Bitcoin Data Analytics: Exploring Research Avenues and Implementing a Hadoop-Based Analytics Framework

Raj Sanjay Shah and Ashutosh Bhatia(✉)

Department of Computer Science and Information Systems,
Birla Institute of Technology and Science Pilani, Pilani, India
{f20171181,ashutosh.bhatia}@pilani.bits-pilani.ac.in

Abstract. Bitcoin is the most successful cryptocurrency since its inception in 2009 [30]. There are 18.1 million BTCs in circulation as of December 2019, which roughly translates to 149 Billion USD [12]. With Bitcoin's substantial market capitalization and unique features like pseudo-anonymity and immutability, it draws much attention from the researchers across the world. Despite this enormous spotlight towards Bitcoin, it remains under-researched because of the large size of the Bitcoin Data, (Roughly 250 GB) and the inability to process this data in small time. To explore avenues for further research, this article presents a survey of the recent advancements done regarding the big data analytics of the Bitcoin Cryptocurrency. Furthermore, we propose an analysis framework based on the Apache Hadoop ecosystem.

1 Introduction

Bitcoin is a decentralized digital currency on top of an immutable distributed ledger called the Block-chain, involving a large number of participants in a peer-to-peer network who validate and certify the transactions. The development of Bitcoin was motivated by the distrust in the current banking system as well as the need for privacy in the digital world. Bitcoin, with its cryptographically backed security, ease of access, minimal transaction costs and minimal setup requirements, soon grew up in popularity and is today being considered by many governments as an acceptable form of currency.

Bitcoin is often termed as a double-edged sword owing to the fact that while it ensures the anonymity of the users' identity, it exposes their transactions to the whole world. Bitcoin achieves Pseudo-anonymity through public and private key hashes that act as the identities of a person. Any Bitcoin user can generate multiple private and public key pairs. As no real-world identities are required to do transactions, Bitcoin has been used for a wide variety of illegal transactions. The identities in Bitcoin are pseudonymous, but at the same time, the entirety of the transaction history of any identity is available, which makes it easier to figure out who the person is. De-anonymization techniques have been extensively studied to understand the level of privacy in the blockchain and various algorithms have been discussed to determine suspect behaviour.

L. Barolli et al. (Eds.): WAINA 2020, AISC 1150, pp. 178–189, 2020.
https://doi.org/10.1007/978-3-030-44038-1_17

Bitcoin's consensus-based mechanism brings us the following question: who decides what to enter into the global transaction logs? All users broadcast transactions to the mining pool. Miners collect these transactions into a block and solve a computationally challenging puzzle (called the proof of work) to determine which miner's block is chosen. This miner receives 12.5 BTCs as an incentive for mining (also known as the block mining reward). The Bitcoin system tweaks the computationally challenging puzzle to make sure that a block is mined roughly every ten minutes; that is, one block is added to the global transaction log every ten minutes on average. The high exchange rate of BTCs to USDs has made mining a very appealing activity for many people around the world. People have started using faster machines for mining, which has led to a continuous increment in the toughness of the puzzle.

Apache Hadoop [2] is a distributed computing eco-system that is scalable, fault tolerant and well suited for processing large amounts of data. This paper discusses the feasibility and drawbacks of an analysis framework built on top of this ecosystem which can be used to run various algorithms on the blockchain data. It is worth noting that no prior work has been done in building an analytic framework on the hadoop ecosystem. Sahoo et al. [35] have proposed a blockchain framework built on the Hadoop ecosystem but their paper proposes a blockchain *built via* Hadoop while we propose a blockchain *analytics tool based upon* the Hadoop ecosystem.

Through a systematic literature review, this article presents a study on the recent developments in various research avenues regarding the Bitcoin blockchain data analytics. It is worth to mention here that the research works which leverage the potential of blockchain technology to solve the problems across various vertical do not come under the scope of this study. The paper starts with a brief background of Bitcoin cryptocurrency in Sect. 2. Section 3 provides a literature review and Sect. 4 gives a detailed explanation of the proposed framework. Section 5 concludes the paper with a discussion on the future research direction in the area of Bitcoin blockchain.

2 Background

Bitcoin first appeared in a white paper titled "Bitcoin: A Peer-to-Peer Electronic Cash System" authored under the name of Nakamoto [30].

The following summarizes the entire process of the Bitcoin network as described by Satoshi Nakamoto:
A user first generates at least one signing key-pair, and publicizes the public key, which represents her address to receive BTCs. There's no limit to the number of addresses an individual can use for transactions. In fact, the ideal number is equivalent to one for each transaction. To make a payment, one then broadcasts a transaction, which indicates the address of the recipient to her peers, who in turn broadcast it to their peers. Eventually, this transaction reaches a miner, who collects the transactions which were broadcasted, into a block, and works on finding a difficult proof-of-work for that block. When a node finds a proof-of-work, it broadcasts the block to all nodes. Double-Spent validity is checked before accepting it into the chain.

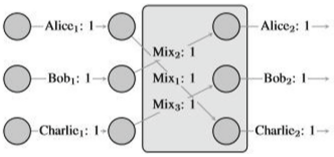

Fig. 1. Example of a bitcoin mixing service

3 Literature Review

The existing research works on Bitcoin cryptocurrency can be classified into following major categories. (1) De-anonymization, forensics and privacy; (2) Finance and Digital Currency Exchanges; (3) Carbon Footprint of Bitcoin and wasted resources; and (4) Mining pools and the degree of centralization. Along with the literature review, we examine the applicability of our proposed framework in each of the above mentioned major categories.

3.1 De-anonymization, Forensics and Privacy

Privacy is needed for user security while de-anonymization and forensics are needed to stop suspect activities. This is the reason why there is great research interest in both aspects of the Bitcoin pseudo-anonymity: Privacy and De-anonymization.

Many research papers focus on de-anonymization and privacy have used Wallet-Explorer [40] to get data of tagged wallet addresses. Wallet-Explorer has tagged more than tens of millions of public keys to various entities like gambling sites, exchanges, mining pools, etc. Wallet-Explorer uses a simple heuristic to merge addresses: two addresses are co-spent in a transaction, then they belong to the same person [34]. While this heuristic is valid only 60–70% of the time, we can still safely assume that both wallet-holders know each other because both have signed using their private keys.

It is important to note that Wallet-Explorer has stopped tagging new services since 2016.

3.1.1 User Privacy and Security

To increase the security of users, various websites and Wikipedia pages suggest not to use addresses more than once [42]. For addresses that receive BTCs, the address should not be used once the coins have been spent.

Implementing this to accept payments seems to be tougher, but services like Mycelium GEAR [16] and Coinbase [11] either pre-load many different addresses for receiving payments or upload a master public key to the server and use that master key to generate addresses. The master private key can be then used to spend the received money.

Other methods to increase privacy are mixing services. Mixing services are best understood with the help of Fig. 1, as explained by Möser et al. [29].

Mixing services make the tracing of Bitcoin transactions tougher. Referring to Fig. 1, suppose Alice1, Bob1 and Charlie1 all own 1BTC. They use mixing services and send coins to Alice2, Bob2 and Charlie2. Now a person cannot trace which coin used to belong to whom as random addresses send coins to Alice2, Bob2 and Charlie2. Mixing services are equivalent to money-laundering services in the context of Bitcoin.

Table 1. Possible types of attacks in various bitcoin management techniques

Type of attack	**Type of wallet**		
	Bitcoin core	SPV wallet	Bank/Exchange wallet
Bait and Switch	No	Yes	Yes
Direxť theft	No	No	Yes
Fabricated transaction	No	Partically safe	Yes
Chain high jacking	No	Yes	Yes
Unintentional transaction suppression	No	Yes	Yes
Intentional transaction suppression	No	Yes	Yes
Rewriting chain	Yes	Yes	Yes

Some of the popular mixing services are OnionBC [32], Bitcoin Fog [15], BitLaundry (discontinued now) [6] and Blockchain.info [7]. Möser et al. [29], in their paper, compared Bitcoin Fog, BitLaundry and Blockchain.info. They used the taint analysis tool of Blockchain.info to determine which mixing service better obfuscates the relationship between the sender and the receiver. They concluded that Bitcoin Fog and Blockchain.info provide enough anonymity and make it harder for a third party to relate between the input and the output addresses while BitLaundry is not reliable mixing service.

Better and complex alternative mixing methods have been suggested in [44] and [26]. Most papers do not consider monitoring the communications between the sender and the mixing services as a possible attack or the possibility that mixing services can be compromised. Wang et al. [41] proposed using escrow addresses to avoid scenarios where the mixing services are compromised and can steal your coins. The generation of the escrow address is done as proposed by Gennaro et al. in [17]. The unlinkability of addresses depends upon the address shuffling protocol used by [41]. The protocol proposed by Qi Wang et al. is based on participant cooperation; therefore, users are not required to pay additional fees to any mixing services.

Reid and Harrigan [34] have used the wallet-to-wallet network to conclude information about Bitcoin users. Egocentric Analysis and visualizations, TCP/IP layer information, and Flow and Temporal analysis have also been discussed by F. Reid and M. Harrigan. These discussions have led to de-anonymization tools

which will be discussed in the following section. Fretter and Harrigan [19] have explored the effectiveness of address clustering heuristics in the context of user privacy. These heuristics have been then used by Harrigan et al. [20] to determine the effect on privacy via airdrops through a cross-blockchain analysis of addresses. They identify situations where a user has unintentionally disclosed information about their address ownership on the Bitcoin, Litecoin and Dogecoin blockchains, via their activity on the clam blockchain. The address-ownership-finding heuristic is not very strong because we do not link real-life identities with the addresses. The case study used by Harrigan et al. [20] found 2000 instances of ownership identification. The amount of information gain is very minimal from the analysis of each airdrop.

Kaushal et al. [25] have evaluated the security risks associated with using Bitcoin wallet services. They have compared the Bitcoin Core software, Bitcoin bank and exchanges and Simplified payment verification wallets (SPV) as shown in Table 1. They have shown that running the Bitcoin Core service [13] (that is running your own full node) is the safest wallet service in the Bitcoin network.

3.1.2 De-Anonymization and Forensics

This section focuses more on the tools and forensic methods that have been developed to study suspects and illegal activities. The major goals of forensic researches are:

- To identify patterns corresponding to suspect behavior given only the Bitcoin public ledger and link these suspect addresses to real-world identities
- Given a suspect address, find other suspect addresses.

Isenberg et al. [22], in their paper "exploring entity behavior on the Bitcoin Blockchain", developed a visual analytics interface to analyze transactions of any particular address. They extracted data from the Bitcoin Core client and stored it in a MongoDB database. They further used the clustering heuristic given by Reid and Harrigan [34] to combine addresses of an individual entity. While this tool is useful for empirical analysis, it has very limited in its capacity as it only gives a better visual of the transactions done by any entity. Another visualization tool has been proposed by McGinn et al. [27] in their paper "Visualizing dynamic Bitcoin transaction patterns". Most visualization tools use bottom-up approach consisting of finding useful information by observing addresses or address clusters and inferring activities like money laundering. The observatory used by McGinn et al. is a 64-screen high-resolution canvas comprising of 132-Million pixels; therefore, because of the large number of pixels available to them, they did not constrain themselves to the bottom-up approach. Their top-down system-wide

visualizations have revealed recurrence of high-frequency patterns of algorithmic distinct denial-of-service attacks on the Bitcoin network and identified phases of such an attack.

Meiklejohn et al. [28] are the first ones to introduce various clustering heuristics. Using these heuristics and visualization tools, they studied the effect of

the Satoshi Dice on the Bitcoin network, analyzed the silk road address 1DkyBEKt5S2GD[1], as well as the Mt. Gox Bitcoin exchange theft. Along with domain experts and tagged addresses, they empirically traced interactions between major institutions trading in Bitcoins. A similar methodology has been used in other papers [38] discussed below (See Footnote 1). An interactive visualization tool has been proposed by Sun et al. [38]. The objective of this tool is to find the relationship map of any suspicious bitcoin account. It allows us to jointly analyze the network of associated accounts. We can also explore the transaction history of any particular address and compare it to the general Bitcoin market. This gives us an insight into user behavior with respect to external influences. For example, we can infer that addresses related to mixing services exchange their BTCs for USDs before government policy changes occur. This tool also features two graphing perspectives:

- Transaction-centric graphs
- Entity-centric graphs

Users can easily retrieve bitcoin transactions in a graph by specifying the range of trading volume and length of the circulation. The view panel of the tool gives the cumulative number of connections and the cumulative number of entities along with various qualitative measures such as entities with most predecessors or successors, and outstanding connections. The historical trend section of the tool gives a detailed overview of the historical Bitcoin statistics like the trade volumes, price and market capitalization.

Hirshman et al. [45] used a K-means clustering algorithm over a preprocessed Bitcoin dataset to find anomalous behavior. They were able to identify patterns of users who conducted transactions in an atypical fashion. They found out that possible addresses related to mixing services receive/spend very large as well as a very small number of Bitcoins.

Other forensic methods are discussed in [39] and [14]. Horst et al. [39] take a look into the Bitcoin clients and study the memory fingerprints of clients Bitcoin core and Electrum when they are not completely encrypted using private keys. The use cases of such analyses are limited because of the dynamic nature of any computer/laptop's memory.

Many companies and online tools provide suspect behavior analysis like Chain-analysis reactor [33] which maps transactions to darknet markets and c-hound [10] which claims to provide an AI-powered sophisticated Bitcoin and other block-chain analysis.

3.2 Finance and Digital Currency Exchanges

Bitcoin has been in the media spotlight from the last ten years because of the high exchange rate between Bitcoins and US dollars and the large fluctuations in the above-mentioned exchange rate. Many people invest in Bitcoin hoping

[1] Complete Address: 1DkyBEKt5S2GDtv7aQw6rQepAvnsRyHoYM.

that the exchange rate further increases. This has led to a large amount of research into predicting future exchange rates. Georgoula et al. [18] use a time-series analysis to understand the relationship between Bitcoin prices and the economic variables, technological factors, and the empirical measurement of the total public reception through twitter feeds. Various interesting results have been obtained through the research:

- Short-run regressions show that the Twitter feed sentiments are directly correlated with the Bitcoin prices.
- The short-run value of Bitcoins is negatively affected by the USD and the euro exchange rate.
- The long-run analysis shows that Bitcoin is negatively related to the Standard and Poor's 500 stock market index (which indicates the general state of the global economy).

While this paper identifies some of the factors affecting the Bitcoin exchange rates, it only finds whether correlations of variables are positive or negative, not the value of each correlation (we cannot compare the effect of two variables with high probability on the Bitcoin prices). Shah and Zhang [36] use Bayesian regression to build a model for price prediction. Jang and Lee [23] discuss a Bayesian Neural Network-based.

Bitcoin has an interesting relationship with various Bitcoin exchanges. Around half of the Bitcoin exchanges close down and thus are unreliable in nature. Yue et al. [46] have provided a multi-functionality tool that performs different types of analysis on Bitcoin currency exchanges. The comparison view of the tool compares multiple exchanges' different indices. This helps a user reliably see the performance of exchanges over time in comparison to other exchanges. The massive sequence view of the tool gives an overview of the Bitcoin exchange market. Users can examine the holistic connections of any particular exchange using the connection view. The data (addresses, historical conversion rates, etc.) of 60 most-used exchanges is stored into a MongoDB database along with the historical data of major Bitcoin news and policy changes. This entire data is then processed to figure out the network standing of each exchange. Temporal analysis of news related to Bitcoin and the historical data of exchanges give us the impact of government policies on the BTC to USD rates. This tool provides comprehensive information about Bitcoin exchanges but is limited as it extracts exchange addresses from Wallet-Explorer [40]. Domain experts determine the news be excluded or included for analysis, which further brings error through human mistake and post-hoc fallacy.

3.3 Carbon Footprint of Bitcoin and Wasted Resources

To maintain an average block mining time of 10 min, the difficulty of the computationally challenging puzzle increases/decreases every 2016 blocks. This has led to a hardware race to get the mining reward.

O'Dwyer et al. [31] have proposed an estimation to determine the energy cost to solve the proof of work puzzle for mining one block. Considering the

current bitcoin puzzle difficulty hash rate, the estimated electricity cost to mine one block according to the formula proposed by [31] is 81,948.8297$. The current mining reward is 101,546.88$.$(12.5BTC * 8123.75\$/BTC)$.

Due to the intrinsic uncertainties in the energy estimation calculations, there are many papers that suggest alternate consensus algorithms that do not require the hash puzzle like the model adopted by ethereum [43] (proof of stake). Variants of proof of stake and proof of work have also been introduced in various other cryptocurrencies like NXT, Peercoin, for proof of stake and litecoin and Dogecoin for proof of work. Barrang et al. [5] have created a new consensus algorithm, which is a variant of the proof of stake algorithm. This algorithm and many others provide solutions to the electricity consumption problem, but they also make the block-chain susceptible to many attacks that are avoided through the proof of work puzzle.

3.4 Mining Pools and the Degree of Centralization

Mining pools bring some levels of centralization with their collective behaviour. BTC.com [9], Slush [37] and ANTPool [1] combined mine a total of 37% of all blocks mined. This gives some leverage to the pools as they can perform a Denial of service attacks by creating a new fork in the block-chain to prevent some transactions. Most research papers discuss other aspects of mining pools: block with-holding attacks and its variants. Block with-holding attack occurs when a pool member finds a block but does not tell the rest of the pool to keep all the rewards. Bag et al. [4] have analyzed a type of block with holding attack and proposed a solution to all such attacks.

The fundamental motive of each mining pool is to increase profits without exceeding 50% of the total bitcoin computation power. More than 50% computation power would lead to loss of user confidence from the Bitcoin system and the BTC to USD prices will drastically drop. The actual mark is much lesser than 50% and closer to 30% according to the authors of Bitcoin and Cryptocurrency Technologies–A Comprehensive Introduction [3]. This profit-maximizing behaviour leads to competition between miners. Consider this scenario: The top 3 mining pools mine a total of 40% of all the blocks mined. A new pool gains momentum and is a threat to the top 3 pools. The top three pools actively conduct a DOS attack by forking all blocks added by the new pool. Other smaller pools and individual miners conform to the "might" of the top pools because they fear retribution from them.

The combined effect of the top pools, individual miners and the smaller pools would lead to the new pool decreasing its members to stay in the mining race, thus also decreasing its computation power. The situation mentioned above occurs in the actual bitcoin network as well, but the frequency of such situations has not been analyzed. Moreover all attacks of similar kind have not been found and analyzed.

4 The Data Analytic Framework

Bitcoin wallet files are stored in BerkeleyDB 4.8 and Blockchain indexes are stored in LevelDB. To convert the data into a more usable format, we used the parser of BlockSci v-0.5 [24]. Running the BlockSci parser over a full node requires a computer with 64 GB of RAM (32 GB minimum). After obtaining the parsed data, we convert the data into a format compatible with CassandraDB with the help of graphsense [21] (Fig. 2).

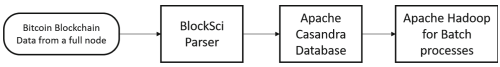

Fig. 2. Overview of the parallel processing framework

Using CassandraDB gives us two major advantages:

1. Cassandra Query Language enables us to query the database, retrieve a part of the database and support major features of any query language.
2. Despite saving the data into the database, we can periodically add the new blocks of the Block-chain without reconverting the entire data. The entire above mentioned parsing phase can be done incrementally.

The above steps convert the data into a form ready for the Map-Reduce stages of Hadoop. Every algorithm in Hadoop is required to have the corresponding map-reduce stages, because all algorithms cannot be written with map-reduce stages, given below are all the algorithms within the scope of the topics discussed in the literature review:-

1. Classifying legal and illegal transactions via the K-nearest neighbour. We applied the KNN algorithm over the data-set uploaded by Brugere [8]. The data-set fulfills one of the two criteria for using KNN on Hadoop: while it had the required tagged data, the data-set contained only around 150,000 records. Because of the lack of sufficient data, we were only able to test the code but were unable to test the speed up as compared to the computation time required on a single computer.
2. Top down approach to finding suspect behaviour. We pass the entire block-chain data into the Hadoop based clustering algorithms. We notice that while there is a significant speed-up by using Hadoop, the total time still comes out to be in days. 150 million transactions take 47 h to process in a Hadoop cluster containing two nodes when running K-means algorithm with the value of K as 5. This can be optimized in the future by using Apache Spark which is more suited to the iterative nature of clustering algorithms.
3. Hadoop is also used to merge wallets efficiently on the basis of the heuristic given by Reid et al. [34].

5 Conclusion and Future Research

There has been a large amount of research based on de-anonymization and privacy risks to users. Many solutions to increment privacy have been given, but such solutions introduce a degree of centralization via third party institutions. De-anonymization and forensic techniques rely heavily on clustering heuristics, visualization modules and then empirical analysis. This study model works well when looking at a bottom-up approach; that is: given a suspicious account it is easy to find other suspicious accounts through transaction linking. But it is very hard to look at just the transaction history, determine anomalous patterns and find out suspicious accounts. Furthermore, most of the tools discussed use Wallet-Explorer tagged addresses. There is no reliable method for tagging suspect bitcoin address to real-life identity.

Financial aspects of the Bitcoin have always been in the limelight. There are many tools and platforms that predict future prices. Digital currency exchanges have also been analyzed in detail, but it is important to note that not all of the wallets of many exchanges are known. Future research can analyze the indirect BTC volume flows between crypto-currency exchanges to estimate the transaction volume between countries.

The final research prospect is the degrees of centralization of the mining pools. The mining pools are some of the most experienced users of the blockchain technology. The strategies and attacks used by them to influence the block mining in their favour, have to be understood to maintain the user confidence in the system.

The current version of the data analysis framework uses Hadoop and CassandraDB that allow for a faster computation of algorithms. Hadoop works the best in batch processing tasks like finding the K-nearest neighbour and merging addresses. At the same time, many scenarios where Hadoop does not work or does not give the most optimal speed up have been enumerated. Further research can be done to find big data ecosystems (like Apache Spark and Apache Flink) that work better in situations where Hadoop fails. Furthermore, the findings of this paper: the advantages of Apache Hadoop for big data analytics of Bitcoin Block-chain can be incorporated with other researched ecosystem to form a complete and self-sufficient framework. This analysis can also be extended to other crypto-currencies like Doge coin and Lite coin.

References

1. AntPool: Mining pools. https://v3.antpool.com/home
2. Apache: Apache hadoop. https://hadoop.apache.org/
3. Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S.: Bitcoin and Cryptocurrency Technologies—A Comprehensive Introduction. Princeton Press, Princeton (2016)
4. Bag, S., Ruj, S., Sakurai, K.: Bitcoin block withholding attack: analysis and mitigation. IEEE Trans. Inf. Forensics Secur. **12**(8), 1967–1978 (2017). https://doi.org/10.1109/TIFS.2016.2623588

5. Berrang, P., von Styp-Rekowsky, P., Wissfeld, M., França, B., Trinkler, R.: Albatross – an optimistic consensus algorithm. In: 2019 Crypto Valley Conference on Blockchain Technology (CVCBT), pp. 39–42 (2019). https://doi.org/10.1109/CVCBT.2019.000-1
6. BitLaundry. http://app.bitlaundry.com/
7. Blockchain.info. https://blockchain.info/q/
8. Brugere, I.: Bitcoin-transaction-network-extraction. https://github.com/ivanbrugere/Bitcoin-Transaction-Network-Extraction
9. BTC.com: Mining pools. https://btc.com/
10. C-hound. https://www.c-hound.ai/
11. Coinbase: Coinbase-wallet. https://wallet.coinbase.com/
12. Coindesk: Bitcoin(USD) price. http://www.coindesk.com/price/
13. Bitcoin core. https://bitcoin.org/en/bitcoin-core/
14. Domingues, P., Frade, M., Parreira, J.: Filtering email addresses, credit card numbers and searching for bitcoin artifacts with the autopsy digital forensics software, pp. 318–328 (2020). https://doi.org/10.1007/978-3-030-17065-3_32
15. Bitcoin fog. http://www.bitcoinfog.info/
16. Gear, M.: Start accepting bitcoin payments. https://gear.mycelium.com
17. Gennaro, R., Jarecki, S., Krawczyk, H., Rabin, T.: Secure distributed key generation for discrete-log based cryptosystems. J. Cryptol. **20**(1), 51–83 (2007). https://doi.org/10.1007/s00145-006-0347-3
18. Giaglis, G., Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D.: Using time-series and sentiment analysis to detect the determinants of bitcoin prices (2015). https://doi.org/10.2139/ssrn.2607167
19. Harrigan, M., Fretter, C.: The unreasonable effectiveness of address clustering. In: 2016 International IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 368–373 (2016). https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0071
20. Harrigan, M., Shi, L., Illum, J.: Airdrops and privacy: a case study in cross-blockchain analysis. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 63–70 (2018). https://doi.org/10.1109/ICDMW.2018.00017
21. Haslhofer, B., Karl, R., Filtz, E.: O bitcoin where art thou? Insight into large-scale transaction graphs. In: SEMANTiCS (Posters, Demos) (2016)
22. Isenberg, P., Kinkeldey, C., Fekete, J.-D.: Exploring entity behavior on the bitcoin blockchain. In: Posters of the IEEE Conference on Visualization (2017)
23. Jang, H., Lee, J.: An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information. IEEE Access **PP**, 1 (2017). https://doi.org/10.1109/ACCESS.2017.2779181
24. Kalodner, H., Goldfeder, S., Chator, A., Möser, M., Narayanan, A.: Blocksci: Design and applications of a blockchain analysis platform (2017)
25. Kaushal, P.K., Bagga, A., Sobti, R.: Evolution of bitcoin and security risk in bitcoin wallets. In: 2017 International Conference on Computer, Communications and Electronics (Comptelix), pp. 172–177 (2017). https://doi.org/10.1109/COMPTELIX.2017.8003959
26. Liu, Y., Li, R., Liu, X., Wang, J., Tang, C., Kang, H.: Enhancing anonymity of bitcoin based on ring signature algorithm. In: 2017 13th International Conference on Computational Intelligence and Security (CIS), pp. 317–321 (2017). https://doi.org/10.1109/CIS.2017.00075

27. McGinn, D., Birch, D., Akroyd, D., Molina-Solana, M., Guo, Y., Knottenbelt, W.: Visualizing dynamic bitcoin transaction patterns. Big Data **4**, 109–119 (2016). https://doi.org/10.1089/big.2015.0056
28. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., Mccoy, D., Voelker, G., Savage, S.: A fistful of bitcoins: Characterizing payments among men with no names. Commun. ACM **59**, 86–93 (2016)
29. Möser, M., Böhme, R., Breuker, D.: An inquiry into money laundering tools in the bitcoin ecosystem. In: 2013 APWG eCrime Researchers Summit, pp. 1–14 (2013). https://doi.org/10.1109/eCRS.2013.6805780
30. Nakamoto, S., et al.: Bitcoin: A peer-to-Peer Electronic Cash System (2008)
31. O'Dwyer, K.J., Malone, D.: Bitcoin mining and its energy footprint. In: 25th IET Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014), pp. 280–285 (2014). https://doi.org/10.1049/cp.2014.0699
32. OnionBC. http://6fgd4togcynxyclb.onion/
33. Chain-analysis reactor. https://www.chainalysis.com/
34. Reid, F., Harrigan, M.: An analysis of anonymity in the bitcoin system. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 1318–1326 (2011). https://doi.org/10.1109/PASSAT/SocialCom.2011.79
35. Sahoo, M.S., Baruah, P.K.: Hbasechaindb - a scalable blockchain framework on hadoop ecosystem. In: Yokota, R., Wu, W. (eds.) Supercomputing Frontiers, pp. 18–29. Springer International Publishing, Cham (2018)
36. Shah, D., Zhang, K.: Bayesian regression and bitcoin. In: 2014 52nd Annual Allerton Conference on Communication, Control, and Computing, Allerton 2014 (2014). https://doi.org/10.1109/ALLERTON.2014.7028484
37. Slush: Mining pools. https://slushpool.com/home/
38. Sun, Y., Xiong, H., Yiu, S.M., Lam, K.Y.: Bitvis: An interactive visualization system for bitcoin accounts analysis. In: 2019 Crypto Valley Conference on Blockchain Technology (CVCBT), pp. 21–25 (2019). https://doi.org/10.1109/CVCBT.2019.000-3
39. Van Der Horst, L., Choo, K.R., Le-Khac, N.: Process memory investigation of the bitcoin clients electrum and bitcoin core. IEEE Access **5**, 22385–22398 (2017). https://doi.org/10.1109/ACCESS.2017.2759766
40. WalletExplorer: smart bitcoin block explorer. http://www.WalletExplorer.com
41. Wang, Q., Li, X., Yu, Y.: Anonymity for bitcoin from secure escrow address. IEEE Access **6**, 12336–12341 (2018). https://doi.org/10.1109/ACCESS.2017.2787563
42. Wikipedia: Address reuse. https://en.bitcoin.it/wiki/Address_reuse
43. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj. Yellow Pap. **151**, 1–32 (2014)
44. Xiao, R., Ren, W., Zhu, T., Choo, K.R.: A mixing scheme using a decentralized signature protocol for privacy protection in bitcoin blockchain. IEEE Trans. Dependable Secure Comput. 1 (2019). https://doi.org/10.1109/TDSC.2019.2938953
45. Huang, Y., Hirshman, Y., Macke, S.: Unsupervised approaches to detecting anomalous behavior in the bitcoin transaction network. URL https://pdfs.semanticscholar.org/2ea6/04d967ca11ec869545ace248c41db6a49855.pdf
46. Yue, X., Shu, X., Zhu, X., Du, X., Yu, Z., Papadopoulos, D., Liu, S.: Bitextract Interactive visualization for extracting bitcoin exchange intelligence. IEEE Trans. Visual Comput. Graphics **PP**, 1 (2018). https://doi.org/10.1109/TVCG.2018.2864814