The Reaction of a Network: Exploring the Relationship between the Bitcoin Network Structure and the Bitcoin Price

Mariano Sorgente
Cristian Cibils

December 10, 2014

Abstract

he Bitcoin protocol and the bitcoin currency have been talked about a lot recently, in the advent of cryptocurrencies and decentralization protocols. We were especially captivated by the explosion of the bitcoin bubble in January 2014, as well as the bubble of 2013. Motivated by these price fluctuations, we decided to analyze the Bitcoin network as the transaction network it is, with the hopes of finding some correlation between the network structure and the bitcoin price. The blockchain, the decentralized ledger, holds a historical record of every transaction in the network. They key intuition in our project is the translation of every transaction into an edge and every address into a node, in order to represent the transaction network fully. Although much prior work has been done in the field, the relationship between the bitcoin network structure and the bitcoin price remains largely unexplored territory. Our analysis leads us to some dead ends, as well as to some very interesting finding about the transaction network.

1 Introduction

What correlation exists between the Bitcoin network structure, and the bitcoin price? To be clear, we will be using the term *Bitcoin*, with a capital B, in reference to the Bitcoin protocol, the underlying principles in the mining network, and the term *bitcoin*, with a lowercase b, in reference to the currency unit, and thus the transaction network that underlies it. We make this distinction because of the dual network nature of Bitcoin. The protocol, which specifies the ledger maintenance and new coin generation, is in itself a network of miners that ensure the smooth flow of financial transactions. The set of all transactions and addresses in the blockchain, however, also can be represented by a graph. . Throughout this project, we will be focusing on the latter.

We were struck by a graph of the bitcoin transaction volume plotted against the bitcoin price (See Figure 1). The graph clearly shows a positive correlation between price and transaction volume, until January 2014. Even more interestingly, there seems to be a negative correlation between volume and price from that point forward.

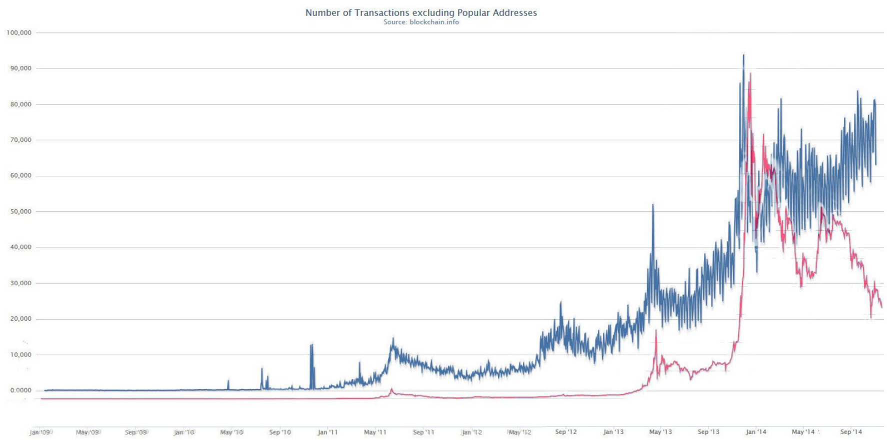


Figure 1. Transaction volume (blue), vs price in USD (red). The y axis represents number of transactions, and the price was scaled to fit such that it's highest point is USD 1,100.

We believe that there are inherent patterns that occur as a reflection of the bitcoin price, with hoarding of wealth being the most prominent. We aim to identify such patterns by extracting information from the bitcoin network on a time lapse, and plotting network structure properties and the price in USD, against time. Some of the features we wish to extract because we believe that they are indicative of users' behaviors are:

1. Average Clustering coefficient: As defined by $\frac{\sum_{i=1}^{n} C(i)}{n}$. Average Clusterintg coefficient is a good indicator for overall network connectedness. Our intuition is that if the network as a whole is highly connected, then the price is likely to increase because the transactions are distributed through the whole network and thus it becomes attractive to new users (process which in turn, increases the price).

2. Alpha-value in power-law distribution of transactions per node: As extracted from a Maximum Likelihood Estimate. Our intuition is that the less concentration of power, or in other words, the heavier the tail, represets less hoarding of coins. This, in turn, encourages investment, as the network sees more liquidity.

3. Number of edges added to network: It makes sense that large price swings may increase trading activity, and this leads to more transactions. It also make sense that number of adresses and transaction size would increase with price changes here. We also explore whether an increase in edges (and nodes) leads to any noticeable change in price.

4. Degree distribution: In the Bitcoin economy, it seems that users try to minimize address reuse. Therefore most nodes only have one or two transactions. There are some exceptions, such as addresses posted on public websites (bitcointalk, youtube, facebook), and specific situations such as paying miner fees. More importantly, there still exist some mobile and

web wallets which don't automatically reuse addrsses. This means that an increase in degree distribution might be a good indicator of adoption, as people use their wallet to buy and sell products and services. Strangely enough, the degree distribution seems to stay stable until around Jan 2014, (the bubble crash), after which it seems to consistenly go up. (Interestingly, opposite of the price. One of our theories is that Jan 2014 is around the time were some regular users adopted Bitcoin. Most if not all of our adoption metrics seem to go up from this time, except for price.)

5. Average balance: The average balance of an address depends heavily on how much money a user is willing to store, as well as the price. Our hypothesis was that when the price goes up, the balance should go down. This is because people will not need to send as much money. (i.e 1 bitcoin = $500 vs 1 bitcoin = $1000, if I want to send $500, I would need to send 1 bitcoin vs 0.5 bitcoins.).

Moreover, other features like average degree may not be so useful, given that it is custom for new addresses to be generated for new transactions. Later in the paper, we graph all of these features and attempt to understand how they have changed, and how they relate to the price.

2 Review of Prior Work

2.1 A Quantitative Analysis of the Full Bitcoin Transaction Graph (Ron et al. 2012)

In [1], Ron et al. present statistical findings from crunching the Bitcoin network from its inception until 2012. Some of the descriptive findings of the paper are the Union Join algorithm, which allowed them to detect entities that use multiple addresses, the distribution of BTC balance amongst addresses and entities, and the number of inactive bitcoins due to abandonment or personal reserves. It was highly insightful to notice several key properties of the Bitcoin network so clearly present in this paper. It was also one of the very few papers that made an allusion to the price of the currency ($12 at the time), and it was noticing this, coupled with the curiosity of what the network looks like two years later, that peaked our curiosity for this project. The main question extracted from this paper was how much of the exceedingly high portion of inactive bitcoins stays inactive to this day We found ouresults for this to be very similar to those found in 2012.

2.2 BitIodine: Extracting Intelligence from the Bitcoin Network (Spagnuolo et al., 2014)

One such paper by researchers at Politecnico Di Milano [2] attempts to create a graph model for the Bitcoin transaction network. In it, they describe the creation of a system called BitIodine which parses the blockchain, clusters addresses, and classifies and labels them. This is very useful because it allows labelling of certain addresses by forum usernames. This allows use of the system to analyze user behavior. Furthermore, BitIodine enables tracking the path between two addresses or nodes. This allows us to calculate average distances in the network, and degree distributions. The source is publicly available, and thus can be used to create a network model and extract useful network metrics over time. However the systems comes with several downsides. It takes the state of the blockchain at a current point in time, and thus is not optimized to work with the entire blockchain history. If we wanted to calculate the degree distribution at a current time, we would have to parse the entire blockchain once again. It also does not provide other network analysis tools such as measuring clustering of nodes. Many concepts and ideas will be useful from BitIodine. In our project we hope to explore similar information regarding the network. However we plan to create a system which is more optimized for querying about this data at a current point in time, and using all of this to find some connection to the price of Bitcoin.

3 Data Collection

3.1 Data Set

Fortunately, the Bitcoin graph is permanently stored with extreme redundancy in the Blockchain, a public ledger of all transactions since the beginning of the protocol. We used open source tools to parse the blockchain and extract time-labeled transaction information. Particularly the bitcoinj api was very useful here. we had to consider many details about transactions, blocks, coinbase transactions, miner fees, inputs, outputs, and hash values in order to parse the blockchain correctly. There were many edge cases and strangely formatted transaction scripts. We wrote a Java program that iterated through all 300,000+ blocks, and all transaction inputs/ouputs in each block, and logged them to a file, to be used by SNAP. This edge file was around 29GB. It contains every single transaction since the Genesis block, and the amount of satoshis (1/100,000,000 BTC) transacted. We also retrieved price data since 2011 from Coinbase's api. Our time data starts at around block 60,000.

3.2 Technical Challenge

The main technical challenge we face is certainly dealing with the size of the blockchain. Parsing everything into the database took significantly more time than anticipated (around 4 hrs), and the computations we made seem to grow in complexity with the size of the network.

Our main python executable iterates through the edge file, creating nodes and adding edges. It also stores a dictionary mapping from bitcoin addresses to $node_ids$. The first important roadblock we encountered writing this was that the program would far exceed or 16GB memory limit. To overcome this, we saved the $adress- > node_id$ dictionary on disk, and compressed the graph by combining duplicate transactions. In the end the program was running with over 10GB of ram, but still underneath our limit.

Another tough challenge was running time of the program. Initially, the running time seemed to scale exponentially with block number. By block 150,000, it had completely stopped. We performed many small optimizations, as well as some larger ones. We only computed expensive operations (such as caculating alpha coefficients), once every few hundred blocks. This would give us enough precision. We also stopped calculating the size of the largest WCC, as this seemed to stay almost equal to the size of the network over time. Finally, we avoided using SNAP functions where possible, as many were algorithmically too slow in complexity for our use case.

4 Method

Our main algorithm for calculating all these metrics can be described as follows:

1. Initialize an empty network N
2. For each block in the Blockchain:
   (a) Parse the blockchain information for that block
      i. Translate every transaction to an edge and every address to a node
   (b) Load the newly parsed information into N
   (c) Compute properties
      i. Number of nodes (addresses)
      ii. Number of edges (transactions)
      iii. Number of edges in last block
      iv. Number of nodes in last block
      v. Average transaction value
      vi. Maximum transaction value (avg over last x blocks)
      vii. Minimum transaction value (avg over last x blocks(
      viii. Degree distribution, and average degree (with SNAP)
      ix. Size of Largest Connected Component (with SNAP)
      x. Alpha value Derived from MLE (with SNAP)
         A. Fit Line with # transactions as X and Probability of Having that # transactions as Y
   (d) Get Price at Timepoint of that block
   (e) Output all these properties and price / time to file
3. Plot the different properties vs time, along with price vs time, and scale them for a more descriptive graph.

5 Results

Our computation yielded remarkable results. We found that there were several key network properties that were very strongly correlated (either positively or negatively) with the price of the currency. Moreover, some graphs will show that some of the extracted features serve as a prediction for the price. We display some of our findings in the figures below:

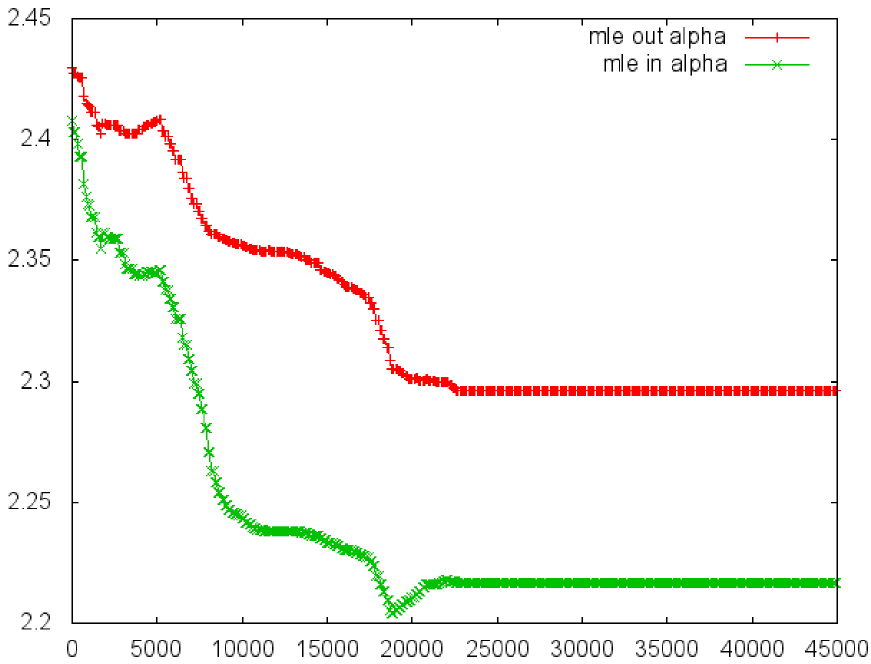
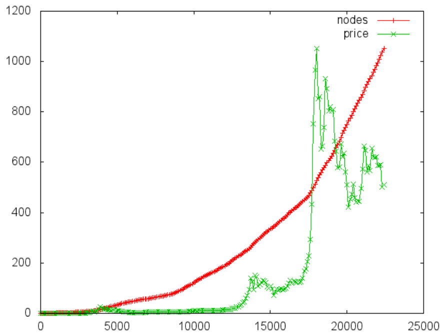
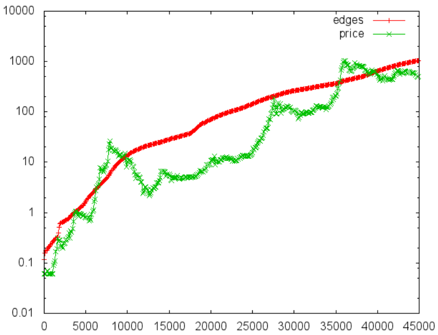


Figure 2. In Degree Alpha and Out Degree Alpha as derived from MLE.

We were very pleasantly suprised to see the alphas converging to approximately two. Our method was that for every 1000 blocks, we would compute the MLE estimate of the power law exponent, given by the ratio of our x-axis being node degree, and y-axis being probability of a node having that degree. We also realized that convergence happened very early in the days of the network, leading us to believe that by the time that the network had a price, the alpha values remain set at 2.216. Unfortunately, this means that there was no observable correlation between the value of alpha and the price of the currency.

The findings start taking interesting shape, however, when we start analyzing the correlation between our extracted data, and the bitcoin price. For reference, we scaled the network property graphs to match the max_y of the bitcoin price, thus the y-axis is always expressed in USD. Moreover, the x-axis is given by block number in the blockchain.





Figures 3 and 4. Nodes in the network vs price, and Edges in the network vs price (logarithmic y-axis).

Two of our most interesting findings were the correlations between nodes in the network, edges in the network, and price. Our data revealed that there was a correlation coefficient r of 0.8467 between the number of nodes and the price, and that there was an r of 0.8322 between the number of edges and the price. Preliminarily, this data indicates that as the network grows, the currency becomes more valuable. (Or the other way around: as the price increases, the media starts talking about Bitcoin, leading to many new adopers and thus the size of the network grows)

Our next exciting finding was the came in the hands of block size (number of transactions per block), which reported an r value of 0.7941

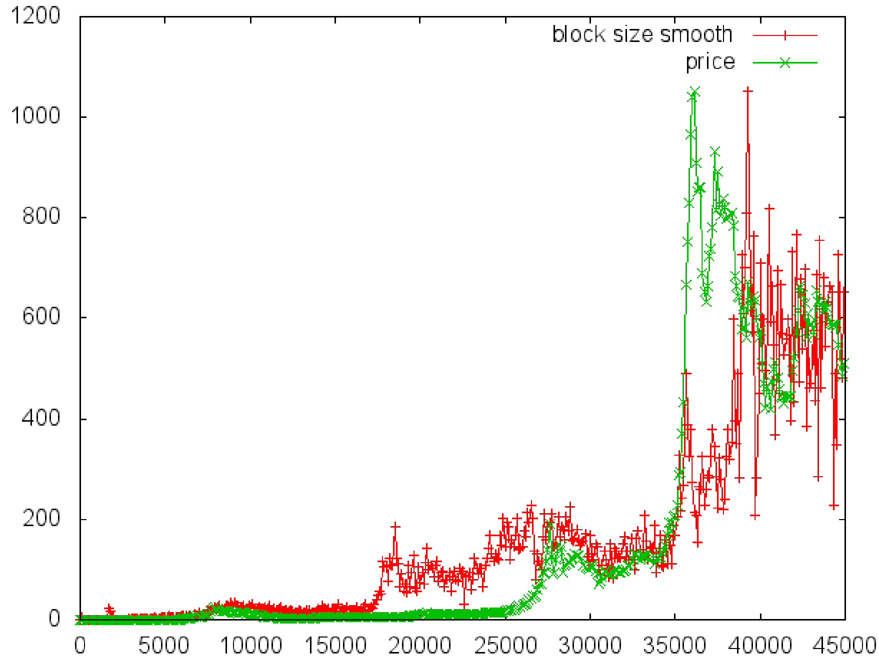


Figure 5. Block size (smoothed over 10 blocks) vs price.

Intuitively, we can reason on this finding as being the relationship between "hype" and price. The more transactions in a block, the busier the market and thus it maes sense that the price would go up.

Another very interesting finding was average node degree, much to our surprise given our justification in the introduction. We found that average node degree indeed offers a strong positive correlation with price, as seen in figure 6.

As can be seen by Figure 6, the average degree stabilizes for a few months. However it starts increasing rapidly after the bubble of 2014. Our theory for this phenomenon is that as more novice users get their first bitcoins, they don't know about the security issues regarding address reuse, and send many transactions to/from one address. Therefore user adoption increases lead to a higher average degree. This is consistent with the finding that user adoption has seen an upward trend througout 2014.

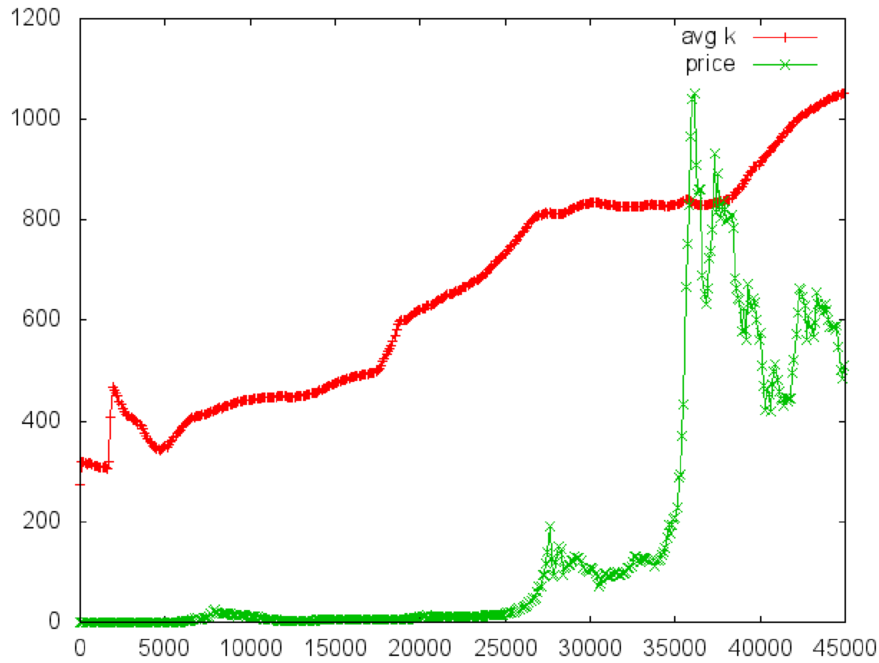


Figure 6. Average node degree vs price.

Note how at the end, we see a strong decrease in price, and a strong increase in node degree.Overall, these two data sets presented a correlation of 0.69
The final interesting discovery was the negative correlation of -0.6873 between average balance and price, as seen in figure 7:

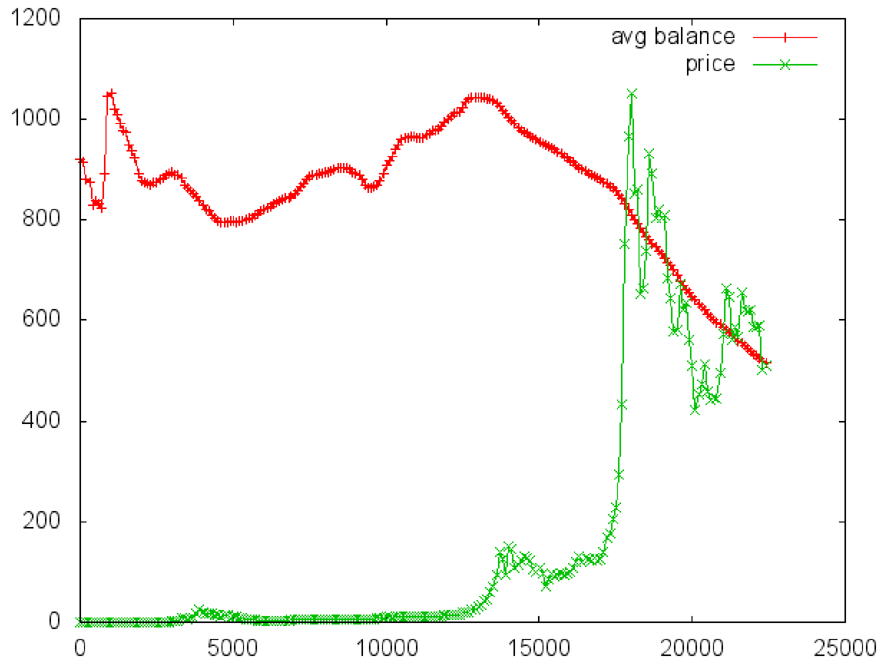


Figure 7. Average node balance vs price.

Our intuition behind Figure 7 is that as people start selling their bitcoins (and making avg k decrease), the value of the currency goes up, but only until a threshold is reached.

6 Discussion

After a lot of hard work put into optimizing for both hardware and software limitations, we were very pleased with our findings. We can see from the graphs that there are network properties that strongly indicate the behavior of the price. However, the most interesting part for us was analyzing what happened during January 2014 (for reference, Jan 14 is the highest peak in the price graph). Although there is much to be said about price speculation being artificially inflated, we see something more. If we look at Figures 6 and 7, we see that near the end, when the price starts plummeting, average node degree increases. We posit that this represents a break from exclusively speculative trading, and a forage into usage-based valuation, interestingly enough, one would be inclined to say that it was this artificial inflation that allowed bitcoin to be traded more (hence the increase in degree), and thus reach usage-based valuation. We are also very excited to see more general results, such as nodes-in-network and edges-in-network as positive indicators for price. We are happy to report that block size is also a part of this group.

Overall, our findings represent a personal breakthrough in handling large scale data. Overcoming hardware (lack of enough memory and processing power) and software (we took SNAP to its limits) limitations was half of the assignment, interpreting the data was the other. We are extremely happy to present our results, and we look forward to other people using our data for research.

7 Final Network Statistics

Here are some of the statistics that we gathered from out dataset, at the state of the Blockchain in October 2014:

1. Blocks: 317,000
2. Transactions: 51,000,000
3. Addresses: 41,000,000
4. Average balance: 2.48 BTC
5. Average node degree: 2.97
6. MLE alpha in: 2.216
7. MLE alpha out: 2.296

References

1 Dorit Ron and Adi Shamir (2012). Quantitative Analysis of the Full Bitcoin Transaction Graph. *Cryptology ePrint Archive, Issue 584*

2 Michele Spagnuolo, Federico Maggi and Stefano Zanero (2014). BitIodine: Extracting Intelligence from the Bitcoin Network. *Financial Cryptography and Data Security 2014*