Blockchain Analysis Tool For Monitoring Coin Flow

Aman Framewala, Sarvesh Harale, Shreya Khatal, Dhiren Patel, Yann Busnel, Muttukrishnan Rajarajan

Blockchain Analysis Tool For Monitoring Coin Flow

Aman Framewala[1], Sarvesh Harale[1], Shreya Khatal[1], Dhiren Patel[1], Yann Busnel[2], and Muttukrishnan Rajarajan[3]

[1]Department of Computer Engineering, VJTI Mumbai, India
Email: amanframewala@gmail.com, sarveshharale10@gmail.com, khatalshreya@gmail.com, dhiren29p@gmail.com
[2]IMT Atlantique, IRISA Rennes, France
Email: yann.busnel@imt-atlantique.fr
[3]City University London, UK
Email: R.Muttukrishnan@city.ac.uk

***Abstract*—While cryptocurrencies like Bitcoin have the potential to break traditional financial barriers, there are growing concerns about such currencies being used to fund illegal activities. Blockchain keeps the complete history of all transactions ever performed and each node replicates it. The humongous data it contains can be analyzed to gain useful insights about user transactions as well as the blockchain as a whole. In this paper, we propose an approach to parse and visualize the data of Bitcoin blockchain in a graph structure and carry out analysis that includes tracking and tracing, address clustering and entity tagging. We also try to find patterns in the data at a macro level to provide insights about the overall system. Thus, these efforts lead to foundation work for an analysis tool for getting insights on the coin flow of any financial system including cryptocurrencies.**

Keywords—Blockchain, bitcoin, tracking and tracing, address clustering, entity tagging,

I. Introduction

Bitcoin has been favored by many people due to its decentralized and pseudo-anonymous nature. The popularity of Bitcoin has continued to rise with over 200k transactions being recorded each day [1][2]. At the same time, Bitcoin is widely used as a means of exchange for dark markets like the Silk Road studied by [3], which was infamous for drugs, human trafficking and also for activities such as money laundering and extortion [4]. This has led to an urgent need for law enforcement agencies to monitor the flow of Bitcoin, detect such activities and further deter them. However, binary-formed data of the Bitcoin blockchain make it cumbersome for the agencies to perform analysis for obtaining usable evidence from scratch [5]. Bitcoin is a pseudo-anonymous currency [6], in which all the transactions are visible and traceable, but the Blockchain does not store an information which allow direct mapping to the real-world entities, thus providing anonymity [7][8]. One of the motives of cryptocurrencies is to provide anonymity and this has led to the formation of new cryptocurrencies like Monero [9] and ZeroCash [10] which enhance the anonymity of users. Other mechanisms like Bitcoin mixing services have also been developed which serve as a tool to provide anonymity by obfuscating the flow of funds [11], thus aiding in money laundering activities.

In this paper, we propose a tool to parse the Bitcoin blockchain data, visualize the transactions and analyze them with ease. It integrates the features of transaction graph analysis [12], address clustering [13], entity tagging [14], tracking tracing [15] and wallet monitoring using alerts into a single tool which is designed to suit the needs of monitoring coin flow. This can help financial institutions and law enforcement agencies in identifying criminal entities and investigating activities like money laundering and ransomware. The major contributions of this work are as follows:

- Establish a concrete methodology for analysis and monitoring of cryptocurrencies.
- Consolidate various analysis functions that can be performed on cryptocurrencies enabling greater auditability.

The rest of the paper is organized as follows: In Section 2, background and related work are presented. Section 3 discusses our proposal with the design rationale and techniques used. Section 4 gives implementation details and discusses results visualization. We conclude the paper in Section 5 followed by the references at the end.

II. Background and Related Work

Nakamoto [16] marks the inception of blockchain and Bitcoin in the world. It proposes the Bitcoin system as a peer-to-peer value transfer system. Bitcoin is a cryptocurrency, based on the UTXO (Unspent Transaction Output) model. Users can transact on the Bitcoin blockchain using Bitcoin accounts. A Bitcoin account is defined by an Elliptic Curve Cryptography key pair [5][17]. The Bitcoin account is publicly identified by its bitcoin address, obtained from its public key using a unidirectional function as shown in Figure 1. Using this public information user can send bitcoins to that address. Then, the corresponding private key is needed to spend the bitcoins of the account.

Table 1 shows a sample private key, its intermediate results and the corresponding Bitcoin address generated.

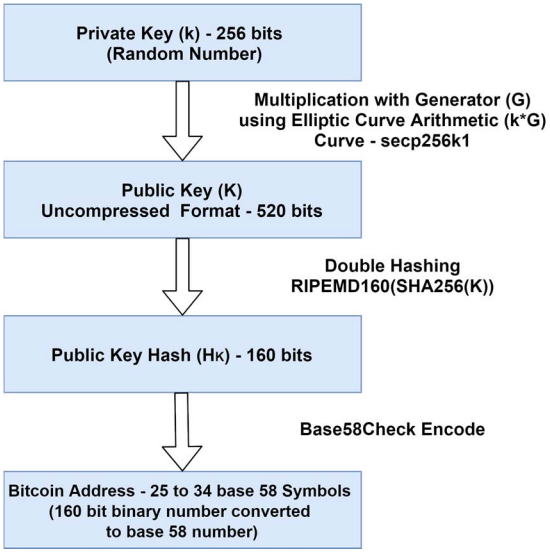
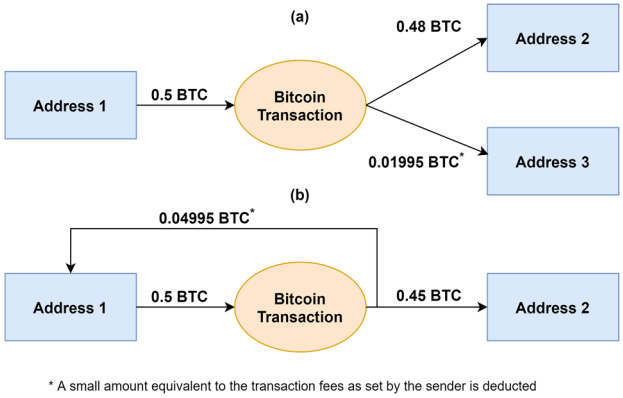


Fig. 1: Bitcoin Address Generation

Private Key	9e524de478970a9621c0e52890805d5f28e362 0892ba6bfa701b026c6ee10a52
Public Key	03ee3b7337eb52d1e8bd7ee271db9aa43a6775 0ff483870ab2753d2e13922970db
Public Key Hash	5355f7bb58765e07a20f978b6e2437e99a5e923
Bitcoin Address	18be54dbyAth7CR4ymeoQBpzwinLW5Qe1K

Table. 1: Bitcoin Address Example

It is easy to understand that any user can create any number of bitcoin addresses (generating the key pairs) using standard bitcoin client software. A transaction in Bitcoin is a transfer of value that is broadcast to the network and collected into a block. A transaction typically references previous transaction outputs (UTXO) as inputs to it and generates new transaction outputs (UTXO). Figure 2 represents typical bitcoin transactions. One can note that a small amount equivalent to the transaction fee gets deducted and is awarded to the miner. Figure 2 (b) shows



* A small amount equivalent to the transaction fees as set by the sender is deducted

Fig 2: Bitcoin Transactions

how change can be returned to the address 1, which gives input to the transaction.

Since blockchains provide auditability, it is possible to view every transaction ever recorded. These transactions can be analyzed to provide insights into emerging trends and sentiments concerning the use of the blockchain.

Spagnuolo et al. [18] propose a framework to automatically parse the blockchain, cluster addresses, classify addresses and users, export and visualize elaborated information from the Bitcoin network. They also implement a classifier that labels the clusters in an automated or semi-automated way, by using several web scrapers that incrementally update lists of addresses belonging to known identities. Cuneyt et al. [19] explore aspects of blockchain analytics such as analysis models, tools and use cases in the modern world. Fleder et al. [12] annotate the public Bitcoin transaction graph by trying to link Bitcoin public keys to real people – either definitively or statistically. The graph is then put through a graph-analysis framework to find and summarize the activity of both known and unknown users. They then use web scraping to find Bitcoin addresses and try to link them to real-world entities. Cuneyt et al. [20] present general algorithms for tracking Bitcoin flows. Ermilov et al. [13] propose heuristic methods for grouping addresses that might probably be controlled by a single entity which is an important step for analyzing transactions. They also recommend using off-chain information to be combined with blockchain information to further refine the results. Hong et al. [21] explores cryptocurrency mixing (laundry) services and proposes a general de-mixing algorithm for common mixing services by exploiting their static and dynamic parameters. Geodell et al. [22] present a study on electronic payment methods majorly focusing on cryptocurrencies and comparing their offered anonymity and auditability. They then propose two schemes of using cryptocurrencies which try to provide an acceptable level of anonymity to users and also providing a good degree of auditability to regulatory authorities. Jourdan et al. [14] propose that identities of Bitcoin address holders can be leaked based on transaction features or off-network information. Balthasar et al. [23] [24] briefly examines some of the most relevant Bitcoin laundry services and studies their

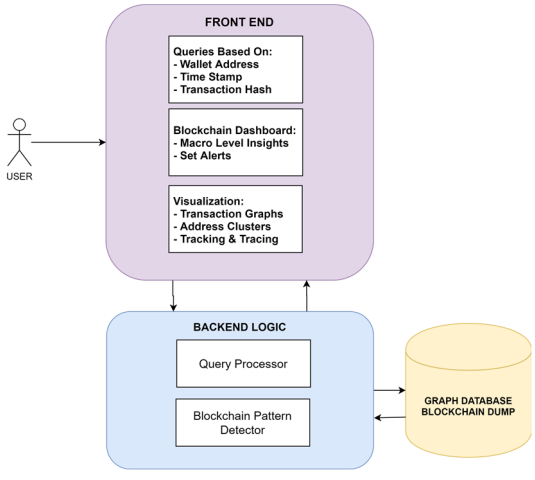


Fig. 3: Overview of the Blockchain Analysis Tool

main features mainly the security and anonymity provided by them. Balsakas et al.[26] provides a comprehensive study of blockchain analysis as a field of study. They explore the features of available blockchain analysis tools and categorizes them based on their provided functionality. It also presents the prevailing challenges on blockchain analysis.

III. Analysis Tool: Our Proposal and Design Rationale

We propose a system that integrates the features of transaction graph analysis, address clustering, entity tagging and tracking tracing into a single tool which is designed to suit the needs of monitoring coin flow. Figure 3 represents an overview of the Blockchain Analysis tool. The front end of the tool provides an interactive web-based GUI provided to make various queries, view statistics representing the current state of the bitcoin blockchain. This tool allows generation of alerts for transactions involving specific wallet address or a given transaction amount. The workflow for the backend of the tool can be seen in Figure 4 and has been described briefly in sub-section C.

A. Blockchain Data Migration Module

This module is responsible for getting the data to a graph database like *Neo4j*[27], where the processing of graph-related queries can be done quickly owing to the intuitive query interface. The process involved transferring the binary bitcoin dump into a database utilizing a parser and using other databases to speed up the process.

Figure 5 depicts our proposed method for the migration of a blockchain dump (i.e. Bitcoin blockchain) into a graph database (i.e. Neo4j). The process for migration of any blockchain to a graph database can be broadly broken down into the following steps:

1) Dump Processing: It consists of the following steps:

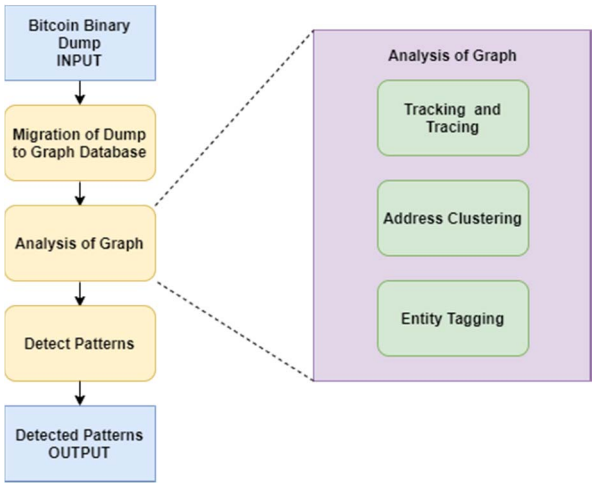


Fig. 4: Workflow for detecting patterns in blockchain

a) Bitcoin Parsing: After downloading the Bitcoin dump data, it needs to be parsed for converting it to a processable format. There are readily available libraries for parsing Bitcoin data which convert raw binary data into a structured form. The parser is used to make transactions from all the blocks available in a readable format.

b) Transaction Deserialization: After getting the transactions, they are deserialized into objects having fields as transaction hash, timestamp, inputs and outputs.

c) Inputs and Outputs Aggregation: In Bitcoin transactions, there might be a possibility that multiple inputs (or outputs) may relate to a single address. Therefore, the inputs (outputs) are aggregated to form a single input (output) from that address. This is done for brevity and convenience.

d) Bitcoin Unit Conversion: Delete Bitcoin transactions contain information about bitcoin amounts involved in the transaction in satoshis (10^{-8} BTC). These values are converted into BTC. This is again for brevity and convenience as there is no specific requirement for processing values at such a granularity.

2) Fields Extraction: After converting the transactions to structured form, essential fields are extracted which are used to migrate the data to Neo4j Graph Database. These fields are extracted and CSV files are created out of them. four types of CSV files are created with the following fields:

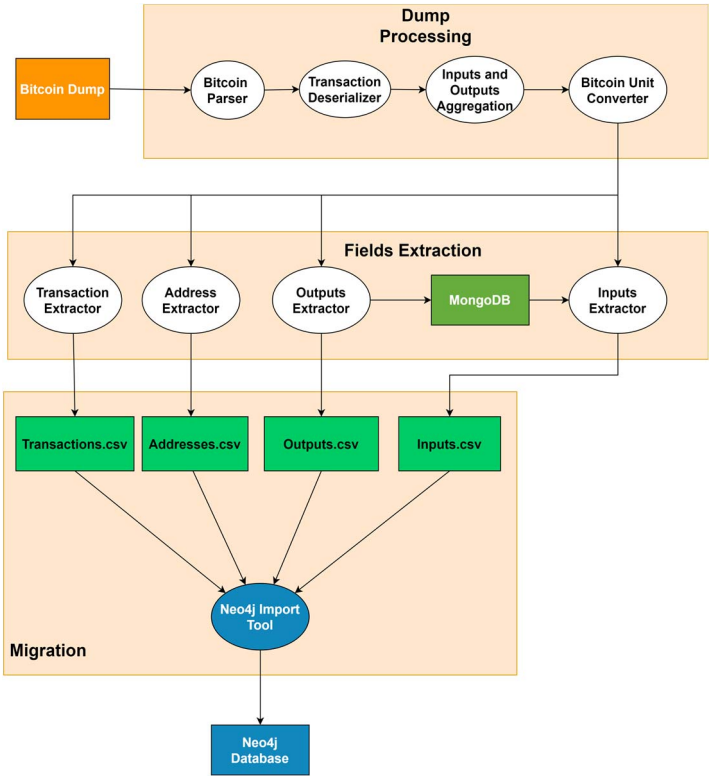


Fig. 5: Workflow for data cleaning and migration to database

Types	Fields
Transactions	Transaction Hash and Timestamp
Addresses	Bitcoin wallet addresses
Inputs	Transaction Hash, Address and Amount
Outputs	Transaction Hash, Address and Amount

Table. 2: Fields Extraction

The UTXOs are saved to MongoDB database which is used for creating Input CSV files. This is due to the structure of the Bitcoin transaction where inputs refer to the previous transaction and its output index. Thus there is a need to keep the UTXOs in a database due to insufficient memory (RAM) during preprocessing.

3) Data Migration: Once all the CSVs are created, they are migrated to *Graph Database* using the *Import Tool* provided by Neo4j.

Since the bitcoin blockchain is continuously appended with new transactions, one needs to run a cron job on a daily basis for synching the database with the latest state of the blockchain. The tool is proposed to have a button to force start a sync in realtime to carry out analysis.

We also propose a mechanism to generate alerts based on a certain wallet adress or transaction amount. The tool would monitor the transactions and provide a notification whenever the condition is met during the syncing of the database.

B. Analysis Of Graph

While monitoring coin flow, it is important to obtain insights from a transaction graph [12]. In this analysis, we have three main subsections. First is tracking and tracing of money through the various wallet addresses [15]. Second being address clustering [13], which tries to group wallet addresses operated by a single logical entity. There are two main ways to cluster addresses namely: (i) Common Spend and (ii) One time change as given in [13]. The third process being Entity Tagging which involves attempts to gain information about some addresses by using techniques like web scraping [25] and usage analysis.

1) Tracking and Tracing: Tracking refers to looking for transactions that use this transaction output and its subsequent transactions (forward direction). Tracing refers to looking for transactions that result in this transaction output and its previous transactions (backward direction).

2) Address Clustering: Address Clustering is the process of grouping multiple addresses such that all addresses are controlled by a single entity using heuristic methods. The entity can be a single person, a group of individuals or an organization. Address clustering may be inaccurate as it is based on heuristics. Figure 6 demonstrates the following 2 heuristics used for address clustering:

a) One Time Change: Change from a transaction is returned to the user through a new address.

b) Common Spending: All the addresses in the inputs of a transaction are controlled by a single entity

3) Entity Tagging: Entity Tagging refers to labeling the address clusters with a real-world entity. This can be done using scraping open-source information. Ex: Tagging a group of addresses operated by a cryptocurrency exchange.

Pattern Detection

In this stage, the behavior of the blockchain is analyzed against various parameters to gain insights at a macro level. The analysis involves market volume and price analysis, mapping of the news events to activities in the blockchain which could be measured as an increase or decrease in demand for the cryptocurrency or sudden rise in acceptance of a given cryptocurrency related to some event.

IV. Result Visualization

The various transaction graphs that are displayed include two type of nodes, the transaction and the wallet address nodes. They are represented by blue and orange color respectively. The former is uniquely identified by their transaction hash, while the latter by their wallet address. The edges represent the relationship that the wallet address has with the transaction. An incoming edge would represent an input to the transaction while the outgoing edge represents the output from a transaction being credited to the given wallet address.

A. Migration to Graph Database

The process of migration was successfully completed adding 470,162,363 transactions, 548,854,187 wallet addresses connected by 793,453,561 aggregated inputs and 1,179,067,970 aggregated outputs. Since there is no limit to the number of wallet addresses a user can create, several of these addresses maybe used for just a couple of transactions to obfuscate the trail. The total disk space used was 400GB including indices. Figure 7 depicts a transaction graph obtained with input given as the wallet address. The graph consists of the given wallet address at the center surrounded by other nodes, which are linked by the transaction hash.

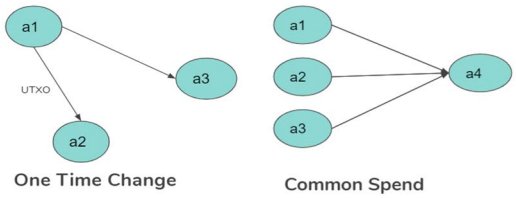


Fig. 6: Address Clustering based on Heuristics

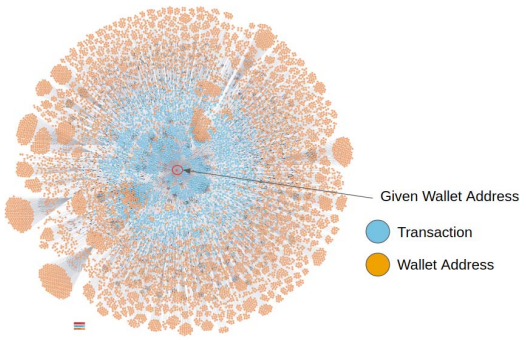


Fig. 7: Bitcoin transaction graph for a Given Wallet

B. Tracking and Tracing

Figures 8 and 9 represent the transaction graphs for randomly chosen wallet address showing the one-hop tracking and the two-hop tracking respectively. The tracking operations were performed for a limited number of nodes to allow for ease in visualization. The visualizations show how the coin flow entering into the given wallet address across several hops, allowing us to reach the point of origination.

C. Address Clustering

Figure 10 depicts an address cluster obtained using the common spend heuristic. All the wallet addresses enclosed in the blue box provide inputs to the same transaction, thus according to the common spend heuristic they are considered to be controlled by the same entity or organization. This allows us to cluster addresses together to aid in entity tagging.

D. Entity Tagging

A group of 307,481 addresses was identified to be belonging to BTC-e.com which is an infamous cryptocurrency exchange. This information was obtained by scraping open source web data.

E. Pattern Detection

Patterns at macro level were monitored to obtain insights about the overall blockchain. We performed a sample tracking analysis in which we obtained the number of addresses in the first, second and third hop transactions originating from the random seed address "1EYSiRC2nUi2xLMTuwkWhHtpTT-VVZ6KNrz". The results are as follows:

- First hop: 82
- Second hop: 105
- Third hop: 140

Analyzing the blockchain at macro level revealed the following statistics:

- Total Transactions: 469608054 [30th October 2019]
- Total Volume of Money in Circulation: 18021375 BTC [30th October 2019]
- Number of Wallets added in a week: 2,62,972

17th November 2019 - 43428228
24th November 2019 - 43691200

V. CONCLUSION

This paper provides a foundation for a blockchain analysis tool to monitor the coin flow in a given blockchain. Currently, this tool is for the Bitcoin blockchain. However, it can be used with any blockchain by adding appropriate migration module in its modular organization. The tool has features including Tracking, Tracing, Address Clustering, and Entity Tagging. Further, it also finds patterns at macro level to gain insights from the data. The results obtained using this tool are insightful and encouraging. Future scope of this work includes support for other cryptocurrencies like Ethereum and making it a universal tool for blockchain analysis.

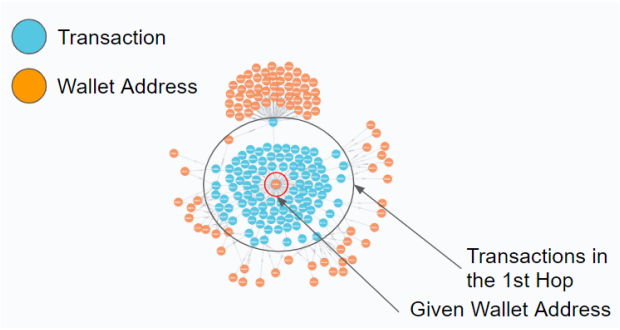


Fig. 8: Hop Tracking (length 1)

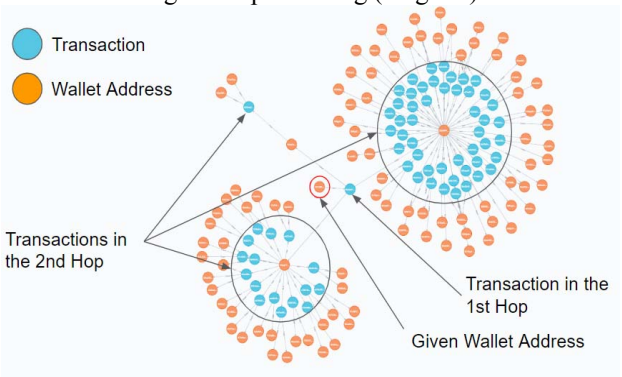


Fig. 9: Hop Tracking (length 2)

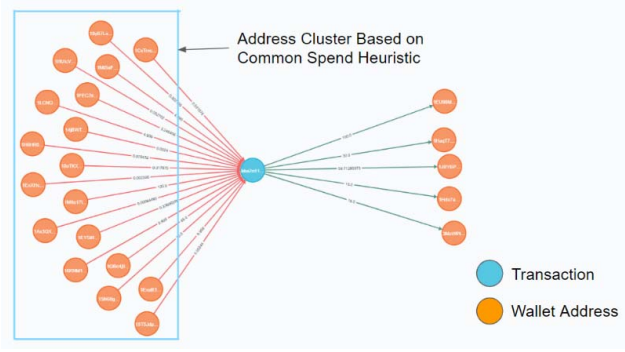


Fig. 10: Address Clustering Result on Data

VI. REFERENCES

1. Blockchain.com (as of 20/01/2020)
   URL : https://www.blockchain.com/en/charts
2. Patel, D., Bothra, J., Patel, V.: Blockchain exhumed. 2017 ISEA Asia Security and Privacy (ISEASP), Surat, 2017, pp. 1-12. Doi: 10.1109/ISEASP.2017.7976993
3. Christin, Nicolas. (2012). Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace. Proceedings of the 22nd International Conference on World Wide Web.
4. Foley, Sean & Karlsen, Jonathan & Putnins, Talis. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?. Review of Financial Studies. 32. 1798-1853. 10.1093/rfs/hhz015.
5. Heaven, D: Sitting with the cyber-sleuths who track cryptocurrency criminals. MIT technology Review. April 2018.
   URL : https://www.technologyreview.com/s/610807/sitting-with-the-cyber-sleuths-who-track-cryptocurrency-criminals/
6. Martins, Sergio, and Yang Yang. Introduction to bitcoins: a pseudo-anonymous electronic currency system. Proceedings of the 2011 Conference of the Center for Advanced Studies on Collaborative Research. IBM Corp., 2011.
7. Koshy, Philip & Koshy, Diana & McDaniel, Patrick. (2014). An Analysis of Anonymity in Bitcoin Using P2P Network Traffic. 8437. 469-485. 10.1007/978-3-662-45472-5_30.
8. Reid, Fergal, and Martin Harrigan. An analysis of anonymity in the bitcoin system. Security and privacy in social networks. Springer, New York, NY, 2013. 197-223.
9. Zero to Monero: First Edition - a technical guide to a private digital currency; for beginners, amateurs, and experts(2018)
   URL : https://www.getmonero.org/library/Zero-to-Monero-1-0-0.pdf
10. Ben-Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., & Virza, M. (2014). Zerocash: Decentralized Anonymous Payments from Bitcoin. 2014 IEEE Symposium on Security and Privacy, 459-474.
11. Seo, J., Park, M., Oh H., Lee K.: Money Laundering in the Bitcoin Network: Perspective of Mixing Services. 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2018, pp. 1403-1405.
12. Fleder, Michael, Michael S. Kester, and Sudeep Pillai. Bitcoin transaction graph analysis. arXiv preprint arXiv:1502.01657 (2015).
13. Ermilov, D., Panovy, M., Yanovich Y.: Automatic Bitcoin Address Clustering. 2017 16th IEEE International Conference on Machine Learning and Applications.
14. Jourdan, Marc, et al. Characterizing entities in the bitcoin blockchain. 2018 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2018.
15. Cai, L., Wang B.: Research on Tracking and Tracing Bitcoin Fund Flows. 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference
16. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. 2008.
    URL : https://bitcoin.org/bitcoin.pdf
17. Andreas M. Antonopoulos. 2014. Mastering Bitcoin: Unlocking Digital Crypto-Currencies (1st. ed.). O'Reilly Media, Inc.
18. Spagnuolo, Michele, Federico Maggi, and Stefano Zanero. Bitiodine: Extracting intelligence from the bitcoin network. International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2014.
19. Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, Murat Kantarcioglu: Blockchain Data Analytics. Journal of IEEE Intelligent Informatics, vol. 20, January 19
20. Cuneyt Gurcan Akcora, Yulia R. Gel, and Murat Kantarcioglu. 2017. Blockchain: A Graph Primer. 1, 1, Article 1 (August 2017). arXiv:1708.08749 [cs.CY]
21. Hong, Younggee, Hyunsoo Kwon, Sangtae Lee, and Junbeom Hur. Poster: De-mixing Bitcoin Mixing Services. (2018).
22. Goodell, G., Aste, T.: Can Cryptocurrencies Preserve Privacy and Comply with Regulations?. SSRN Electronic Journal. 2. 10.2139/ssrn.3293910
23. Balthasar, T., Hernandez-Castro, J.: An Analysis of Bitcoin Laundry Services. NordSec 2017, LNCS 10674, pp. 297–312, 2017.
24. Möser, Malte, Rainer Böhme, and Dominic Breuker. An inquiry into money laundering tools in the Bitcoin ecosystem. 2013 APWG eCrime Researchers Summit. IEEE, 2013.
25. Saurkar, Anand V., Kedar G. Pathare, and Shweta A. Gode. An Overview on Web Scraping Techniques and Tools. International Journal on Future Revolution in Computer Science & Communication Engineering (2018)
26. Balsakas, A., Franqueira, V.: "Analytical Tools for Blockchain: Review, Taxonomy and Open Challenges," *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, Glasgow, 2018, pp. 1-8.
27. S. Jouili and V. Vansteenberghe, "An Empirical Comparison of Graph Databases," *2013 International Conference on Social Computing*, Alexandria, VA, 2013, pp. 708-715.