. The actual exchange node can easily be spotted on the graphical representation of the network. It is the big node that lies in the middle of the network and from which connections to the entire network are made.

Vertex degrees

- **Degree distribution**

Figure 17 shows the indegree distribution. This distribution shows an outlier with an enormous indegree, the miner. For this reason, figure 18 shows an alternative representation where the miner was left out. These 2 figures already hint towards a scale free behaviour. To get some further insight in the behaviour of the network, figure 19 shows the indegree distribution on a logarithmic scale. An imaginary line could be drawn on this graph, proving the scale free behaviour of this network.

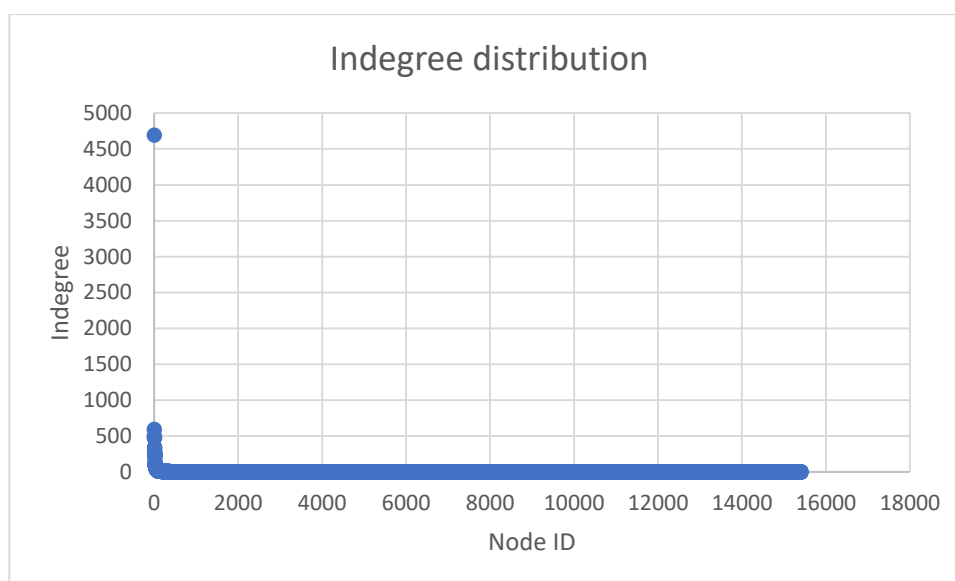


Figure 17: Indegree distribution

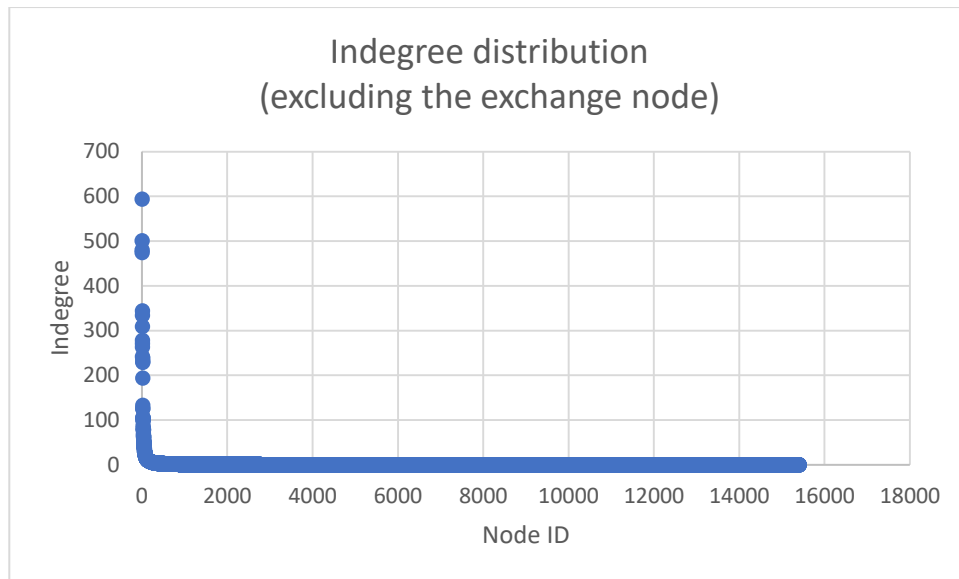


Figure 18: Indegree distribution (excluding the most popular node)

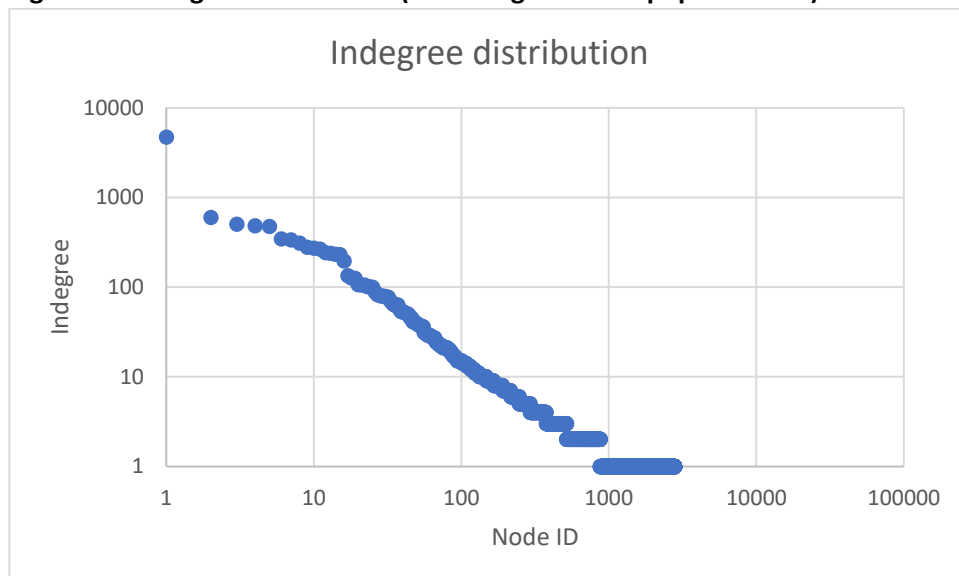


Figure 19: Indegree distribution (log scale)

The outdegree distribution shown in figure 20 also reports an outlier, which can be associated with the exchange node. Notice that the degree of the outlying vertex in the outdegree distribution is much lower than the outlier on the graph of the indegree distribution. Also, the miner is not an outlier on the outdegree plot while the exchange node was not an outlier on the indegree plot. This is in accordance with the expectations, since the miner approved a lot of blocks and got rewarded for doing so. This results in the outlying indegree. At the same time, the exchange node was supposed to provide the network with bitcoins, explaining the outlying outdegree. Plotting the outdegree on a logarithmic scale, figure 21 proves the scale free behaviour once again. Though, a bit more attention should be given to the tails here. Apparently, a lot of nodes only have an outdegree of 1 or 2. This could be explained by people investing in the currency itself or people trying out just one transaction. Another part of these nodes with an outdegree of 1 can be explained by the rewards that are paid to approve the blocks. Furthermore, the fact that more than 10.000 addresses report an outdegree of 0 embeds the critique that is often given to the bitcoin network. Namely, the currency was created to be used for

payments. I.e. 'one should be able to buy a bread using bitcoins.' But apparently, many people just bought some bitcoins as an investment, without actually trading with the coins.

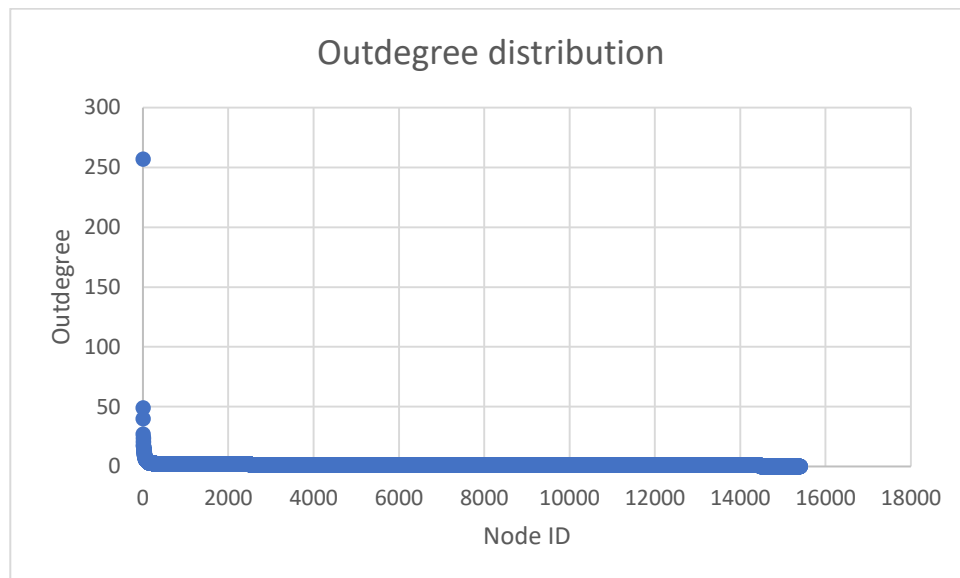


Figure 20: Outdegree distribution

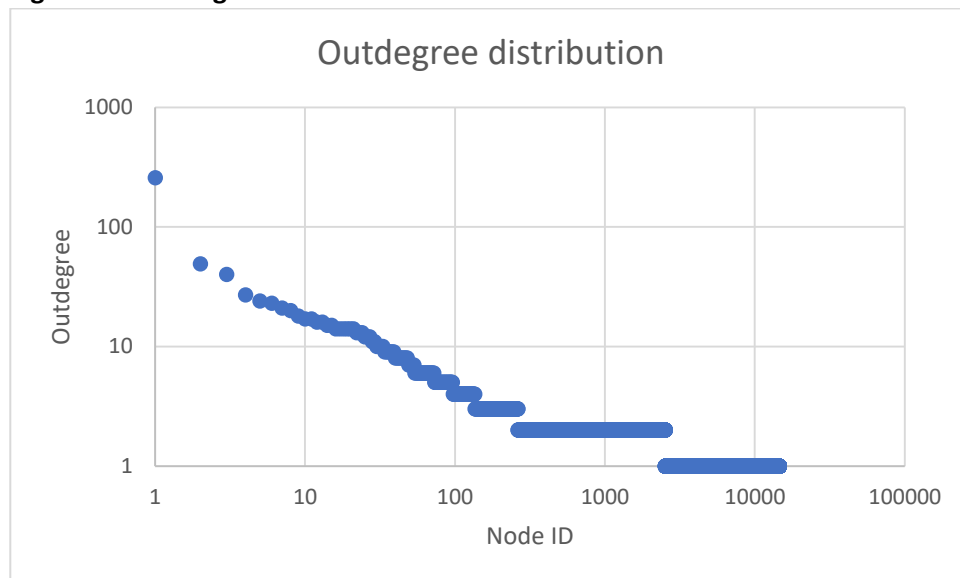


Figure 21: Outdegree distribution (logarithmic scale)

- **Correlation(indegree, outdegree)**

Note that this statistic measures something different than the often reported 'degree correlation'. The correlation between the indegree and outdegree reports whether members with a high indegree tend to have a high outdegree as well and vice versa. In this case, a correlation of 0.134417 is reported. This means that nodes with higher indegrees also tend to have higher outdegrees. This is not surprising, since an active participant in the bitcoin network will probably both receive and send out coins to other members while trading. At the same time, members who buy bitcoins solely as an investment will be characterised by low indegrees and outdegrees, thereby explaining the positive correlation.

- **Average vertex degree**

The average vertex degree is still quite low with a value of 1.18. Once again, this indicates that there are some nodes being very active on the bitcoin network, but a lot of the members are not very active at all. For this reason, they bring the average vertex degree down to almost 1. Although the bitcoin is still in its infancy at this moment, some criticism to bitcoin can be found again here: the amount of people actually using the currency for payments, which was its intended function, is quite low. In reality, many people just try out one transaction, or they do some speculation on the currency. This is reflected in the very low average vertex degree, meaning that on average, a person isn't really involved in many transactions. The nodes that are involved in many transactions are the traders and the exchange node. This behaviour will remain a problem for the bitcoin network.

- **Degree minima and maxima**

Minimum indegree	= 0
Maximum indegree	= 4692
Minimum outdegree	= 0
Maximum outdegree	= 257

The node having an indegree of 4692 can be identified as the important miner. The address ID of this node is 50215. The node with ID 65393 has the highest outdegree and is the exchange node. This exchange node was started in March 2010, when the amount of transactions on the chain was about 50.000. In other words, the address ID of this exchange node should be originated somewhere around this time frame. For this reason, we looked up if this was true and we checked the size of the exchange node at 50.000 transactions. The exchange node was indeed originated around this time frame. The graph after 100.000 transactions shows that this exchange node has a reasonably high indegree and outdegree, therefore connecting well with the rest of the network. In other words, the exchange node had been started up very effectively, since the users of the bitcoin network immediately connected with it and started using the exchange node for its intended purpose.

Distance statics

- **Average path length (directed)**

Bearing in mind the analysis of the previous networks, we can assume that many nodes cannot be reached in this graph as well. This assumption is also backed by the big number of vertices we find with an indegree or outdegree of 1. These nodes, amongst others, cannot be reached from anywhere in the graph. For the nodes that can be reached, the average path length is 8.323.

- **Average path length (undirected)**

When making abstraction of the directions of the links, a more general average path length can be calculated. The average path length is 5.801 in this case. Note that this does not mean that every node can be reached from every other node in 5.8 steps on average. This is because many nodes cannot be reached at all due to the directions of the links. Interpreting this value of 5.8 boils down to comparing it with the amount of hops the algorithm takes from the exchange node. Given the 10 iterations of the algorithm, 9 hops from the exchange node were made. (The algorithm starts counting from the initialisation

of the first node) This results in the fact that the network succeeds to interconnect some important nodes to make the shortest path significantly smaller than 9.

- **Eccentricity (undirected)**

This metric was also computed by making abstraction of the directed links. The eccentricity for the undirected graph is shown in figure 22. This figure shows that only some of the nodes are in the centre of the network. Table 2 gives a summary of these nodes in the centre of the network. Adding the indegrees and outdegrees to the table, one node that immediately catches the attention is the node with address ID 65393. Observing the fact that this node has the highest outdegree in the network and the fact that the indegree is also significantly higher than the indegrees of most other nodes, this node probably plays a crucial role in this network.

For this reason, the history of bitcoin was looked up to see whether some special events happened around the time when this node was originated. Strikingly, the first bitcoin exchange was originated around the same time frame. This bitcoin exchange was set up as a central market place where one would be able to buy and sell bitcoins. The behaviour of node 65393 fulfils this property in a perfect manner. However, identifying node 65393 as the bitcoin exchange rendered node 50215 unexplained – since we believed node 50215 to be the exchange node in the beginning. Therefore, the decision was made to look further into this node.

Previous networks already reported some nodes that were probably miners in the early phase of the network. These nodes were characterised by a high indegree and the difficulty to grow a network around them. Node 50215 shows the exact same behaviour: the node has a very high indegree but is not very heavily connected to the rest of the network. Therefore, the assumption that this node was an important miner in this phase of the bitcoin network was made.

Furthermore, we found a directed path from this node to the exchange node through an intermediary node. Figure 23 provides a general overview of the path between the miner and the exchange node. The miner is the lowest hub on the figure here while the exchange node lies in the upper left corner. Even though it is quite tedious to show the exact directions of the arrows, we included some screenshots to prove our point (making the arrows bigger didn't work in Gephi). The directed path from the miner towards the intermediary node is shown on figure 24. The path from this intermediary node towards the exchange node is illustrated on figure 25. Note that the arrow does not point towards the intermediary node here, so the direction must be towards the exchange node.

Two possible explanations for this behaviour could be assumed. The first reasoning for this phenomenon goes as follows: possibly, the bitcoin developers set up some addresses to mine the initial blocks. Though, the mining of those blocks granted them some rewards. Since the exchange node needs bitcoins to distribute them to the network, there had to be some kind of mechanism to transfer the collected mining fees towards the exchange node. This was done through the intermediary node, which represents a kind of 'wallet' of the exchange node. Namely, exchange nodes often have different 'wallets' where money is stored to decentralise the money supply. For this reason, the mining node

probably transferred the money towards this wallet, which was then claimed by the exchange node when needed.

A second explanation involves a miner that is not owned by the BTC developers. Possibly, an early BTC enthusiast decided to dedicate a server to mine some blocks. This assumes that this person had an idle server in its possession. The transactions with the BTC exchange could then be explained by trading the mining rewards with the bitcoin exchange through a wallet.

To make a clear-cut decision about what is actually happening here, the history of bitcoin should be investigated in further detail. However, finding information about the first years of the blockchain's existence is rather hard since bitcoin wasn't very popular at this time yet.

A short analysis of the network created after 200.000 transactions will be discussed in the next section. The most important finding after 200.000 transactions is the appearance of a second mining node which reports a similar behaviour. The mining node stores some of its fees into wallets, which are then transferred to the wallet of the exchange node. For this reason, the next section discusses this behaviour in further detail.

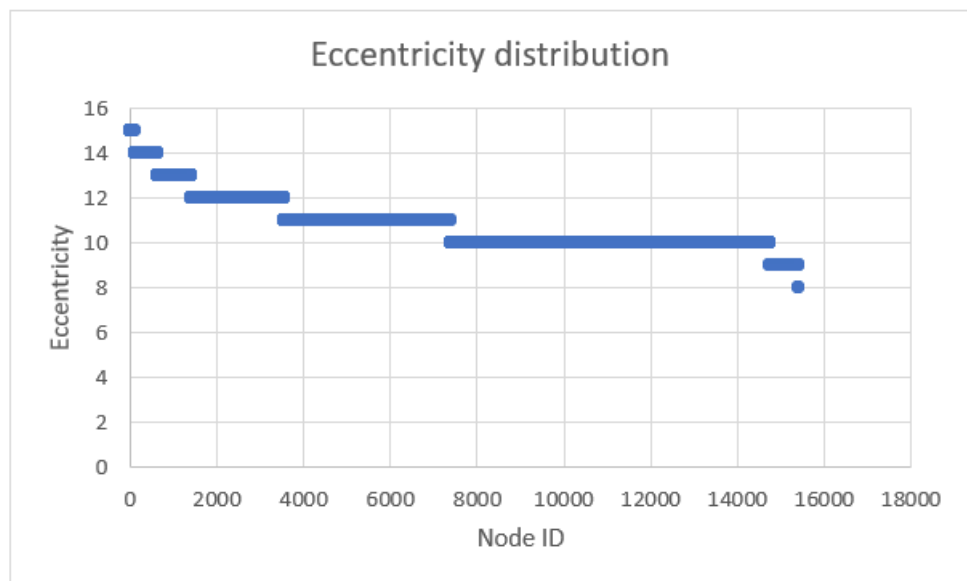


Figure 22: Eccentricity distribution

Number		Address ID	Degree	Eccentricity	Indegree	Outdegree
4763		40601	89	8	80	9
4977		65393	499	8	242	257
5159		86439	3	8	1	2
5202		86130	18	8	10	8
5724		73372	12	8	9	3

Table 2: centre of the network



Figure 23: The path between the miner and the exchange

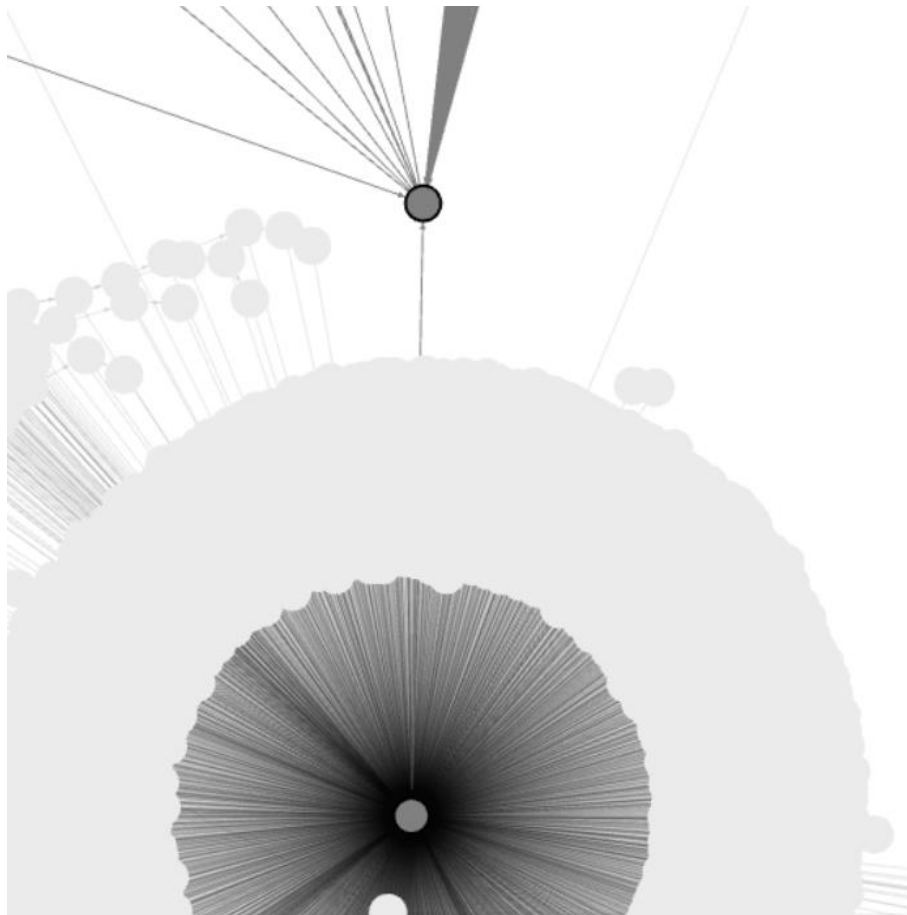


Figure 24: Directed path from the miner towards the intermediary node

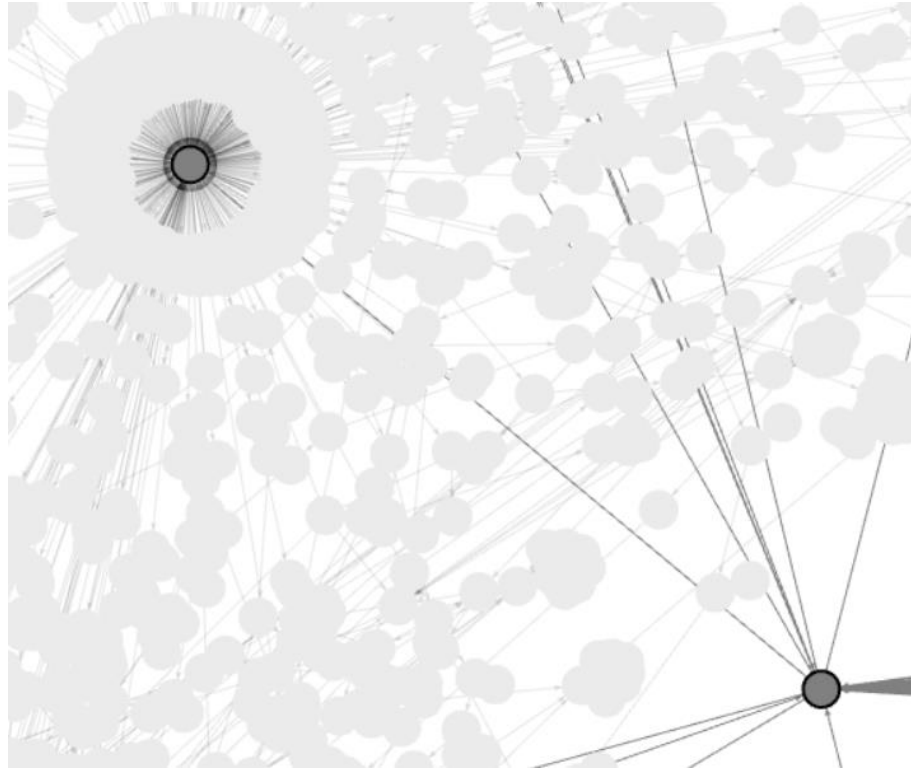


Figure 25: Directed path from the intermediary towards the bitcoin exchange

- **Radius (undirected)**

The smallest eccentricity in the network is 8. Since 5 nodes report this eccentricity, the centre of the network consists of these 5 nodes. (see table 2) Note that this eccentricity metric was calculated for the undirected variant of the actual graph. In other words, this does not mean that the nodes in the centre can reach every other node in the network within 8 steps. In terms of transactions, this radius does not mean that an amount of bitcoins can be transferred from the core to every other node in the network in a maximum of 8 steps.

- **Diameter (undirected)**

The diameter of the network can be distilled from figure 22 and is equal to 15. Again, the diameter was calculated for the undirected variant, so the same conclusions as for the radius apply.

Clustering

- **Clustering coefficient**

The clustering coefficient is 0.032. This is in line with the clustering coefficients we found earlier and this means that in general there is not a lot of clustering going on in this network. Although, when looking at the visual representation of the graph, one can detect some major hubs as 'black dots'. The reason why this clustering coefficient stays very low is because the dots around these hubs are not interconnected with each other. In other words, some hubs have a lot of neighbours, but the neighbours are often not each other's neighbours. In terms of bitcoin terminology, some traders are very active on the network, but the parties they trade with aren't interconnected with each other to a high degree. Once again, the fact that the currency is not mainly being used to perform payments can

be pinpointed as a major reason for this behaviour. Since one can imagine that in a ‘town where everyone uses bitcoins’ the interconnectedness will be very high. Therefore, we believe that if the currency ever manages to be used as a payment instrument, the expected interconnectedness in these clusters will be higher. This could possibly happen as the bitcoin grows to a more mature currency.

- **Network density**

The network density measures the degree to which the network is a complete graph. Noticing the fact that the graph has 14.469 nodes and 18.243 edges, a low network density is expected. This is because a network with 14.469 nodes would require 104.448.746 edges in order to be complete. The network density of this bitcoin network is 0.0001574.

Centrality

- **Betweenness centrality**

Looking at the visualisation of the network, the betweenness centrality plot given in figure 26 was expected. Some nodes function as hubs and because of this reason, they are on the shortest path between various other nodes. The actual Excel tables add some more flavour to the interpretation of these betweenness centralities though. As discussed above, following from the huge indegree, the miner had the highest betweenness centrality in the network after 70.000 transactions. But one could question whether the property of having both a high indegree and outdegree would create an even higher betweenness centrality for the exchange node. Since, the exchange node has way fewer other nodes referring to him, but the high outdegree increases the probability to find a lot of directed paths in the network. Therefore, table 3 brings the juice to make the right decision here. Apparently, the exchange node overtook the miner in terms of betweenness centrality, meaning that in terms of betweenness this node is now the most important node in the network. Finally, analysing figure 27 ends the conclusion for the betweenness centrality. One can assume that roughly 400 nodes can be found when expanding directed paths from the exchange node and miner node. The corresponding pseudo-algorithm was already explained in the previous section and will not be repeated here. After these 400 nodes, the betweenness centrality drops very quickly, just like in the previous section and can be explained by paths that exist in a similar way from smaller clusters in the network.

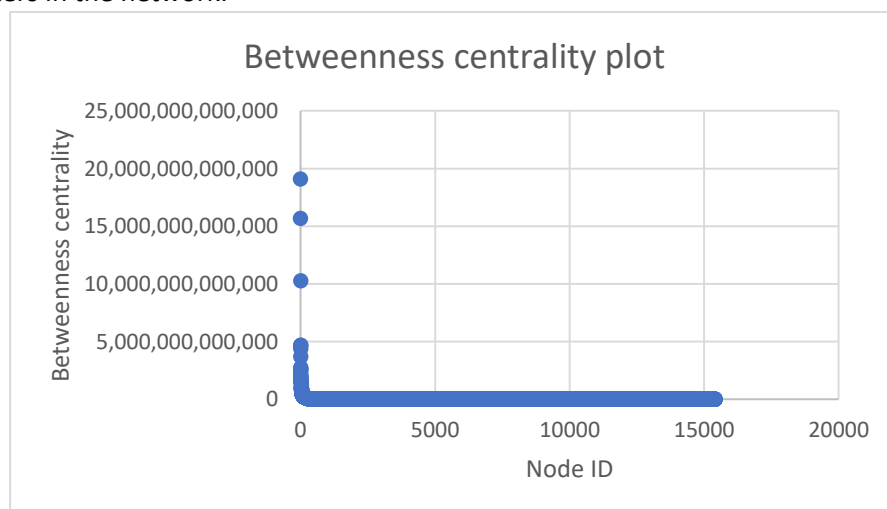


Figure 26: Betweenness centrality plot

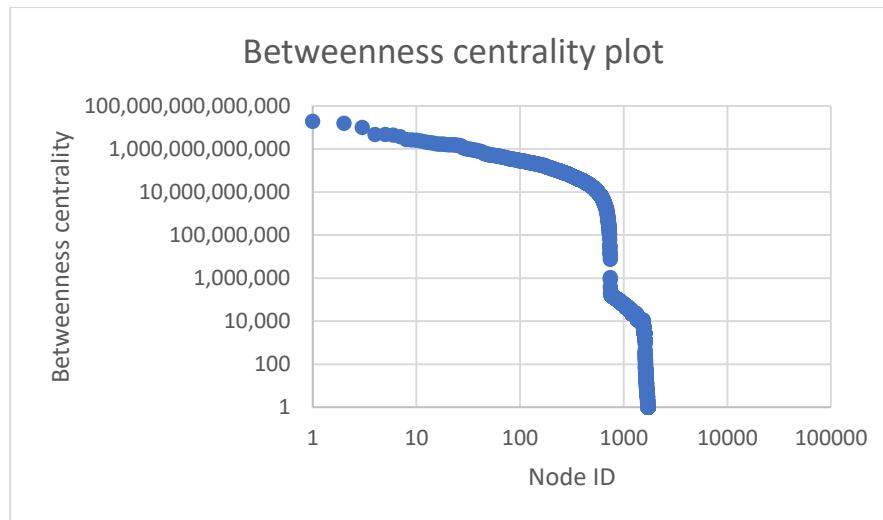


Figure 27: Betweenness centrality plot (log scale)

Id	Label	indegree	outdegree	Betweenness centrality
4977	65393	242	257	19.092.503.482.001
1	50215	4692	17	15.698.622.049.812

Table 3: Betweenness centrality table

Page rank

- **Page rank distribution**

On figure 28, it seems like the page rank distribution slowly becomes a linear plot when using a logarithmic scale. The outlier with the highest page rank is the miner here. The tail is represented by the many nodes with an indegree of 0. Many of these nodes refer to the miner. Note that, however being the most important node in terms of betweenness, the exchange node is not the most important node according to the betweenness centrality. This could be explained by the extraordinary high indegree of the miner, making the miner seem like a more important node according to the page rank algorithm.

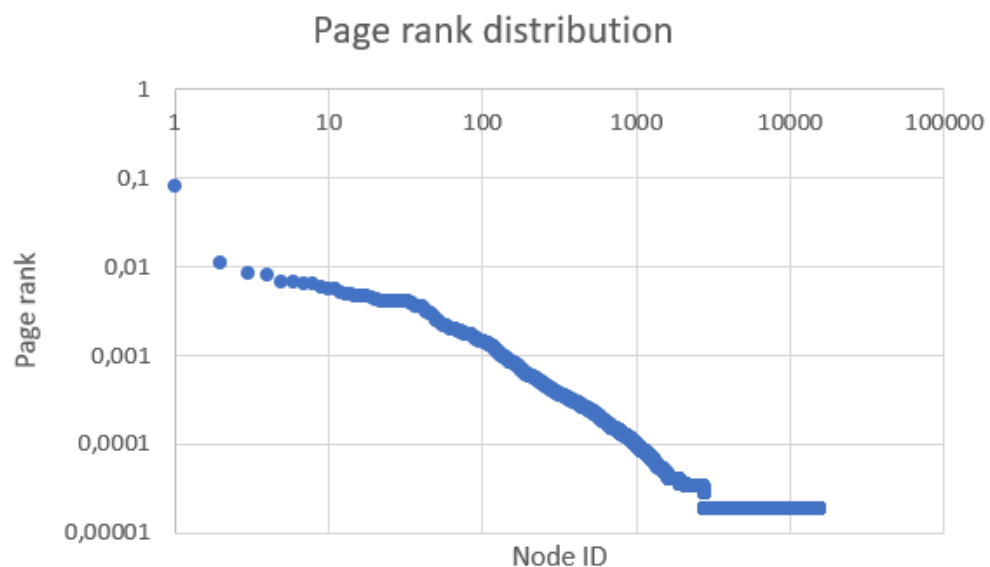
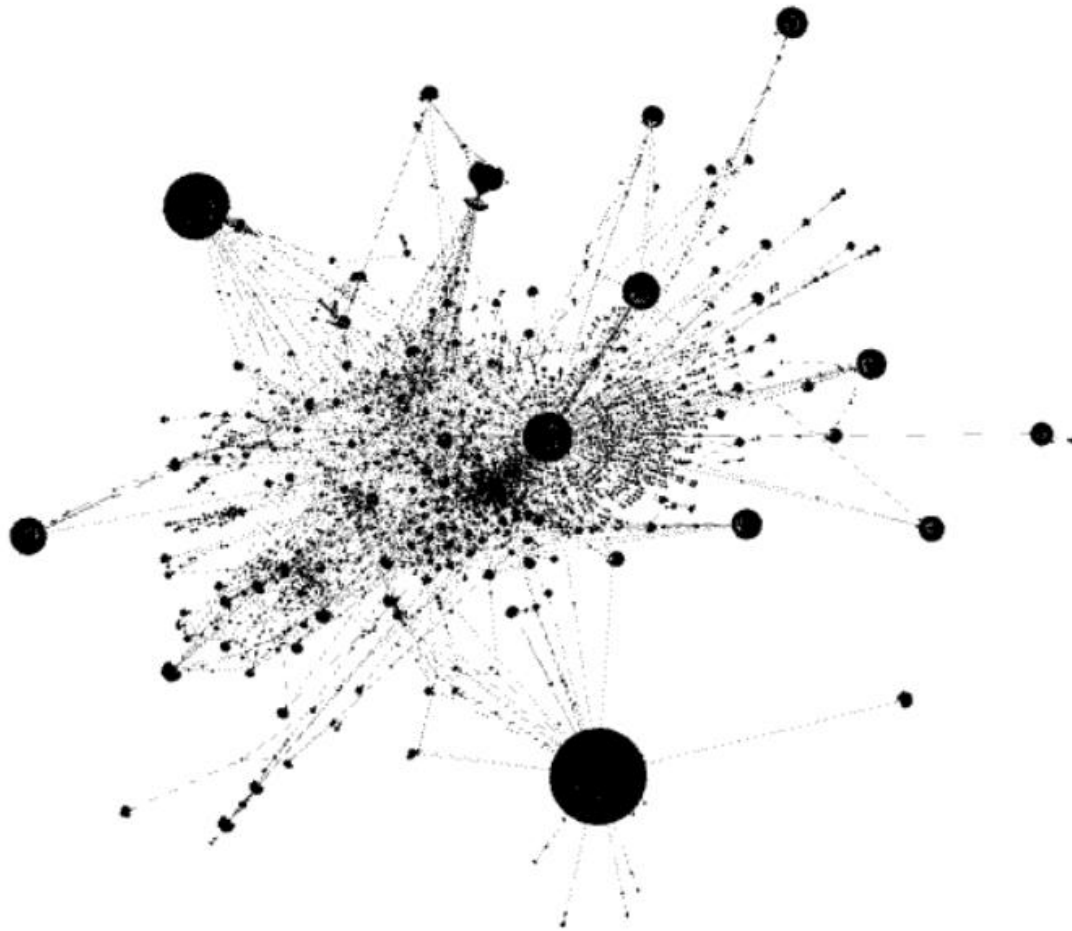


Figure 28: Page rank distribution

Networks after 200.000 transactions (2 years)

This graph represents the bitcoin network after 200.000 transactions. The bitcoin exists 2 years now. On this graph, 3 nodes immediately catch the attention. First of all, the bitcoin exchange can be spotted right in the middle of the graph. The exchange node is very well connected to many of the nodes in the network. Furthermore, there are 2 nodes represented by a very high indegree, the biggest one being the miner we already encountered after 100.000 transactions. The second one is probably also a miner that was created later as the network grew bigger. This assumption is made due to the similar behaviour of those 2 nodes. This similarity will be discussed in this part of the report.



Mining behaviour

To cut to the chase, this network most likely has 2 addresses with the mining of the blocks as their sole purpose. Again, this could be explained by 2 possible scenarios.

The first scenario assumes that these nodes were initialised by the bitcoin developers in order to grow the network rapidly in the early stage. The reasoning goes as follows: Figure 29 shows the neighbours of the exchange node. Some of them are traders, others are most likely different wallets of the exchange node. This mechanism of using wallets could have been used to keep the money supply artificially decentralised. Furthermore, figure 30 shows that the miner with ID 50215 is connected with one of these neighbours of the exchange node, probably a wallet of the exchange node. Even more important, the direction of the path goes from the miner, through the wallet, towards the exchange node. Therefore, this scenario implies that the bitcoin developers who started up these mining addresses transferred the mining rewards from the blocks they mined with this address towards the exchange node. The history of the bitcoin lines up with this idea since the first 'mining pool' was originated way later than this node with ID 50215. In other words, this mining node was most likely in possession of the BTC developers. Realising the fact that this node generated some revenue in the form of mining rewards, they most likely decided to transfer these coins towards their bitcoin exchange node.

Looking at the second mining node, the similarity in behaviour is striking. The second node also reports directed paths towards the exchange node. This is illustrated in figure 32. For this node however, transactions are first transferred to multiple different nodes, probably to decentralise the traffic. This was probably done to make the transaction volumes harder to notice. One could assume that these nodes are the 'wallets' of the mining node. Then, all these wallets contain directed paths towards a wallet of the exchange node. This can be seen on figure 31. From this wallet, the direction of the coins is towards the exchange node again. So in a nutshell, the second node probably mines blocks, stores the rewards into different wallets which send the coins to a wallet of the exchange node. This way, the earned rewards reach the bitcoin exchange again.

However, the second scenario assumes that these miners were not necessarily owned by the BTC developers themselves. Since any early bitcoin enthusiast could have had an idle server and could have decided to use it to mine the blocks with. In this scenario, the interpretation of the directed paths changes a bit. Now, the miner trades the mining rewards with the bitcoin exchange through the use of different wallets.

Concluding the behaviour of both mining nodes, these nodes probably accelerated the initialisation of the bitcoin network in some kind of way due to their ability to mine a lot of blocks. The reason why these addresses could mine so many blocks can be explained by the fact that computer servers have a way bigger combined computational power than the power of an individual miner. Consequently, the probability that a server mines a block is bigger than for an individual miner. Therefore, dedicating a server to solve hashes with a short hash length can perfectly explain these high indegrees.

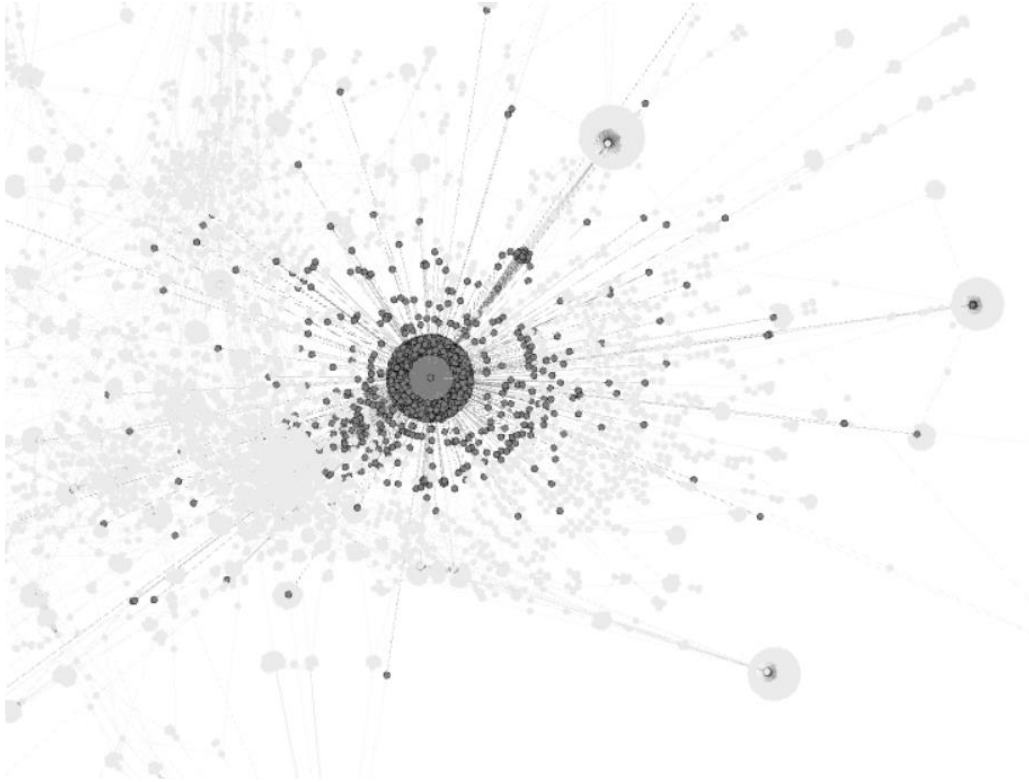


Figure 29: The neighbours of the exchange node

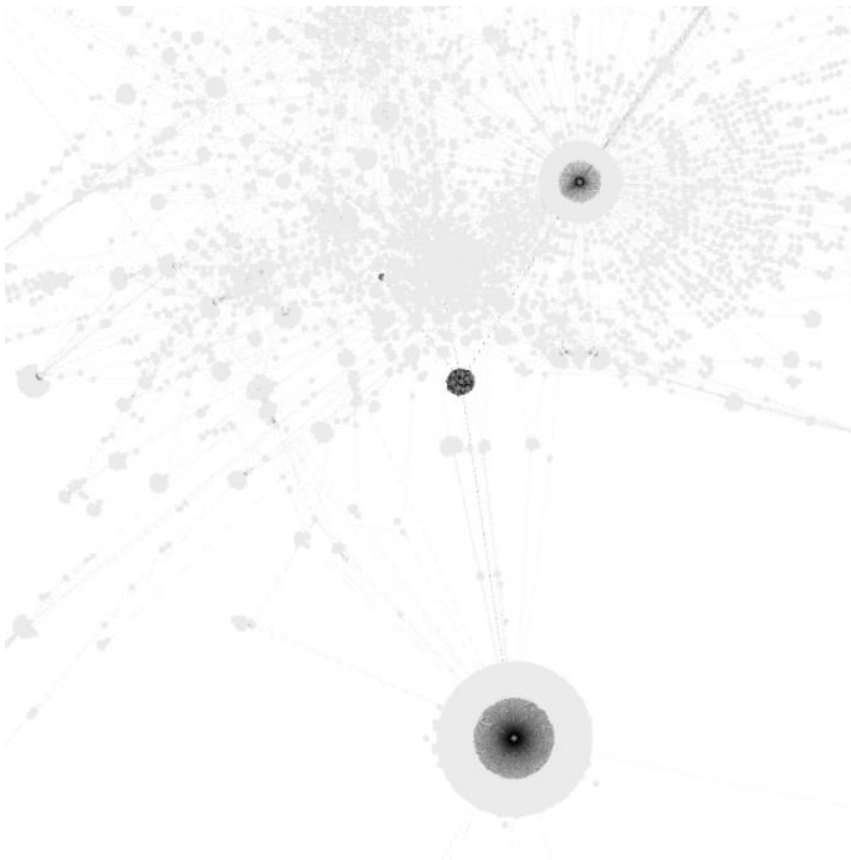


Figure 30: The miner puts BTC in one of the wallets of the exchange node



Figure 31: The neighbours of the exchange node

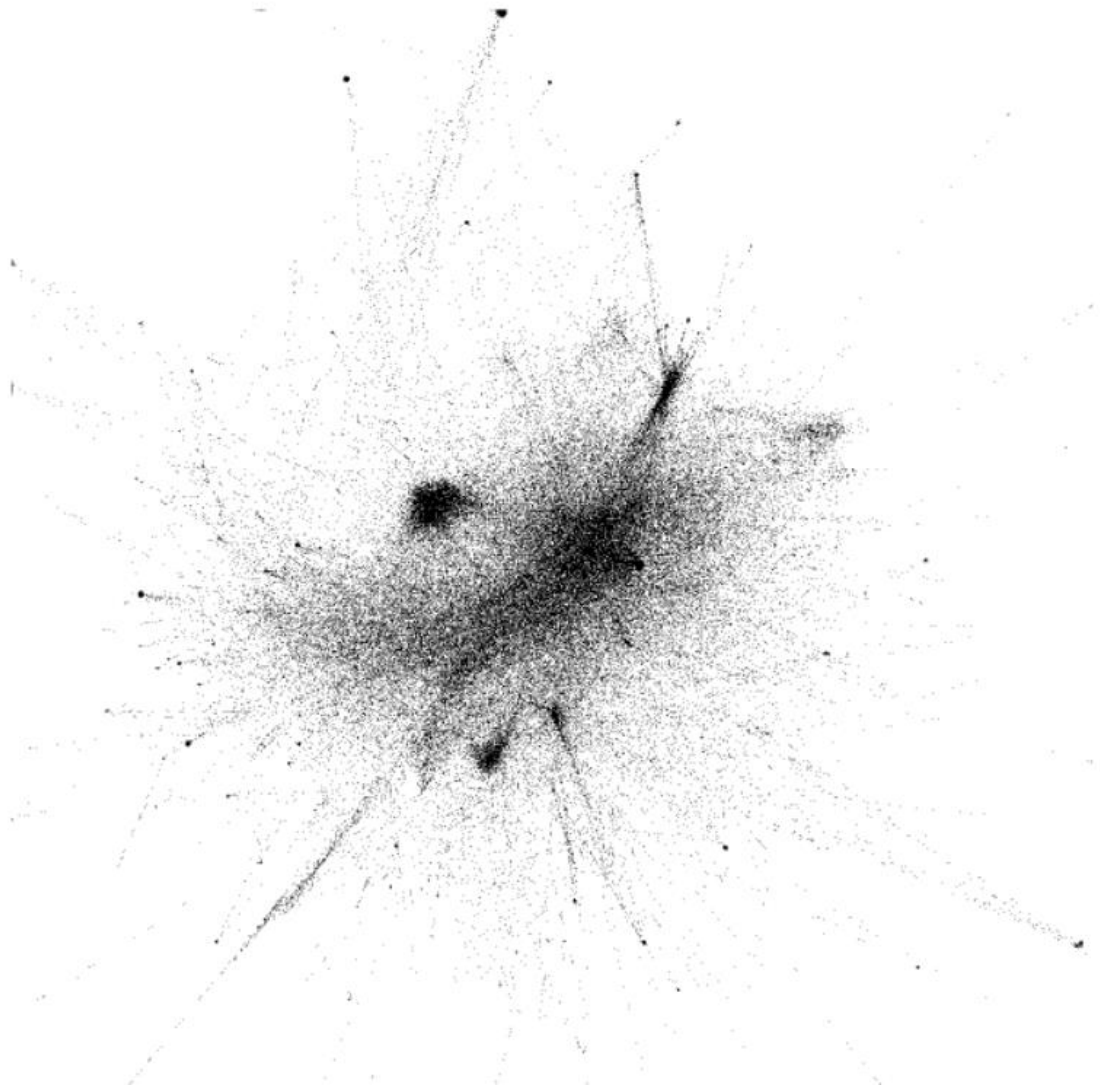


Figure 32: The miner puts money in his wallets

Network after 500.000 transactions (2 years and 5 months)

Slush Pool

In an attempt to find the first mining pool, Slush Pool, the algorithm was used to analyse a dataset of 500.000 transactions. The network exists 2 years and 5 months now. However, extracting Slush Pool from these data is harder than one would think. Since, mining pools allow everyone to join the pool so the blocks aren't necessarily mined by the same address. Either way, from a network analysis point of view it is still interesting to report the key statistics for this network.



Vertex degrees

- **Degree distribution**

The indegree distribution is perfectly scale-free. (See figure 33) On the outdegree distribution, 2 outliers are reported. (See figure 34) One of them can be identified as the first bitcoin exchange. The second one is probably the second bitcoin exchange since this node also has a very high indegree. Therefore, the node shows the same behaviour as the first BTC exchange. Furthermore, the tail of the outdegree distribution shows that there are a lot of participants who are not involved in transactions with many other participants in the bitcoin community.

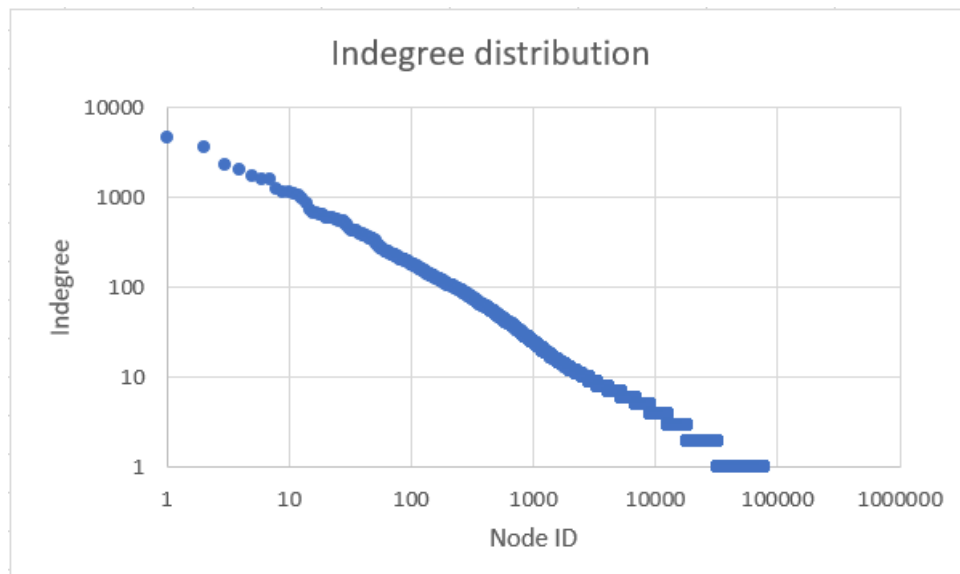


Figure 33: Indegree distribution (log scale)

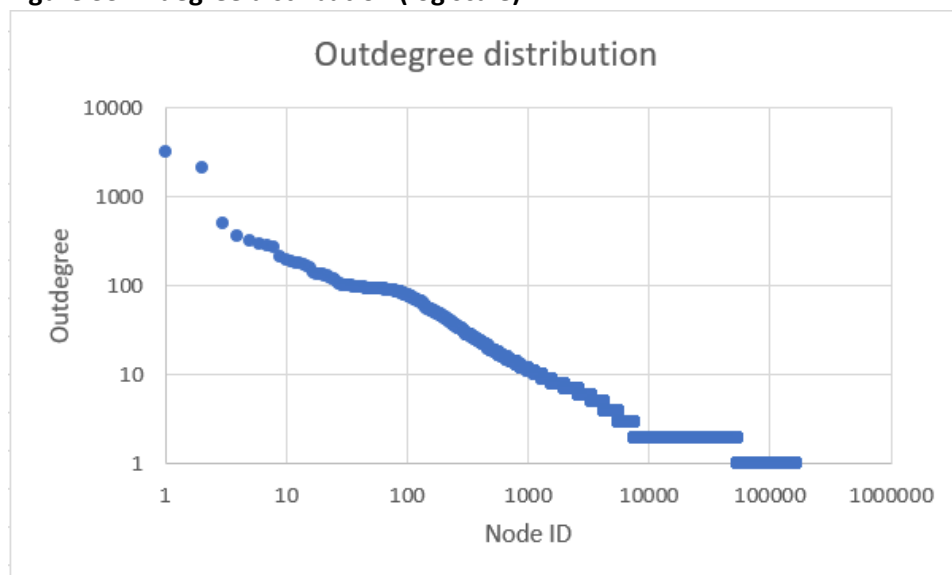


Figure 34: Outdegree distribution

- **Vertex degree analysis**

To interpret the meaning of the most important nodes in this network, table 4 provides the necessary information. The nodes in green are the exchange nodes. Their property of having both a high indegree and outdegree is explained by the role they are fulfilling in the network. The nodes in blue are the 2 miners from the analysis after 200.000 transactions. They are recognisable by their high indegree and relatively low outdegree. The higher outdegree of node 93840 can be explained by the fact that this node used different wallets to send the BTC towards the exchange node. Node 50215 took more risk by only using 1 particular path towards the BTC exchange. Finally, the nodes indicated in red are all characterised by a high indegree and low outdegree. (Node 281061 is debatable) Therefore, all these nodes are probably important miners on the bitcoin network. Furthermore, note that Slush Pool was founded on November 27, 2010, when about 200.000 transactions were completed on the bitcoin network. For this period, the algorithm computed that the highest allocated address ID was 155.703. Therefore, one would conclude that nodes with ID's bigger than roughly 160.000 are possibly part of the Slush Pool initiative. Although, miners with a lower address ID could have decided to join slush pool, which makes it very hard to make clear-cut decisions. For these reasons, it is important to stress that this report cannot pinpoint which exact addresses do belong to Slush pool in this period and which ones do not.

Node ID	indegree	outdegree
328077	3601	3248
50215	4588	21
65393	2340	2114
93840	2035	70
206088	1752	12
121903	1620	58
281061	1152	502
422845	1577	13
96169	1236	65
391201	1168	0
160125	1116	15
339609	1048	6
112908	990	6

Table 4: Nodes sorted on indegree

- **Correlation(indegree, outdegree)**

The correlation between the indegree and the outdegree is soaring here with a value of 0.551942. This concludes that participants with a higher indegree also tend to have a higher outdegree and vice versa. This is expected since these participants represent active members on the bitcoin network.

- **Average vertex degree**

The average vertex degree is equal to 1.534953. This means that an average participant on the bitcoin network doesn't perform transactions with many different other participants.

- **Degree minima and maxima**

Minimum indegree:	0
Maximum indegree:	4588
Minimum outdegree:	0
Maximum outdegree:	3248

Important to note here is that the maximum outdegree is not represented by the first bitcoin exchange, but by the node with address ID 328077. Since this node also has a very high outdegree, this node is probably the second bitcoin exchange that was started up. The node with the highest indegree is node 50215. This is the node we identified as an important miner earlier.

Distance statics

- **Average path length (directed)**

The average path length is equal to 49.317. Note that only 9 hops from the exchange node were taken, so an average path length of 49.317 is quite high. However, due to the fact that this graph is directed, probably some complex routes must be set up between some nodes to create a path. From a practical point of view, it is not interesting to analyse the meaning of this average path length in further detail.

Clustering

- **Clustering coefficient**

The average clustering coefficient is equal to 0.026. In line with the previously encountered networks, the clustering is very low for this network as well.

- **Network density**

Given the low clustering coefficient, one could expect a low network density as well. The graph density is equal to 0.00001696, which is basically 0. This is also in line with previously encountered network densities.

Centrality

- **Betweenness centrality**

Figure 35 shows the betweenness centrality of this graph on a logarithmic scale. The behaviour is in line with the previous betweenness plots in this report. Interestingly, the betweenness centrality seems to be linear until a certain value, after which it drops to 0 quite quickly. As a last mention, the 2 nodes with the highest betweenness centralities are the 2 bitcoin exchanges. They are followed by nodes with high indegrees and low outdegrees, which is a typical behaviour for a miner.

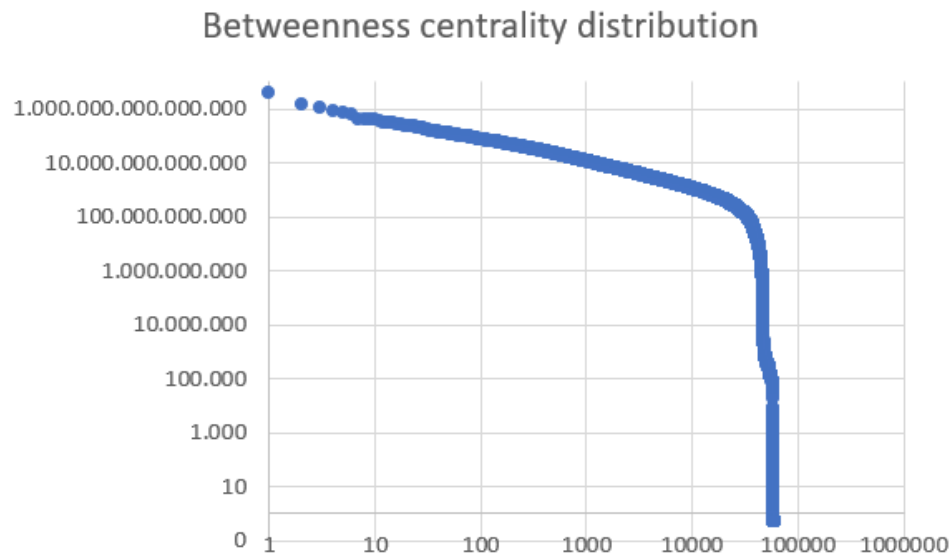


Figure 35: Betweenness centrality distribution

Page rank

- **Page rank distribution**

Figure 36 shows the linear behaviour of the page rank distribution on a logarithmic scale. Important to note here is that the miner still has the highest page rank, followed by the first bitcoin exchange. In other words, the first BTC exchange became more important in terms of betweenness centrality, but the high indegree for the miner still makes this node more important considering the page rank method. However, this could change over time as the network grows further.

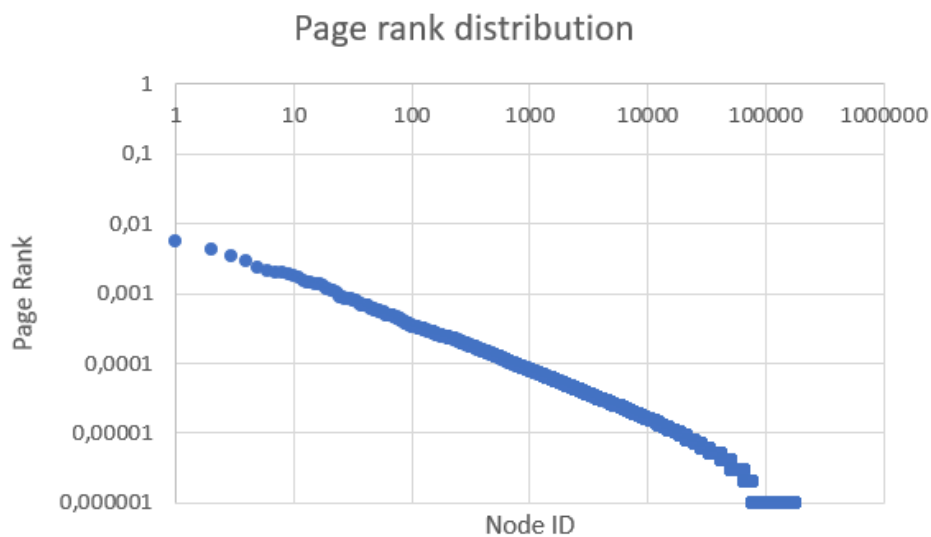


Figure 36: Page rank distribution

Conclusions

The given dataset makes the bitcoin network look like a network that was scattered in the beginning but grew towards a connected network after the introduction of the first bitcoin exchange. However, making this conclusion would not take the missing values in the input file into account. Since the bitcoin was founded in a very mysterious way, the missing values in the input file can probably be attributed to Satoshi Nakamoto, the presumed founder of the bitcoin. Since we believe that the founder has a big impact in how strongly the network is connected, this would probably result in a network that is connected to a higher degree.

The scale-free behaviour of the network is one of the main important findings of this research from a network analysis perspective because it represents the general 'feel' of the network. Considering the eccentricities of the nodes, the first bitcoin exchange became the core of the network quite quickly. A striking observation here is that it was even in the centre of the graph that was built around the miner node. In addition, the exchange also overtook the miner in terms of betweenness centrality, meaning that it appeared on the biggest number of shortest paths in the network. Looking at page rank distributions, the biggest miners seem to be more important than the exchanges for the given datasets. However, this might change when the BTC evolves to a more mature currency. In terms of clustering and density, the network provides low values throughout the entire analysis. This means that in general, participants in the blockchain network don't perform transactions with each other's 'neighbours'. Finally, considering the degree correlations participants with a higher indegree also tend to have a higher outdegree and vice versa. This means that participants who act as the recipient of many transactions also tend to send out more transactions and vice versa. Note that no conclusions about causality can be made here.

Considering the transaction behaviour, this study illustrates some critique that is often given to bitcoin. Namely, bitcoin was intended to be used for payments but in reality, it was rather an investment for the people who bought the coins. This can be derived from the low average vertex degrees and by the fact that a huge proportion of the nodes report an outdegree of 0.

Furthermore, the roots towards centralised mining were also found in this analysis. The BTC developers probably wanted to speed up the initialisation of the network by hashing a lot of the blocks themselves by means of a dedicated server. For this reason, the idea of creating mining pools as the hashes became more difficult was a step that is easy to take. These mining pools control the mining of the blocks to a high degree nowadays. The decentralisation of these mining pools is debatable but this debate falls outside the scope of this report.

At the same time, this report also critiques the Proof-of-Work mechanism which is used to decentralise the mining of the bitcoin blocks. Namely, this report already indicates that it is possible to hash way more blocks than the other participants in the network by investing in infrastructures with a high computational power. Individual miners with low computational powers make no chance in a developed Proof-of-Work system. Therefore, we conclude that the Proof-of-Work system is merely a system that measures how fast one's cumulated computational power is and how efficient one's hash solving algorithms are. Therefore, it provides a bad mechanism to pursue decentralised mining in a blockchain environment. Perhaps other mechanisms like Proof-of-Stake could bring a solution.

As a last mention, finding the Slush Pool was harder than previously anticipated. Probably, the Slush Pool founders used multiple servers to mine the blocks. This could explain the fact that multiple addresses with a high indegree and low outdegree were encountered after 500.000 transactions. However, further investigation needs to be done to make distinct conclusions about Slush Pool.

Limitations and going forward

As announced throughout the report, the mismatch between the input and the output file render the possibility to make clear-cut conclusions about the network difficult. Especially in the earliest phases of the bitcoin history, this mismatch seems to leave its mark on the analysis. However, when assessing bigger datasets, general properties such as scale-freeness can be proven without issues. This would indicate that these datasets provide a good representation of the actual network that is originated by performing bitcoin transactions. However, to solve the 'mismatch problem' in the early phase of the network, future research could imply tracking down the missing transactions of the input files. A further step could include extending the dataset with the missing values where possible. This way, more accurate conclusions about the early behaviour could potentially be made. Although, finding the remaining input address ID's is probably equally difficult as revealing the true identity of Satoshi Nakamoto since he probably gave the inputs to the first transactions himself. However, from a network analysis perspective, this should not really provide a huge problem. Since, assuming that these unknown inputs all came from Satoshi, one could extend the model by making a new address and linking it with every transaction that has no match in the input file. This way, the actual network around this Satoshi persona could be formed. Although, this approach raises 2 important questions. First of all, it assumes that all the missing input comes from Satoshi himself, which could already be a wrong theory to start with. Although, for the initialisation of the network, this theory is probably true since wiping his own address from the input files can be explained by the pseudonymous persona he created around himself. The second important consideration questions how far in time one should go with this approach. Since, attributing the first missing input data to Satoshi Nakamoto is probably the right way to go, but relating all future missing data to Satoshi might be a too far-fetched. For this reason, one should try to come up with a relevant time period to link missing input lines to Satoshi Nakamoto. Therefore, a tentative analysis of the input file and output file can be consulted in the addendum. This analysis already offers a good starting point. The conclusion of this small analysis shows that the input file especially lacks information in the very beginning of the currency's existence. These findings are in line with the theory that says that Satoshi Nakamoto did not want to disclose his own address ID. At the same time, it provides a first step to decide until when the missing input addresses could be attributed to Satoshi.

Another important limitation is the fact that for this research only 500.000 transactions were observed, thereby not covering the entire lifespan of the bitcoin network. The reason for this was two-fold. The first reason can be motivated by the computational efforts that are inherent in analysing huge datasets. In addition to this, data must be looked up in multiple files, thereby making amount of data that has to be processed even bigger. For example, analysing 1.000.000 transactions boils down to iterating over a certain amount of transactions in two files: the txin.dat and txout.dat file. As shown in the analysis of the files in the addendum, from a certain point in time the amount of lines that must be read in starts to grow faster than the number of actual transactions. This is due to the fact that multiple addresses start to participate in the same transaction. For even bigger datasets, some of the transactions should even be looked up in a third file: the multiple.dat file. For these reasons, computational efforts generally behave exponentially. The second reason is identified by the actual scope of this research itself. The research attempted to shed light upon the very beginning of the bitcoin network. Therefore, it was not needed to observe more transactions.

Going forward, one could extend the model by adding weights to the graph. These weights could help identifying the degree to which who trades bitcoins with whom. Thereby the transaction behaviour of the different parties in the network could be analysed in further detail. Another extension would be adding more shortcuts to the algorithm to make the computational time more robust to handling bigger datasets. Thereby, one could analyse bigger datasets within a reasonable computational time. Currently, the model needs 60 hours to process 1.000.000 transactions.

Furthermore, Slush Pool could be investigated in further detail. This report offers the foundation towards looking for the first mining pool. However, more historical and network analysis needs to be done to make distinctive decisions.

REFERENCES

- Alvarez, J. (2019, March 19). Who is Satoshi Nakamoto? We Look at The Possible Candidates. Retrieved from <https://blockonomi.com/who-is-satoshi-nakamoto/>.
- Bashir, I. (2017, March 17). Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks. Packt.
- Block Explorer. (2019). Blockchain Explorer. Retrieved from <https://www.blockchain.com/explorer>.
- Coincodex. (2018). Lawsuit for Over \$10 Billion in Bitcoin Allegedly Stolen by “Fake Satoshi”. Retrieved from <https://coincodex.com/article/1357/lawsuit-for-over-10-billion-in-bitcoin-allegedly-stolen-by-fake-satoshi/>.
- Financial Times. (2018, November 22). Bitcoin’s crash is not the end of cyber currencies. *Financial Times*. Retrieved from <https://www.ft.com/content/ea80a128-ee6a-11e8-8180-9cf212677a57>.
- Kelly, J. (2017, November 29). Factbox - Things you might not know about bubbly bitcoin. *Reuters*. Retrieved from <https://www.reuters.com/article/uk-markets-bitcoin-factbox/things-you-might-not-know-about-bubbly-bitcoin-idUSKBN1DT30L>.
- Kharif, O. (2019, April 24). John McAfee Vows to Unmask Crypto’s Satoshi Nakamoto, Then Backs Off. *Bloomberg*. Retrieved from <https://www.bloomberg.com/news/articles/2019-04-23/john-mcafee-vows-to-unmask-crypto-s-satoshi-nakamoto-within-days>.
- MIT. Bitcoin network dataset. Retrieved from <https://senseable2015-6.mit.edu/bitcoin/>.
- Palatinus, M. Industry Pioneers. Retrieved from <https://slushpool.com/about/facts/pioneers>.
- Rathod, A. (2018, November). Who is Satoshi? Bitcoin’s Early Developer Jeff Garzik, Has Some Theories. Retrieved from <https://toshitimes.com/who-is-satoshi/>.
- Song, R. (2018, April 2). Why Bitcoin is Different. *Medium*. Retrieved from <https://medium.com/@jimmysong/why-bitcoin-is-different-e17b813fd947>.
- Song, R. (2018, April 10). Mining Centralization Scenarios. *Medium*. Retrieved from <https://medium.com/@jimmysong/mining-centralization-scenarios-b74102adbd36>.
- Torpey, K. (2019, April 29) Here’s What The Bitcoin Community Should Do About The Fake Satoshi Nakamoto. *Forbes*. Retrieved from <https://www.forbes.com/sites/ktorpey/2019/04/29/opinion-heres-what-the-bitcoin-community-should-do-about-the-fake-satoshi-nakamoto/#7dd9d9c21a8f>.
- Toshendra, K. S. (2018, January 25). What are alternative strategies for Proof-of-work? *Blockchain council*. Retrieved from <https://www.blockchain-council.org/blockchain/what-are-the-alternative-strategies-for-proof-of-work/>.
- Wikipedia. (2019, April 2). History of bitcoin. Retrieved from https://en.wikipedia.org/wiki/History_of_bitcoin.
- Wikipedia. (2019, May 2). Satoshi Nakamoto. Retrieved from https://en.wikipedia.org/wiki/Satoshi_Nakamoto.
- Zainuddin, A. Blockchain Scalability Solutions: Overview of Crypto Scaling Solutions. Retrieved from https://masterthecrypto.com/blockchain-scalability-solutions-crypto-scaling-solutions/?fbclid=IwAR39bcQZnBtQ6PVE8F-wdXPIoMAdY7oZurfJYzM8fY180Mkm_WapFaVs3uQ.

ADDENDUM

Input file and output file analysis

Analysing the amount of lines that must be read in to obtain a certain transaction ID, something interesting can be reported about the input and the output file. Figure 37 shows that the input file misses a lot of input data for the first 50.000 transactions on the network. This can be derived from the structurally lagging behaviour of the amount of lines that must be read in to obtain a certain transaction ID. Since many transaction inputs are omitted, the amount of lines that must be read in is significantly lower than the cumulated transactions. Later, the input file seems to follow the output file more nicely. On the other hand, Figure 38 shows that the output file behaves normally. This is the output file. This phenomenon is once again supported by the fact that Satoshi probably wanted to keep his identity as a secret by hiding his address in the input file.

However, when looking at both files for a longer time horizon, the behaviour changes a little bit. Namely, at a certain point in time, it was possible to perform transactions between different parties within the same transaction ID. For example, 5 different parties could set up a transaction in which 3 of them received bitcoins while the 2 others sent the bitcoins. For this reason, the bitcoin developers even introduced the multiple.dat file which was dedicated to store the data of such transactions. Therefore, figure 39 and figure 40 show a drastic increase in the amount of lines that must be read to analyse a given amount of transactions. This is because every address that is linked to a certain transaction corresponds with a new line in the input and output files.

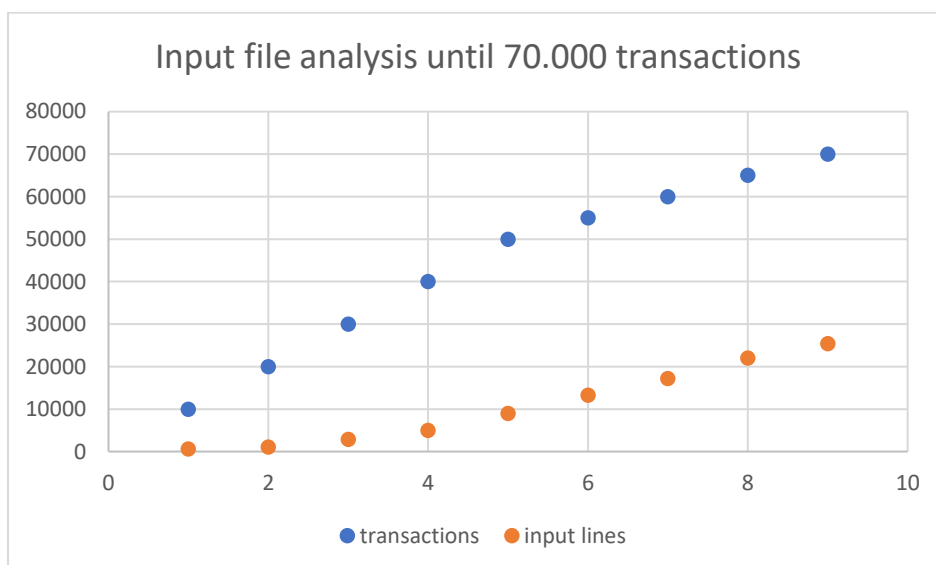


Figure 37: Input file analysis until 70.000 transactions

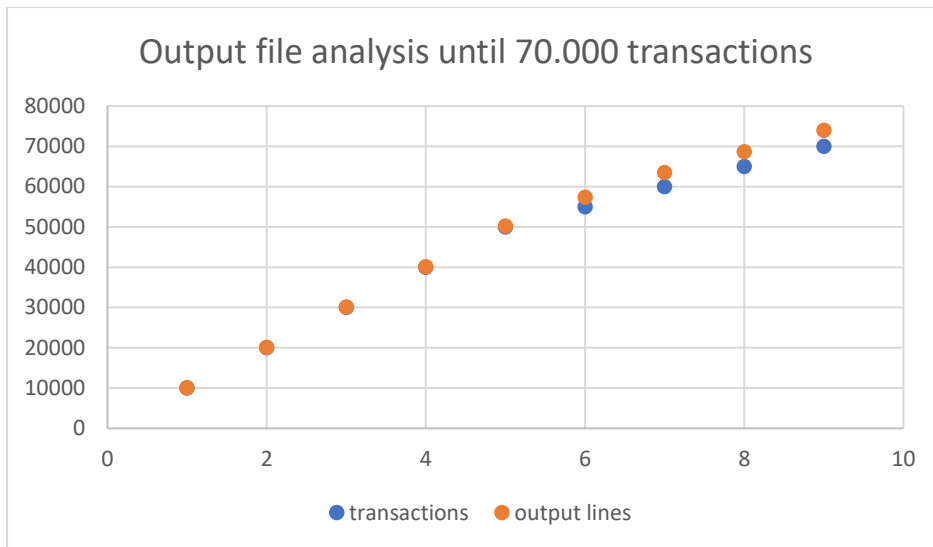


Figure 38: Output file analysis until 70.000 transactions

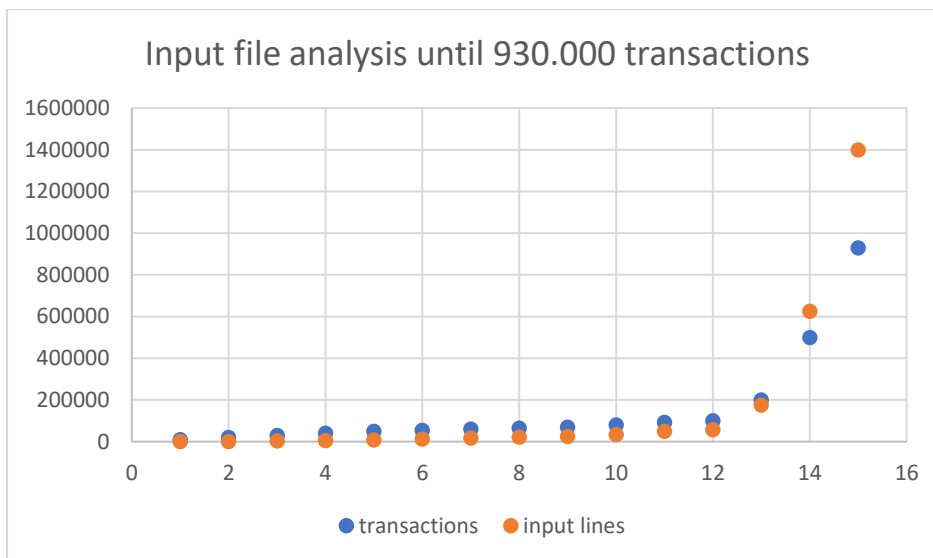


Figure 39: Input file analysis until 930.000 transactions

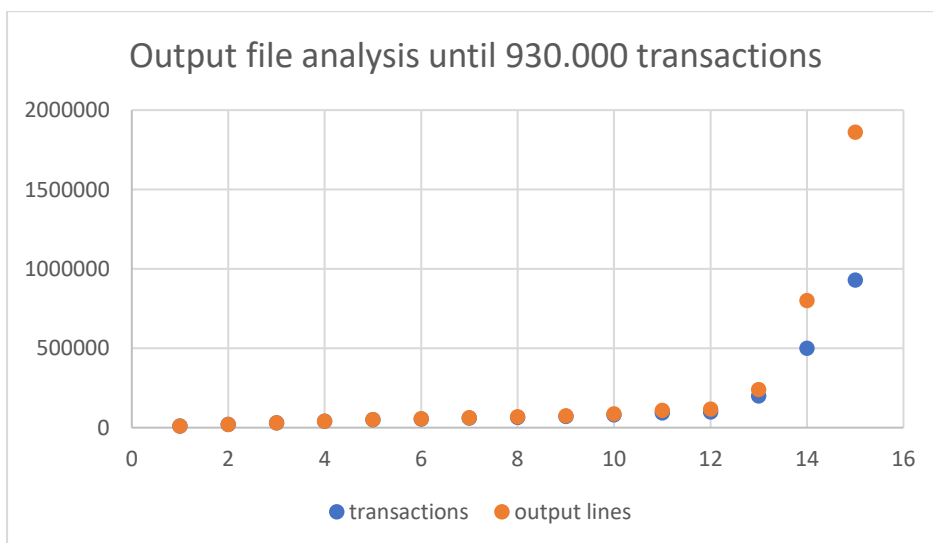


Figure 40: Output file analysis until 930.000 transactions