مجلة
كلية التــراث الجامعة

مجلة علمية محكمة
متعددة التخصصات نصف سنوية
العدد الحادي والأربعون
30 نيسان 2025
ISSN 2074-5621

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

An Analysis of Blockchain Data Analytics

Ahmed Hamid Ahmed

College of Computer Sciences.
University of Technology- Iraq, Baghdad, Iraq

Abstract—Blockchain technologies are responsible for many of the modern world's unique experiences, including cryptocurrencies and managing the world's food supply. Because these systems are exponentially expanding, as is the data stored by the relevant Blockchains that drive them. As such, research into this field is also booming and responsible for several key developments related to Blockchain methodologies. The following paper is an outline of the topics, challenges and current uses of Blockchain Data Analytics, including predicting the values of cryptocurrencies (here: Bitcoin) and detecting crimes committed on the Internet.

Keywords—Blockchain, Bitcoin, Financial Analytics, e-crime, cryptocurrency

I. INTRODUCTION

Blockchain is a phenomenon of the modern age of computing, driven by the rise of cryptocurrencies that require easy transactions and anonymity. In essence, Blockchain is nothing more than a distributed public ledger where transactions between two parties are logged without the need of a central authority. Thus, it is possible for two parties who have not been in direct contact to produce a transaction which cannot be modified and will be permanently recorded on the ledger to be seen by everyone using a Blockchain. It can be easily seen how this breakthrough is relevant to finances, especially the cryptocurrencies that drive much of the modern financial discussion.

Bitcoin is at the forefront of this discussion, both in terms of driving the cryptocurrency boom and in innovating Blockchain to be used for its purpose. With the onset of Bitcoin, Blockchain 1.0 was established [1]. As of the time of this writing, there are over one thousand cryptocurrencies which involve Blockchain to some extent. Because of the massive impact that Blockchain has had on business and finance, it has been compared to the equally revolutionary double-entry accounting that turned the business world on its head when it first appeared [2]. Despite this connection to Bitcoin and the subsequent cryptocurrencies, Blockchain is also used across a wide spectrum of industries and services. Some of these uses include:

- Identity services (Hypr, Bitnation)
- Voting (Social Krona, FollowMyVote)
- Copyright management (Blockphase, LBRY)
- Provenance (Chronicled, Everledger)

Because Blockchain is a relatively new technology and is directly related to an industry which is shifting by the hour, it remains a challenge to accurately predict where and how it will be used in the future. However, it is without a doubt that Blockchain is here to stay and will find its way into numerous applications before long.

There are two types of Blockchains: private and public. The public Blockchains are of the most interest to the most people as Bitcoin (and similar cryptocurrencies) represent this type. Here, any node can join the network without seeking permission, allowing for every transaction to be easily

viewable by every node on the Blockchain network. Conversely, private Blockchains are (as the name implies) privately created for a specific industry, company or organization, which means that only participants with the necessary permissions can read and write data on the network. As public Blockchains are more relevant to the modern world, the focus of the below information will be limited to examining this type.

Although the basic chain structure of Blockchain remains the same throughout each evolving type, some developing Blockchain solutions incorporate innovative data structures. This makes analyzing this technology to learn about trends in the field a worthy endeavor for anyone interested in both modern computing and modern business. However, this analysis is not without its questions and concerns. The following research will be driven by these three questions:

1. How is the data stored on Blockchains modelled and stored?
2. Which types of tools are needed to properly analyze Blockchain data?
3. How can the insights provided by analyzing this data be of assistance to future research?

Based on these three research questions, this paper will attempt a thorough and comprehensive look into Blockchain Data Analytics. Section II will present a brief history of public Blockchains and how the technology was/is tied to Bitcoin. After, Section III will examine common Blockchain data structure models. Finally, Sections IV and V will take a brief look at modern research which displays how Blockchain Data Analytics has been used for several purposes, both licit and illicit, including: cryptocurrency modelling, e-crime detection, illegal economic activities, and human trafficking. This paper will end with our Conclusion (Section VI) that summarizes the ideas presented throughout this writing.

II. BRIEF HISTORY OF BLOCKCHAIN

In 2008, Satoshi Nakamoto first outlined what would become known as “Blockchain” in the white paper titled “Bitcoin: A peer-to-peer electronic cash system” [3]. Although he did not outright name his technology “Blockchain” but rather used the phrase “a chain of blocks,” the technology that runs beneath Bitcoin, and the thousands of similar cryptocurrencies spawned in its wake, have taken upon this name.

Although the subsequent cryptocurrencies are similar to Bitcoin in many respects, they have differed in some key points as well. For example, ZCash focuses more on privacy through its shielded pool which hides transactions and Litecoin attempts to make mining more balanced through its modified algorithm. While the use and value of these cryptocurrencies is still being debated by professionals of all types, it cannot be argued that their widespread introduction to society also paved the way for a more uniform acceptance of Blockchain technology.

The current implementation, Blockchain 2.0, came about in 2014. It has paved the way for Smart Contracts (another term for software code) to be publicly used and saved on Blockchains. Thus, online transactions are able to pass between entities without change, without delay, and in a way that is publicly verifiable. The upcoming Blockchain 3.0 is predicted to force Blockchain technology further into everyday use through IoT integration [4]. Even though Blockchain is an ever-changing technology, its use has reached the levels of, and in some cases even surpassed,

traditional business and transaction practices. This trend is set to continue once Blockchain becomes more commonly applied throughout sectors and industries.

III. BLOCKCHAIN DATA MODELS

As there are two types of Blockchains (public and private), there are also two categorizations for public Blockchains. The first type is known as "unspent transaction output" (UTXO) and is used for services such as Bitcoin and Litecoin. The second type is account-based Blockchains, used in services such as Ethereum, which is a fund-raising platform for tech start-ups. Within both types, data blocks correspond to a finite number of transactions. However, these transactions differ between the two types of public Blockchains. These differences will be outlined below.

A. Unspent Transaction Output Blockchain Data

Unspent transaction output (UTXO) Blockchains came first onto the market. Based on their connection to Bitcoin, they are also the more valuable type based on market capitalization, as Bitcoin alone accounts for between 45 and 60 percent of the cryptocurrency market [5]. Within this type, each data block is responsible for a financial transaction, encoding the transferring of coins between parties. Every transaction uses inputs and produces outputs, here referring to spending coins from the input and transferring coins to the output. Within UXTO, coin supply and block creation are connected. For every transaction, coins are created and gifted to the block miner as a reward.

It should be mentioned that there are three rules responsible for shaping data on these Blockchains, each established by the original designer of Bitcoin, Satoshi Nakamoto [3]. The first is the Source Rule which states that input coins from different transactions may be combined before being spent either on a single transaction or spent separately. The second, or the Mapping Rule, dictates that every coin payment has to display proof of funds through a reference to former outputs. Through this rule, it is possible follow a trail of previous payments. That said, it is often impossible to determine the origin point of a single coin because every transaction has separate lists of the respective inputs and outputs. The third rule, called the Balance Rule, makes it so that the coins received of one transaction must all be spent on a single transaction. All coins not sent to the output location are deemed the transaction fee and are passed to the miner who created the block. It is still possible for the one spending the coin to keep the change and forego this transaction fee if they create a new address and send the leftovers there. Because of these rules outlined by Nakamoto at the start of the Bitcoin adventure, UTXO Blockchains are not networks, but rather forward branching trees.

UTXO Blockchains contain more than the data related to the transaction. Metadata is also stored in these Blockchains, dating back to the original Bitcoin, where Nakamoto left the message "The Times 3 January 2009 Chancellor on brink of second bailout for banks" [3]. In 2014, every Bitcoin transaction began including a field (called "*OP_RETURN*") which stores log information in 80 bytes [6].

B. Account-based Blockchain Data

In contrast to the Balance Rule of the UTXO Blockchain, in account-based Blockchains, addresses are able to spend a portion of its coins and keep the balance. This Blockchain type uses only one input and one output address. Because it is free to create addresses, one address is often used to send and receive coins several times.

The most notable account-based type is Ethereum, which was founded in 2015 [7]. Although this Blockchain also has a currency similar to Bitcoin (Ether), it was designed around storing data and code on the Blockchain as opposed to being a straight financial service. The service uses a proprietary coding language called Solidity to write this code, which is known as a "Smart Contract." The code is then compiled to bytecode before being executed along the Ethereum Virtual Machine. These Smart Contracts involve both code and agreements and are self-executing Turing complete contracts. The unpreventable execution of code can also be publicly verified, like the UTXO types.

Two types of addresses are used for account-based blockchains: externally owned and contract addresses. The first type are controlled by the users, while the second type are controlled by the smart contract code. While most transactions involve a user address (externally-owned contract) uploading code to a contract address, these transactions can also originate from a contract address itself (i.e., a contact address transmitting to another contract address). Uploading these contracts instruct every node in the Blockchain to store the relevant code locally as it is stored in the Blockchain and then duplicated at every node.

Returning to Ethereum, this Blockchain copies Bitcoin in a key way: allowing users to pass messages within the smart contracts using the *input data* field (which contains the function names and parameters). The metadata stored in the code allows for code execution as parameters are fed to the stored function. Because this code execution happens at every node across the world, Ethereum has earned the moniker as being the "World Computer."

Account-based Blockchains involve two types of transactions: transfer and internal. The first type transfers the relevant cryptocurrency between addresses. The second type are the result of smart contracts changing their state related to their address. For this type, there are two ways it can occur, either by parsing the message and updating the relevant state or by executing every contract transaction individually through a full node. They both have their advantages and disadvantages. While the first type is unable to locate transaction failures, the second type takes up both large chunks of time and resources.

IV. METHODOLOGY AND TOOLS OF BLOCKCHAIN DATA ANALYTICS

Two types of graphs have been employed when analyzing Blockchain data (and UTXO data specifically). The two approaches that guide this analysis are known as the transaction graph approach and the address graph approach. Both types utilize a single type of node, but they differ in some key areas, which will be outlined below.

The transaction graph method ignores the addresses and produces edges around each transaction node [8]. In contrast, the address graph approach ignores the transactions, producing edges around the address nodes [9]. There is a limitation to this method due to the aforementioned Mapping Rule of UTXO Blockchains that forces each input of a transaction to be connected to each output address. Massive cliques can form when a multitude of addresses are used for a given transaction. While both of these methods have their uses, neither accurately represents Blockchain data. This is because the loss of data relevant to either addresses or transactions have an effect on accurate predictions [10].

It is possible to losslessly encode network subgraphs using K-chainlets as this allows nodes to be either addresses or chainlets and incorporates the local higher order structures of the Blockchain graph [11]. "Chainlets" are subgraphs that can be used as the foundation of

Blockchain analysis in place of edges or nodes. Through subgraphs, including nodes and edges counts as a single choice, making it possible to use the subgraph as a standalone data unit. In addition, the shape of the subgraphs is unique, related to their specific task within the network. This is shown in Fig. 1 below.

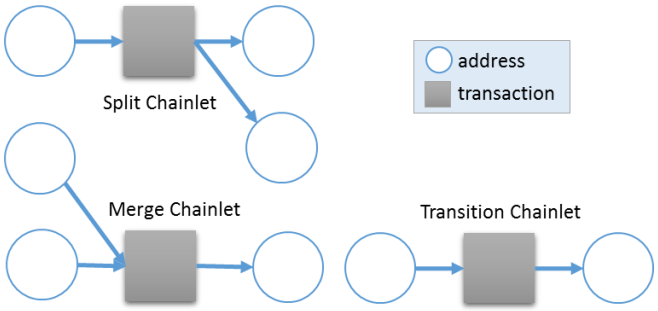

Fig. 1. Chainlets use their input and output to encode transactions. It is possible to combine these chainlets and use them for machine learning [11].

With the boom in popularity of Bitcoin, much work has gone into predicting its price through such means as clustering coefficients, average account balance, counting the new edges [12], then moving on to network temporal behavior and flow. Such studies have proven at Bitcoin is a scale-free network, noticing disassortative behavior in that although active entities are always in flux, there is always some active entities in play [13], with coin exchange sites serving as the most central network nodes [14].

By contrast, account-based blockchains can better make use of the abovementioned graph analysis tools because the transactions are one-to-one. Regardless, analyzing the network involves careful attention to distinguishing internal and ordinary transactions to accurately model the [buy/sell] relationships on the graph [15]. Using Ethereum as an example presents another concern of graphing account-based blockchains: the cryptocurrency involves overlapping layers of token networks that must each be represented with their own graphs that indicate user/contract addresses as nodes. It is possible for networks of Ether tokens to share nodes but not edges, meaning dozens of edges may be shared by only two nodes.

A. Tools

Researchers face issues when data querying Blockchain data, as the fact that data blocks are written onto files (.dat files for Bitcoin and levelDB for Ethereum) causes this process to take a large chunk of time. Although Blockchain query languages do exist, they are not yet commonplace tools. Private tools for this purpose also exist, such as those found on Chainanalysis and Santiment. Only the websites “blockchain.com” and “etherscan.io” have public tools available, though their range of usage remains limited. Two common tools research employ for their analysis are Biva and the BlockSci project.

V. APPLICATIONS OF BLOCKCHAIN DATA ANALYTICS

Regardless of the evolving landscape of Blockchains, cryptocurrencies remain the focal point for analyzing the technology. This is because no matter what other systems develop, Bitcoin and

Ethereum continue to dominate the market space. In addition, through a proper analysis of these online financial tools, researchers both professional and private are seeking to use the coins for economic development. Thus, the two main applications remain price prediction and detecting e-crimes related to these currencies.

A. Price Prediction

As mentioned, the largest question surrounding Bitcoin is how it compares to more traditional economies in terms of fluctuations and the impact it would feel from a recession or crash. Thus, researchers focus heavily on how the price of Bitcoin is affected by transactions and addresses. Accurately predicting the price of this behemoth cryptocurrency revolves around studying the transactions and related graphs, with features such as average transaction amount revealing unsatisfactory predictive measures [10]. Other methods focus instead on how the number of new edges, the clustering coefficient, and the average balance impacts the price of Bitcoin, using Blockchain chainlets as the measure of prediction. This has shown some promising results.

Beyond serving the ability to predict Bitcoin prices, chainlets can also strongly indicate price risk. It has become possible to condition "extreme chainlets" with relative daily loss distribution in order to predict the outliers to some degree of accuracy [16]. These extreme chainlets make it possible to account for extreme Bitcoin losses in the risk models as they represent transactions from either more accounts to less addresses or more addresses to less accounts. Looking at transactions of this type, it is possible to notice movement of Bitcoin to and from other currency types.

B. Detecting E-Crimes

Bitcoin first came into prominence as the primary currency of preference for illicit Internet activities within the Dark Web. Such crimes as drug distribution, human trafficking, blackmail, money laundering [17], ransomware [18], and more have been attributed to individuals using Bitcoin. Due to the anonymity it provides, Bitcoin allowed users and vendors of illegal products and services to exchange payment and avoid detection. Actually, cryptocurrencies are in fact "pseudo-anonymous" currencies because while users are not forced to identify themselves, all of the transactions are publicly available for anyone to see. Criminals then, in the path to true anonymity, have taken measures to distance their online presence with their real-life identity. One such measure is using a service such as Tor to enhance their online privacy. Another way members of the Internet's black market stay hidden is to operate in plain sight. That means, they ensure that all of their transactions appear as normal as possible in terms of amount, frequency, and time. This has not stopped law agencies across the world from taking measures to locate these criminals who hide behind computer screens in order to carry out their activities that have a major impact on everyday life.

Due to the Mapping Rule outlined above, intelligence agencies have trouble following the trail of coins. Heuristic measures have come into the fold, but have provided mixed results in tracking the flow of coins [19]. Thus, researchers have turned their attention to locating clues that will help them locate the ways criminals mix their coins along the Bitcoin Blockchain. With these developments, criminals have shifted their attention to more truly anonymous cryptocurrencies, such as Monero and Zcash.

This difficulty has thwarted neither researchers nor policing agencies from using the tools at their disposal to trace such crimes as money laundering. For example, a measure known as anti-

money laundering (AML) was used on Bitcoin to identify the crime through locating successive transactions where money was transferred [17]. The relevant Blockchain addresses used can be connected in order to identify the suspects of such suspicious financial patterns. Such a pattern often includes repeated transactions that transfer coins from black addresses to online exchanges. This allows the criminals to cash the coins out without authorities potentially confiscating them first. In addition to money laundering, this pattern also helps locate ransomware payments [18].

Researchers have resorted to unsupervised learning because of the lack of wallet addresses that have been identified as malicious, false, or targets of crimes such as ransomware attacks. Such methods include using the ransom addresses that are known to locate associated wallet addresses by connecting the co-spending behavior or locating similar spending patterns. As the field of detecting Internet crimes that are funded by cryptocurrencies is ever-expanding, the techniques are also in constant flux. Other techniques for locating criminal behavior include: adaptive penalization, oversampling, and Bayesian networks.

VI. CONCLUSION

From the earliest days of Bitcoin adoption, Blockchain has become one of the most interesting, important, and constantly changing areas of research focus that connects digital events to those that effect the offline world. Because the systems that drive Blockchains are always evolving and growing, the amount of data stored on this technology is also expanding at a rapid pace. While this paper has done its best to summarize the field as best as possible, further research is required to truly represent the phenomenon. Besides Bitcoin, novel cryptocurrencies appear on and vanish from the marketplace at an ever-expanding rate. This implies that Blockchain technology is here to stay and that it is a facet of the modern world that must have serious research efforts poured into it.

REFERENCES

[1] M. Swan, *Blockchain: Blueprint for a new economy*. O'Reilly Media, Inc., 2015.

[2] P. Vigna and M. J. Casey, *The age of cryptocurrency: how bitcoin and the blockchain are challenging the global economic order*. Macmillan, 2016.

[3] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[4] S. C. Alliance, "Smart contracts: 12 use cases for business & beyond," 2016, http://digitalchamber.org/assets/ smart-contracts-12-use-cases-for-business-and-beyond.pdf.

[5] F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," *IEEE Communications Sur- veys & Tutorials*, vol. 18, no. 3, pp. 2084–2123, 2016.

[6] M. Bartoletti and L. Pompianu, "An analysis of bitcoin op return metadata," in *International Conference on Financial Cryptography and Data Security*. Springer, 2017, pp. 218–230.

[7] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.

[8] M. Fleder, M. S. Kester, and S. Pillai, "Bitcoin transaction graph analysis," *arXiv preprint arXiv:1502.01657*, 2015.

[9] M. Spagnuolo, F. Maggi, and S. Zanero, "Bitiodine: Extracting intelligence from the bitcoin network," in *International Conference on Financial Cryptography and Data Security*. Springer, 2014, pp. 457– 468.

[10] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," *No Data*, 2015.

[11] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu, "Forecasting bitcoin price with graph chainlets," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PaKDD)*. Springer, 2018, pp. 765–776.

[12] M. Sorgente and C. Cibils, "The reaction of a network: Exploring the relationship between the bitcoin network structure and the bitcoin price," *No Data*, 2014.

[13] M. Ober, S. Katzenbeisser, and K. Hamacher, "Structure and anonymity of the bitcoin transaction graph," *Future internet*, vol. 5, no. 2, pp. 237– 250, 2013.

[14] A. Baumann, B. Fabian, and M. Lischke, "Exploring the bitcoin net- work." in *WEBIST (1)*, 2014, pp. 369–374.

[15] W. Chan and A. Olmsted, "Ethereum transaction graph analysis," in *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 2017, pp. 498–500.

[16] C. G. Akcora, M. Dixon, Y. R. Gel, and M. Kantarcioglu, "Bitcoin risk modeling with blockchain graphs," *Economics Letters*, pp. 1–5, 2018.

[17] R. Moser, M.and Bohme and D. Breuker, "An inquiry into money laundering tools in the bitcoin ecosystem," in *eCrime Researchers Summit*. IEEE, 2013, pp. 1–14.

[18] D. Y. Huang, D. McCoy, M. M. Aliapoulios, V. G. Li, L. Invernizzi, E. Bursztein, K. McRoberts, J. Levin, K. Levchenko, and A. C. Snoeren, "Tracking ransomware end-to-end," in *Tracking Ransomware End-to- end*. IEEE, 2018, pp. 1–12.

[19] S. Meiklejohn, M. Pomarole, G. Jordan, D. Levchenko, K.and McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 127–140.