MA 703 Final Report

Yukun He

Bitcoin Online Rating Dynamic Network Analysis

Abstract

In this project, I will perform descriptive analysis for a network of users who trade using Bitcoin on a platform called Bitcoin Alpha. Through the analysis, I want to explore the consequential overtime effects on the interaction among users and the behavior of specific groups of individuals of the trading network. The phenomenon was caused by two closely related milestone events in the middle of an eight-week period from the Bitcoin development history. Discoveries from the analysis can help us understand more about this popular type of cryptocurrency.

Introduction

A cryptocurrency (or crypto currency) is a digital asset designed to work as a medium of exchange that uses cryptography to secure its transactions, to control the creation of additional units, and to verify the transfer of assets. (Greenberg, 2011) Cryptocurrencies are a type of digital currencies, alternative currencies and virtual currencies. Cryptocurrencies use decentralized control (McDonnell, 2015) as opposed to centralized electronic money and central banking systems. (Allison, 2015) The decentralized control of each cryptocurrency works through a blockchain, which is a public transaction database, functioning as a distributed ledger. (D'Agnolo, 2015) A blockchain is a continuously growing list of records, called blocks, which are linked and secured using cryptography. (Narayanan, Bonneau, Felten, Miller, Goldfeder, 2016) Each block typically contains a hash pointer as a link to a previous block, a timestamp and transaction data. (Investopedia, 2016) By design, blockchains are inherently resistant to modification of the data. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way". (Iansiti, Lakhani, 2017)

Among all the cryptocurrencies, Bitcoin is the most well-known and successful one. The concept of Bitcoin was first introduced by Satoshi Nakamoto in the year of 2009, immediately after the global economics crisis, based on his design and publication of an open-source software and a peer-to-peer network constructed on that software. It is the first decentralized digital currency, as the system works without a central bank or single administrator. (Brito, Castillo, 2013)

In the most recent decade, the Bitcoin has caused a fever which swept through the global market. It's high security, high convenience, and less artificial controlled market prices attract either corporate or individual investors. Also, Bitcoin provides an alternative way of trading for a vast amount of people in the world who do not have access to bank accounts or are facing fiat cash hyperinflation. From February 9th in 2011 to December 17th 2017, the price per Bitcoin has increased from 1 dollar to 19,783 dollars. Consequentially high amount of user activity creates ideally large amount of network activities, which can be interesting to analyze.

For this project, my focus will be on the consequential influence from two historical milestone events that closely took place in the history of Bitcoin development. On November 29th, 2013, the price per Bitcoin pasted the 1,000 dollars mark. The value per Bitcoin almost doubled in one week period. This rapidly increasing in buying was mainly caused by Chinese individual investors, who wanted to transfer their money into untraceable oversea accounts. Bitcoin value reaches a peak at 1,216.70 dollars. The price will not break 1,000 dollars for the next three years. (Crane, 2017) Several days later, in the afternoon of December 5th, 2013, People's Bank of China released a policy about deriding the usage of Bitcoin as a type of currency. The policy mentioned that Bitcoin did not have true meaning and cannot be treated the same as fiat cash. Consequently, People's Bank of China forbade Chinese finance institutions and payment institutions from trading Bitcoins or pricing with Bitcoins. In one hour after the releasement of this policy, the price of Bitcoin decreases by 35%. This day is called the darkest day for Chinese Bitcoin developers.

Data and Data Manipulation

My network dataset is acquired from the Stanford Network Analysis Project. It is a who-trusts-whom network of users who trade Bitcoins via a web-based platform called Bitcoin

Alpha. Users on this platform are anonymous, such that a reputation record for each user will keep the trading environment healthier and more trustworthy. Users' reputation is collected and calculated to prevent transactions with fraudulent and risky users. Members of Bitcoin Alpha rate other members in a scale of -10 (Highly Untrustworthy) to +10 (Highly Trustworthy) in steps of 1. This dataset represents a weighted signed directed network, with 3,783 nodes and 24,186 edges. An overview of the raw dataset is in the following format:

Table 1: Raw data

	source	**target**	**rating**	**time**
1	160	1	10	2014-03-13 04:00:00
2	95	1	9	2013-11-16 05:00:00
3	377	1	7	2014-10-31 04:00:00

Each row of the raw data is a record of rating from one user to the other, which can also be considered as a record of Bitcoin transaction between two users. The raw data contains 4 variables. The source column is the user ID of the source of the transaction, or the rater. The target column is the user ID of the target of the transaction, or the rate. The rating column is the source's rating for the target, regarding to that specific transaction. The rating ranges from -10 to +10 in steps of 1. The time column is the time of the transaction. This raw data includes records of transactions from 2010 to 2016. Each unique user ID is a vertex of the network, and each recorded transaction is an edge. Number of edges each year is in the following table:

Table 2: Edges per year

year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Number of edges	98	7603	7250	6121	2735	362	17

To make the dataset more relevant to my topic of focus, several modifications were made. 20 levels of rating are too much, so I calculated a mean score of rating for each unique user and categorized users into five reputation levels based on their average score received. From -10 to -6, the users are "Highly Untrustworthy"; from -6 to -2, the users are "Somehow Untrustworthy"; from -2 to +2, the users are "Neutral"; from +2 to +6, the users are "Somehow Trustworthy"; from +6 to +10, the users are "Highly Trustworthy".

These levels are assigned to vertices in the network as a vertex attribute. Then, I filtered out rows of data in November and December of 2013 to create two datasets for each month. Rating value of each transaction is considered as edge weight in the network. The finalized dataset is in the following format:

Table 3: Revised data

	source	target	rating	day	source_level	target_level
1	user 888	user 1886	1	25	Somehow Trustworthy	Neutral
2	user 649	user 12	1	22	Neutral	Somehow Trustworthy
3	user 649	user 125	2	29	Neutral	Neutral

Analysis Results

Firstly, we treat two months' dataset as two complete networks. To get a sense of the distribution of the reputation level as a vertex attribute across vertices, the status of the frequency of each level in both months is summarized in two tables:

Table 4: November frequency per level

Highly Trustworthy	Somehow Trustworthy	Neutral	Somehow Untrustworthy	Highly Untrustworthy
2	294	628	40	6

Table 5: December frequency per level

Highly Trustworthy	Somehow Trustworthy	Neutral	Somehow Untrustworthy	Highly Untrustworthy
2	230	423	16	3

We see that in both months users in Neutral reputation level make up roughly 60% of the frequency and users in Somehow Trustworthy make up roughly 30% of the frequency, while the other 3 levels make up the rest roughly 10% together. The core users are in Neutral or Somehow Trustworthy reputation level. Therefore, many of the transactions were rated positive buy the users. In November, the mean rating of each transaction is 4.7257. In December, the mean rating of each transaction is 3.2430. Hence, the operation situation of Bitcoin Alpha platform was in a relatively safe and healthy condition. However, when we compare two tables, there is a significant drop in the number of users in all levels. And the mean rating also decreases from November to December. These are probably caused by the policy limitation from People's Bank of China at the beginning of December.

Similarly, we can consider each transaction from one user to another as an edge. Summarizing these transactions in a two-way matrix allows us to get some rough preliminary sense as to the distribution of such edges throughout the network:

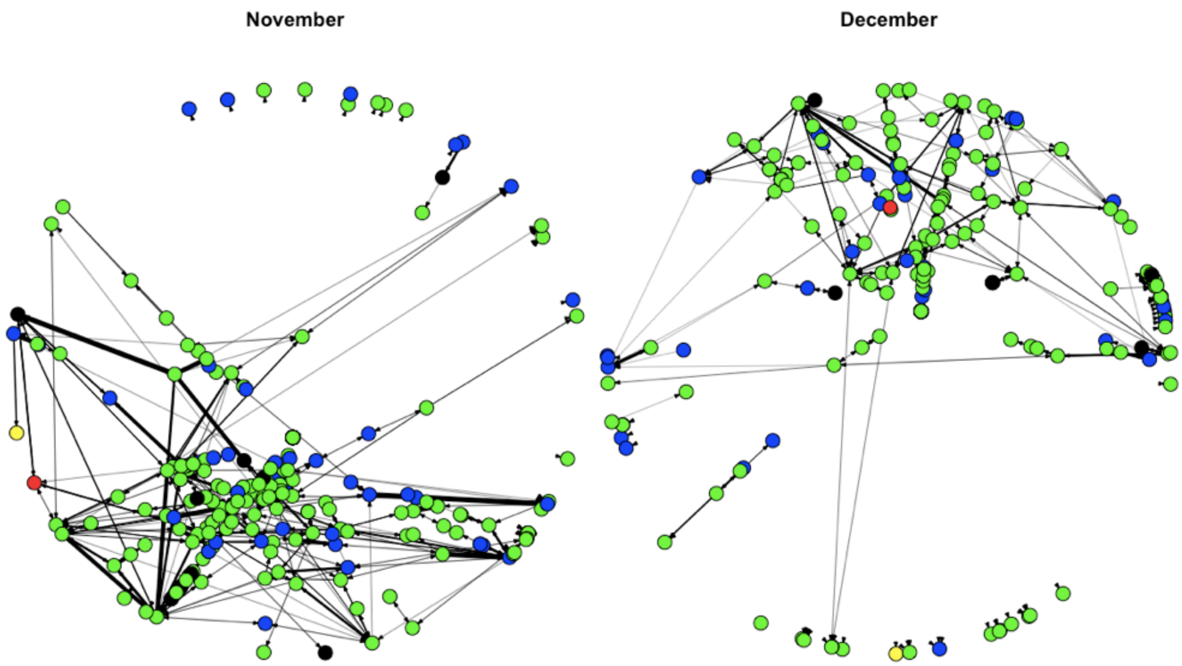
Table 6: November edges per contact

	Highly Trustworthy	Somehow Trustworthy	Neutral	Somehow Untrustworthy	Highly Untrustworthy
Highly Trustworthy	0	2	0	0	0
Somehow Trustworthy	2	42	188	17	3
Neutral	0	188	207	23	3
Somehow Untrustworthy	0	17	23	0	0
Highly Untrustworthy	0	3	3	0	0

Table 7: December edges per contact

	Highly Trustworthy	Somehow Trustworthy	Neutral	Somehow Untrustworthy	Highly Untrustworthy
Highly Trustworthy	0	2	0	0	0
Somehow Trustworthy	2	41	137	8	1
Neutral	0	137	138	8	2
Somehow Untrustworthy	0	8	8	0	0
Highly Untrustworthy	0	1	2	0	0

We can easily observe that the largest number of transactions was between Neutral reputation level users and between Neutral and Somehow Trustworthy reputation level users. When we compare the two tables, we can observe that the interaction between groups of users decreases from November to December. This can be directly related to the policy limitation from People's Bank of China at the beginning of December. Furthermore, the number of users in each reputation level per month and the majority of user reputation levels are more directly represented by the directed and weighted network visualizations in the layout of Fruchterman.Reingold, where five levels of reputation are represented by yellow, blue, green, black, and red vertices:



Next, the 2-month time span is divided in to 8 weeks. Hence, we can have a closer and more detailed dynamic analysis rather than the static network with less real insight. The number of transactions per week and the mean rating per week is in the following table:

Table 8: November number of transactions per week

Week 1	Week 2	Week 3	Week 4
76	80	157	145

Table 9: December number of transactions per week

Week 1	Week 2	Week 3	Week 4
108	103	53	55

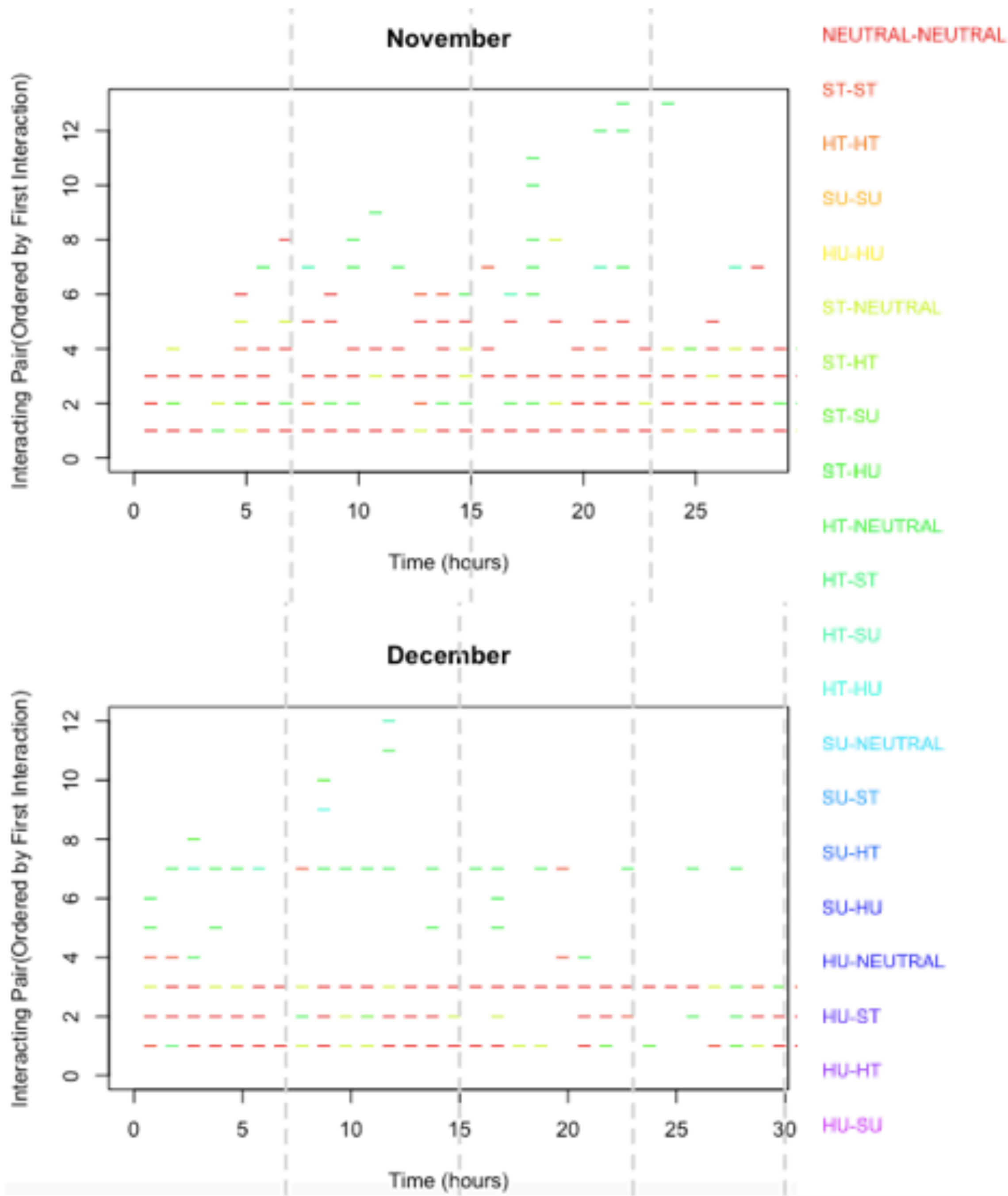
Table 10: November mean rating per week

Week 1	Week 2	Week 3	Week 4
0.6414	0.6751	1.3249	1.2236

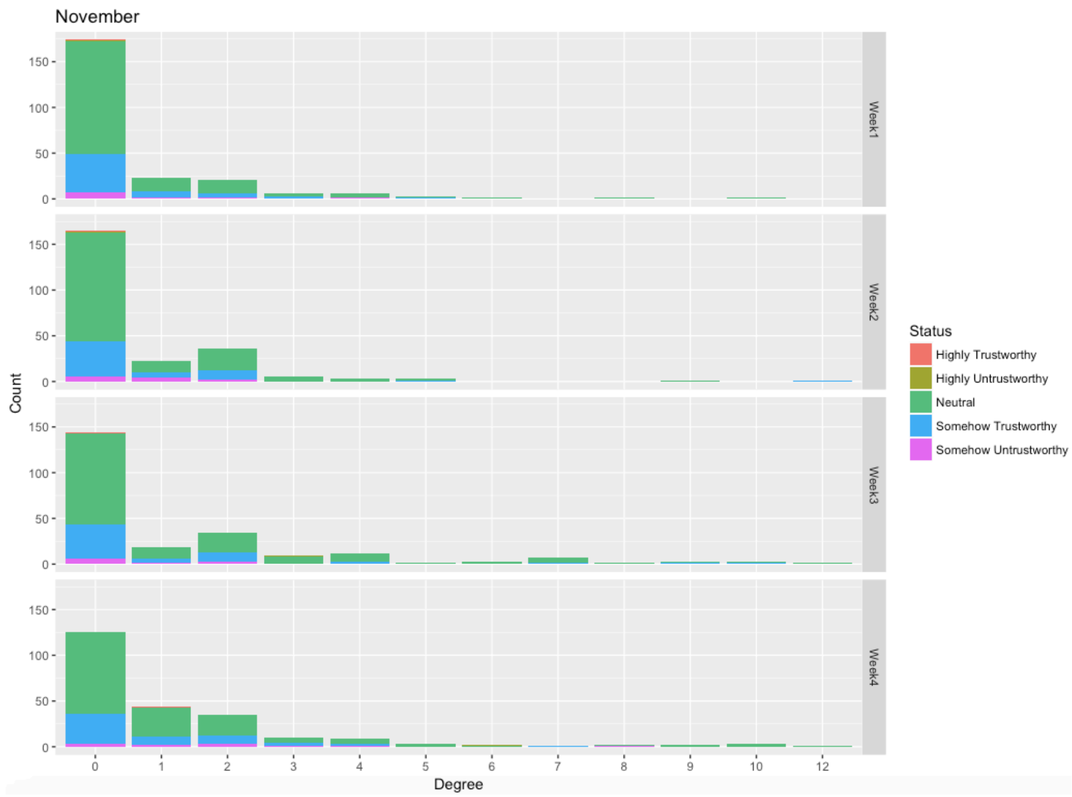
Table 11: December mean rating per week

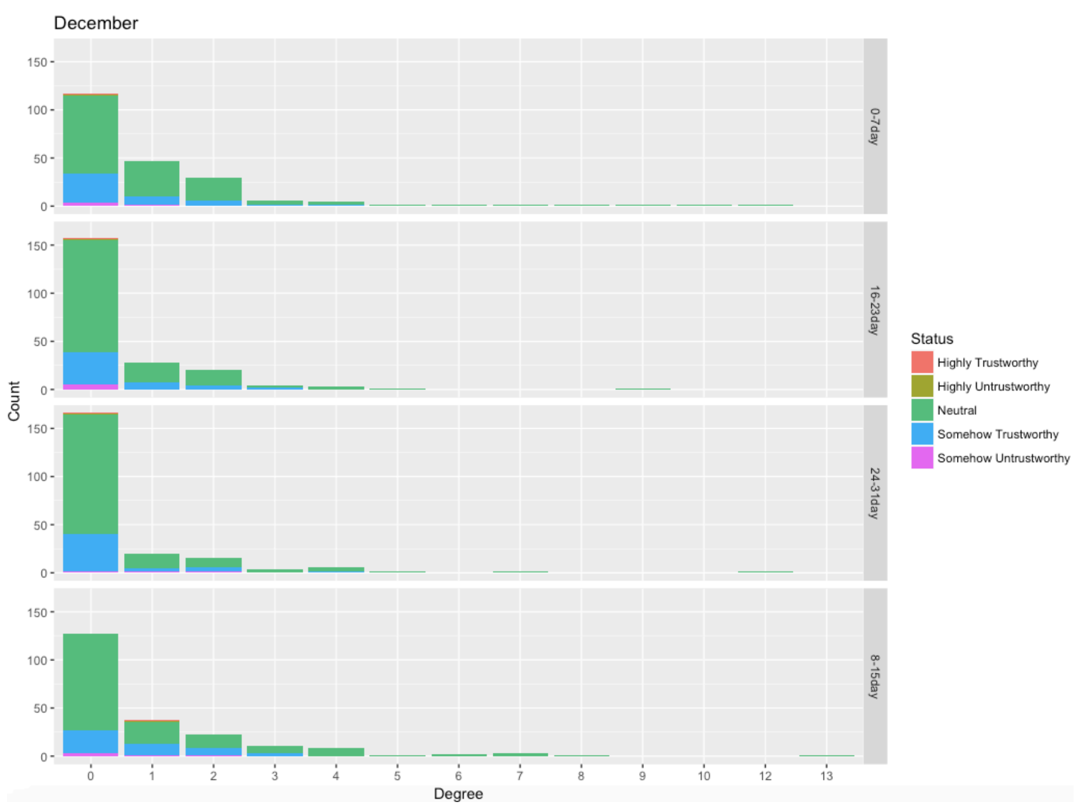
Week 1	Week 2	Week 3	Week 4
1.0093	0.9626	0.4953	0.5140

It is not surprisingly to find out that the number of transactions and mean rating in Week 3 and Week 4 in November rapidly increases. The significant increase in price per Bitcoin in late November attracts and motivates more users with higher reputation level to become active in the market. Also, users are more willing to give higher ratings when the market is performing well. On the other hand, number of transactions and mean rating decreases remarkably when it comes into December. The number of transactions keeps decreasing and finally reaches a lowest value, 55 recorded transactions, of the two months. The policy limitation from the People's Bank of China notably decreased the number of users with higher reputation level or even increased the number of users with lower reputation level at the same time, and discouraged their enthusiasm towards Bitcoin trading. Also, users attitude toward Bitcoins were more negative and passive. What's more, more temporal information can be observed in the following timeline visualization. For example, the presence and absence of dynamic edges in the dynamic network. Interactions between Neutral reputation level users and other levels users are quite constant throughout the 8-week time span, but the number of edges per interaction is remained at a low level. On the other hand, the interactions between Somehow Trustworthy level users and other levels users are not as frequently as the former ones. But the number of edges per interaction is always at a higher level. Hence, although Somehow Trustworthy level and Neutral level users are both considered as the core of the network, their trading behaviors are different.



In the last part of the analysis, I examine the degree distribution for each week in the 2-month period, visualized as bar charts, where the stacked bars have been colored to indicate the relative frequency of vertices of each of the five levels. From the degree distribution, I conclude that during any 7-day period there is a substantial fraction of the users studied that were not involved in any recorded transactions.


November
Count
Degree
Week1
Week2
Week3
Week4
Status
Highly Trustworthy
Highly Untrustworthy
Neutral
Somehow Trustworthy
Somehow Untrustworthy


December
Count
Degree
0-7day
16-23day
24-31day
8-15day
Status
Highly Trustworthy
Highly Untrustworthy
Neutral
Somehow Trustworthy
Somehow Untrustworthy

Conclusion and discussion

After the descriptive analysis, I conclude that the two consecutive milestone events happened in November and December, 2013 do have profound and various effect on the user network of the Bitcoin Alpha platform. The significant increase in price per Bitcoin in late November attracts and motivates more users with higher reputation level to become active in the market. Also, users are more willing to give higher ratings when the market is performing well. On the other hand, the policy limitation from the People's Bank of China notably decreased the number of users with higher reputation level or even increased the number of users with lower reputation level at the same time, and discouraged their enthusiasm towards Bitcoin trading. Also, users attitude toward Bitcoins were more negative and passive. The number of transactions keeps decreasing and finally reaches a lowest value, 55 recorded transactions, of the two months.

I would like to continue to pursue this field. Hence, I think the most important future step will be modeling of the dynamic network. What's more, I think the igraph package is limited to produce interpretable network visualizations when facing a large amount of edges and vertices. However, I do see examples of large network visualizations produced by package visNetwork that are more interpretable.

Reference

1. S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. Edge Weight Prediction in Weighted Signed Networks. IEEE International Conference on Data Mining (ICDM), 2016.
2. Andy Greenberg (20 April 2011). "Crypto Currency". Forbes.com. Archived from the original on 31 August 2014. Retrieved 8 August 2014.
3. McDonnell, Patrick "PK" (9 September 2015). "What Is The Difference Between Bitcoin, Forex, and Gold". NewsBTC. Archived from the original on 16 September 2015. Retrieved 15 September 2015.
4. Allison, Ian (8 September 2015). "If Banks Want Benefits of Blockchains, They Must Go Permissionless". NewsBTC. Archived from the original on 12 September 2015. Retrieved 15 September 2015.

5. Matteo D'Agnolo. "All you need to know about Bitcoin". timesofindia-economictimes. Archived from the original on 2015-10-26.
6. Narayanan, Arvind; Bonneau, Joseph; Felten, Edward; Miller, Andrew; Goldfeder, Steven (2016). Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton: Princeton University Press. ISBN 978-0-691-17169-2.
7. "Blockchain". Investopedia. Archived from the original on 23 March 2016. Retrieved 19 March 2016. Based on the Bitcoin protocol, the blockchain database is shared by all nodes participating in a system.
8. Iansiti, Marco; Lakhani, Karim R. (January 2017). "The Truth About Blockchain". Harvard Business Review. Harvard University. Archived from the original on 2017-01-18. Retrieved 2017-01-17. The technology at the heart of bitcoin and other virtual currencies, blockchain is an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way.
9. Jerry Brito & Andrea Castillo (2013). "Bitcoin: A Primer for Policymakers" (PDF). Mercatus Center. George Mason University. Archived (PDF) from the original on 21 September 2013. Retrieved 22 October 2013.
10. Joy Crane, (December 28th,2017), "How Bitcoin Got Here: A (Mostly) Complete Timeline of Bitcoin's Highs and Lows", NYMAG.COM
11. Kolaczyk, E. D. (2009). Statistical Analysis of Network Data: Methods and Models. Springer Science & Business Media.
12. Kolaczyk, E. D., & Csárdi, G. (2014). Statistical analysis of network data with R (Vol. 65). Springer.