Group 3: Analysis of Bitcoin Blockchain First Draft

Group Members

- Hangyu Kang (hkang98)
- Shengwen Yang (syang382)
- Bodi Yang (byang89)
- Fengxia Dong (fdong6)
- Sam Tauke (sstauke)

Introduction

For the past decade, Bitcoin (BTC) has been the pre-eminent cryptocurrency. Unlike the public banking system we are used to, the recorded transactions in blocks are open to the public and everyone shares the same transaction records once the transaction has been done between the users. It uses block-chain technology for security. Each block is a record of the transactions that have happened since the last block. And those blocks are connected in a chain, each block pointing to the blocks before and after it. Thus, we call the cryptographic technique the block-chain. A block is finished once a bitcoin miner has solved the mathematical equation linked to that block. The miner who solves the hash problem associated with a block is then allowed to push the verified list of transactions from the block to the blockchain and is rewarded with BTC as payment. Because the size of the block is dependent on how fast miners can solve the hash problem, the amount of time between new blocks varies. The BTC protocol aims for each block to take about 10 minutes and every 2016 blocks the protocol adjusts the problem difficulty to get closer to this 10-minute average.

Additionally, blocks are limited to at most 1 mb of transaction data. Therefore, it is possible that the amount of transactions since the last block can exceed the amount that can fit into the current working block (in cases of large numbers of transactions or a long time to solve block). To solve this, individuals will pay fees to miners to ensure that their transactions are included in the next block, so that they do not have to wait for a long time before they can consider their transaction as valid.

Goal of the Project

Given that it is an entirely new form of currency, there is much that is unknown about BTC. In this project, we plan to answer the following questions:

1. What is the average number of transactions/transaction quantity/time elapsed per block? As we discussed in the Introduction, the size of the block is dependent on how fast miners can solve the hash problem. While it generally takes miners about 10 minutes to solve the hash problem, what can the average number of transactions, transaction quantity, and time elapsed per block be?
2. How has the number of BTC transaction / total BTC transacted changed over time? Has it been strictly increasing or have there been dips?

Bitcoin was not popular when it was launched in 2009. However, it has recently become the world's largest cryptocurrency by market cap. We are interested in examining the pattern of the evolving path of the number of transaction and the amount of Bitcoin transacted.

3. Is there a relationship between price and transaction volume? Does the speed of mining increase when prices increase?
   A defining characteristic of BTC has been its high price volatility, especially in its early period, as illustrated in **Figure A** below. It is of interest to examine if people are more likely to transact BTC and if the miners are motived to solve hash problems quicker when prices are higher.

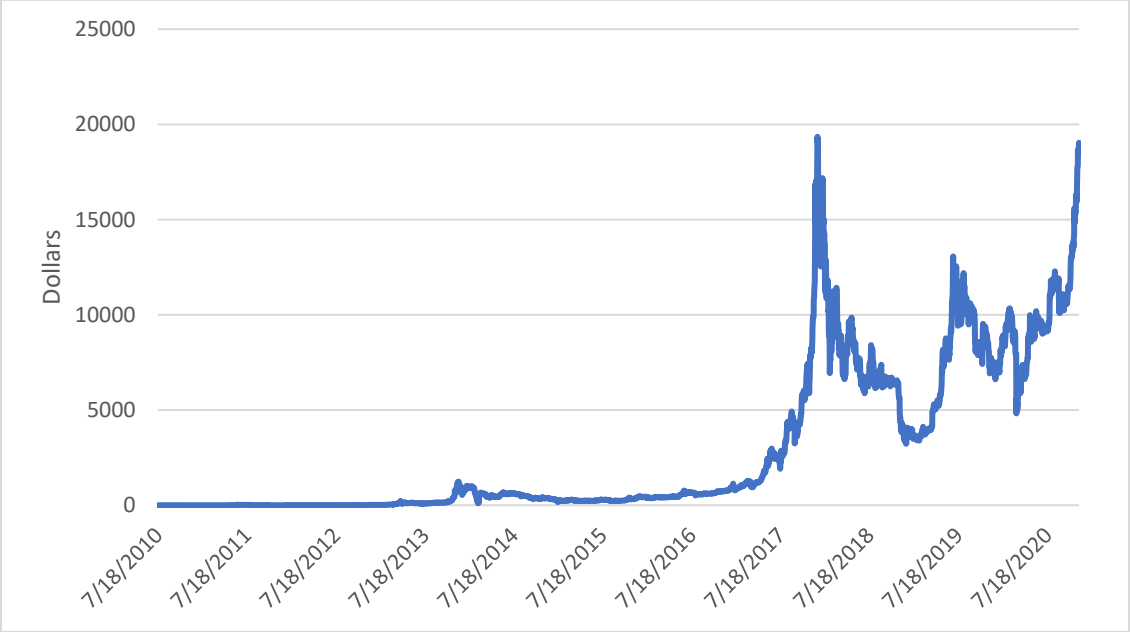

Figure A: Price of BTC from7/18/10–11/24/20

Source: investing.com

Data Background and Preprocessing

Our data come from a user on kaggle who scraped the BTC ledger to pull all blocks that have inputs from the genesis block in January 2009 through block 507,999 in February 2018. The data are available here.

A block is considered to have no input if no transactions occurred between the mining of the previous block and the mining of that block. We expect (and observe) that this would occur more frequently in earlier period when there were fewer people using BTC. Indeed, the data show periods of several hundred blocks without any transactions. The data are organized in a series of csv files that each has five columns: block height, input hash, output hash, sum of BTC transacted, and timestamp. The data are stored in 73 CSV files that are 72.41 GB in total. In addition, we collected minute prices of bitcoin from investing.com for the period of 07/18/10 to

11/24/20 and put them on kaggle. Both sets of data are then downloaded from the kaggle API using bash scripts for our analysis.

We preprocessed the raw data with Linux Shell. We cleaned the raw data by removing the punctuation characters (quotation, parenthesize, etc.) to make it neat, and sorting the data (based on time).

Analysis Results

We structured our analysis through a .dag file that splits the 22 csv files into 22 different condor jobs and compiles the results. We conducted analysis with R.

To answer our first question, we calculated average transactions and the amount of BTC transacted over the whole period, ranging from January 2009 to February 2018. We found an average of 442.32 transactions and 10,541.23 BTC transacted per block. Additionally, we found that the average elapsed time per block was 28.51 minutes.

To answer our second question, we calculated the average transactions and number of bitcoin transacted per block over several time periods. Blocks per day is calculated as total number of blocks with at least one transaction divided by the total number of days with at least one transaction. As shown in **Table 1**, transactions and number of BTC transacted per block had changed over time. We see a steady increase in transactions per block from 212.68 over blocks 190,001 through 209,999 to 846.48 over blocks 367,001 through 369,999, reaching the highest over blocks 364001 through 367000. A similar pattern emerges for the number of BTC transacted per block and the Blocks per Day. Further, we see a convergence towards the intended 10 minutes per block. All of these changes are consistent with the increased interest in BTC.

Block Range		Date Range		Transactions Per Block	Number BTC Transacted per Block	Blocks Per Day	Minutes Per Block
190001	209999	7/19/2012	3/12/2014	212.68	30,499.25	32.73	44.00
210000	224000	9/6/2012	3/13/2014	315.24	10,845.86	25.08	57.41
224001	234000	8/24/2012	5/1/2013	329.25	10,460.62	39.71	36.26
247001	260000	7/15/2013	3/16/2014	278.73	4,415.73	53.02	27.16
270001	280000	11/14/2013	3/16/2014	368.72	7,417.12	81.39	17.69
283895	290000	2/3/2014	3/16/2014	403.84	4,686.17	147.68	9.75
290001	300000	3/18/2013	5/10/2014	388.48	4,085.02	23.70	60.75
310001	320000	7/8/2014	9/10/2014	418.81	3,756.97	154.77	9.30
330000	337000	11/12/2014	1/1/2015	591.70	6,636.03	138.36	10.41
337001	343000	12/28/2014	2/11/2015	639.56	8,115.18	131.33	10.96
343001	349000	1/30/2015	3/24/2015	678.43	6,570.97	111.66	12.90
349001	354000	3/22/2015	4/27/2015	732.36	6,511.63	136.67	10.54
354001	360000	4/22/2015	6/8/2015	751.52	7,720.16	124.30	11.59
360001	364000	6/4/2015	7/5/2015	848.53	9,202.25	124.71	11.55
364001	367000	7/4/2015	7/26/2015	1,034.93	8,547.54	129.59	11.11
367001	369999	7/24/2015	8/15/2015	846.48	8,072.96	130.36	11.05
Full Period				**442.32**	**10,541.23**	**50.52**	**28.51**

Table 1: BTC Transactions 7/19/12–8/15/15

There are certain data anomalies that need to be explored further. For instance, we see a very high average number of BTC transacted per block in the first range of blocks. Is this reflective

of the true trading patterns during this period or were there outliers that skewed the numbers? Additionally, we see block ranges with nonsensical entries like dates in the future. This could be due to an error in our process or potentially indicative of an error by the individual who scraped the original ledger data. Due to data anomalies, certain block ranges are excluded.

Conclusion

While far from groundbreaking, the results of this initial inquiry confirm our expectations. We observe that the volume (in terms of both number of transactions and amount transacted) has increased steadily over time. We also find that the minutes per block has tended towards 10 minutes per block.

In the future, we intend to investigate both the data anomalies that we observe here and the relationship between price, transaction volumes, and mining rates.