Blockchain Data Analytics

Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, and Murat Kantarcioglu

***Abstract*—Many novel applications, ranging from cryptocurrencies to food supply chain management, drive consumer and industrial adoption of Blockchain technologies. As these applications proliferate, so does the complexity and volume of data stored by Blockchains. Analyzing this data has emerged as an important research topic, already leading to methodological advancements in the information sciences. In this invited paper, we provide a brief overview of Blockchain Data Analytics, focusing both on the emerging research challenges and on the novel applications – from Bitcoin price prediction to e-crime detection.**

***Index Terms*—Blockchain, Bitcoin, Ethereum, Financial Analytics, Anomaly Detection, Time Series Analysis, Blockchain Data Analytics.**

I. INTRODUCTION

THIS decade has been marked with the rise of Blockchain based technologies. At its core, Blockchain is a distributed public ledger that stores transactions between two parties without requiring a trusted central authority. On a Blockchain, two unacquainted parties can create an unmodifiable transaction that is permanently recorded on the ledger to be seen by the public. As legendary venture capitalist Marc Andreessen states “the consequences of this breakthrough are hard to overstate” [1].

The first application of Blockchain has been the Bitcoin [2] cryptocurrency. Bitcoin’s success has ushered an age known as the Blockchain 1.0 [3]. Currently there are more than 1000 Blockchain based cryptocurrencies, known as **alt-coins**. These developments have ignited public interest in Blockchain technology. Some observers compare the inception of Blockchain to the invention of double entry accounting that revolutionized the business world [4]. The emerging Blockchain based applications include voting (FollowMyVote, Social Krona), identity services (Bitnation, Hypr), provenance (Everledger, Chronicled) and copyright management (LBRY, Blockphase). Although it is hard to predict the future impact of Blockchain, it is safe to say that it will enable many important and diverse applications.

Private blockchains are created by industry/organizations and only allow write and read access to the participants with necessary permissions. In contrast, public blockchains, such as Bitcoin, allow any node to join the network without permission and all transactions can be observed by all the nodes that are part of the Blockchain network. In this article we restrict our attention to public blockchains, where data is publicly accessible.

Each blockchain solution utilizes a chain structure, but may also employ novel data structures. Clearly, this information may be analyzed to provide novel insights about emerging trends. This raises questions such as: *1) How to represent and model the data stored on blockchains 2) What are the novel analytical tools needed for analyzing Blockchain data? 3) What insights could be gleaned from the transactions stored on public blockchains?*

We address the above questions by offering a short introduction to Blockchain analytics. We first provide a brief history of public blockchains. Then we discuss the common Blockchain data structure models and provide some insights into several important analytical methods and tools. Finally, we briefly discuss recent studies that use Blockchain data analytics for cryptocurrency modeling, detection of e-crime, human trafficking and illicit economic activity.

II. HISTORY

Blockchain was devised and outlined by Satoshi Nakamoto in his “Bitcoin electronic cash system” [2] white paper in 2008.

Although Nakamoto mentioned “a chain of blocks” only, the term Blockchain has become the name of the technology underlying Bitcoin. The success of Bitcoin led to the creation of hundreds of similar cash systems, which came to be known as cryptocurrencies. These off-shoots differ from Bitcoin in a few aspects. For example, Litecoin modified the block mining algorithm for fairness in mining, and ZCash introduced a shielded pool to hide transactions for better privacy.

Although their success is still a hotly debated topic, cryptocurrencies paved the way for broad Blockchain adoption. Since 2014, Blockchain 2.0 led to the creation of Blockchain platforms where software code, called Smart Contracts, can be stored and executed on a Blockchain publicly. These contracts allow unstoppable, unmodifiable and publicly verifiable code execution as transactions between online entities. Blockchain 3.0 is expected to further immerse the technology into our daily lives with IoT integration [5].

Blockchain continues to evolve, but its applications have already matured to rival, and already in some cases, replace more traditional institutions as avenues of global activity. For example, the Ethereum blockchain has become a major fundraising medium for tech start-ups; initial coin offerings (ICOs) of Ethereum tokens have reached 45% of second quarter IPOs [6] in the US.

III. BLOCKCHAIN DATA MODELS

Public blockchains can be broadly categorized as unspent transaction output (UTXO) based (e.g., Bitcoin, Litecoin) and account based (e.g., Ethereum) blockchains. In both types

The authors are from University of Texas at Dallas Computer Science and Mathematical Sciences departments and Illinois Institute of Technology Applied Mathematics department. Corresponding author e-mail: (cuneyt.akcora@utdallas.edu).

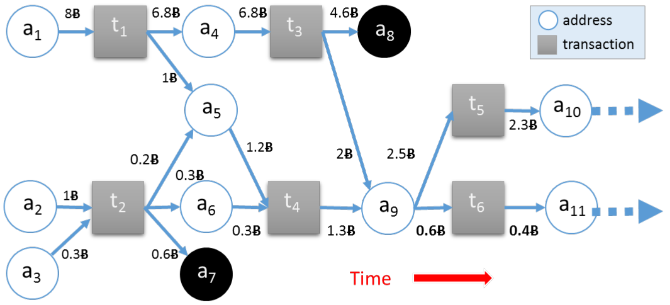

Fig. 1. A network of 11 addresses and 6 Bitcoin transactions. Block boundaries are not shown. Coins at addresses a_7 and a_8 remain unspent. The difference between input and output amounts (e.g., 0.2B at t_1) are collected as the transaction fee. Most crypto-currencies have the same data model as Bitcoin.

of blockchains, a data block consists of a finite number of transactions, but the transactions have differing characteristics. Below, we briefly discuss these two different type of blockchain transaction data.

A. The Unspent Transaction Output Based Blockchain Data

The unspent transaction output (UTXO) based blockchains are the earliest and most valuable (in terms of market capitalization) blockchains: Bitcoin alone constitutes 45-60% of the total cryptocurrency [7] market capitalization. In UTXO blockchains each data block contains a (financial) transaction that encodes a transfer of coins between multiple parties. Each transaction consumes (i.e., spends coins from) some inputs and creates (i.e., directs coins to) new outputs. Fig. 1 shows an example UTXO network, where transaction t_1 encodes a transfer of bitcoins from the address a_1 to a_4 and a_5. In UTXO, coin supply is tied to block creation; a certain amount of coins are created and given to the block miner as the block reward. Bitcoin started with 50B per coin and halves the block reward every 4 years. This geometric series will result in a total of 21 million bitcoins.

We emphasize **three rules** that shape data on UTXO blockchains. These rules are due to the design choices by Satoshi Nakamoto in Bitcoin [2].

Source Rule: Input coins from multiple transactions can be merged and spent in a single transaction (e.g., the address a_5 receives coins from t_1 and t_2 to spent in t_4 in Fig. 1), or spent separately (e.g., in Figure 1, a_9 spends coins received from t_3 and t_4 in t_5 and t_6).

Mapping Rule: Each coin payment must show proof of funds by referencing a set of previous outputs. Although this allows anyone to trace back a history of payments, it is not always possible to locate where a specific coin originates from. This is because each transaction lists a set of inputs and outputs, separately. For example, t_2 has two inputs and three outputs, but an explicit mapping between inputs and outputs does not exist. Coins flowing to a_5 may have come from either a_2 and a_3, or both. As a result, a transaction can be considered a lake with in-flowing rivers, and out-flowing rivers (i.e., emissaries).

Balance Rule: Coins received from one transaction must all be spent in a single transaction. Any amount that is not sent to an output address is considered to be the transaction fee, and gets collected by the miner who creates the block. In order to keep the change, the coin spender can create a new address (i.e., change address) and send the remaining balance to this new address. Another option is to use the spender's address as one of the output addresses, and re-direct the balance. As a community practice, this reuse of the spender's address (i.e., **address reuse**) is discouraged. As a result, most nodes appear in the graph two times only; once when they receive coins and once when they spend it. The change address, if created, becomes the new address of the coin owner.

Due to these rules, the unspent transaction output based blockchains should be considered as forward branching trees, rather than networks.

UTXO blockchains also contain non-transactional data. In the first Bitcoin block, Nakamoto had left the text message "The Times 3 January 2009 Chancellor on brink of second bailout for banks". Adding metadata to Bitcoin transactions have been a topic of discussion and since 2014, each Bitcoin transaction contains a field (*OP_RETURN*) that is designed to store log information in 80 bytes [8].

Improving on the metadata functionality, The Namecoin blockchain has been created in 2011 to store key-value pairs for a decentralized namespace. Namecoin data blocks store registrations or updates for the .bit domain names, which are independent of the ICANN. A domain expires 35,999 blocks (200-250 days) after it is registered as a key-value pair in the Blockchain. Besides the domain registration (i.e., /d), Namecoin has a public online identity namespace (i.e., /id), along with other proposed services.

Storing the full Blockchain to reach .bit domain addresses in real time has discouraged Namecoin adoption. Although online explorer sites and browser extensions have been created to help Internet users with .bit domains, Namecoin namespaces have historically remained underutilized, and most blocks are empty of any key-value pair [9].

B. Account Based Blockchain Data

In account based blockchains, an address can spend a fraction of its coins and keep the remaining balance. In these blockchains, a transaction has exactly one input and one output address. Although address creation is free, mostly a single address is used to receive and send coins multiple times.

Created in 2015, Ethereum [10] is currently the most valuable account based blockchain. Similar to Bitcoin, Ethereum has a currency: Ether. However, the Ethereum project's main goal is to store data and software code on a Blockchain. The code (a smart contract) is written in the proprietary coding language Solidity, which is compiled to bytecode and executed on the Ethereum Virtual Machine. Smart Contracts are self-executing Turing complete contracts which contain code and agreements. An analogy is the MYSQL snippets stored on a database. However, Smart Contracts also ensure unstoppable, deterministic code execution that can be verified publicly.

Account based blockchains use two types of addresses; externally owned addresses (governed by users) and contract addresses (governed by smart contract code). A transaction to upload the Smart Contract code to a contract address is usually initiated by an externally owned address (i.e., user address), but it can also be initiated by a contract address. The code at the address is stored in the Blockchain and replicated at all Blockchain nodes. In other words, uploading the contract forces other nodes to store the code locally.

Similar to the log field in UTXO blockchains, each Ethereum transaction contains an *input data* field which is used to pass messages (i.e., function names and parameters) to smart contracts. The code is executed by feeding parameters to the stored function. This execution occurs at all nodes, worldwide. For this reason, Ethereum is called the World Computer.

Contract creation is expensive, but born by the contract creator. Subsequently, other users or contracts can create transactions directed to the contract address to call the functions contained in the contract. Costs of operations (such as multiplication = 5 and addition = 3) executed by the contract are summed up in terms of the execution fee called “gas”, and billed to the address that created the transaction, in ethers. The currency ether acts as the digital oil of Ethereum World Computer.

Smart Contracts gave rise to Smart Contract based tokens: exchanged data units. Holding tokens allow users to get serviced by a company in real life. For example, the Storj token stores files on your hard disk, and pays you a fee through Ethereum. Furthermore, tokens can be bought or sold online and act as value stores. In this worldwide market, tokens that are valuable are arbitrated in fiat currencies. These prices can be viewed on online exchanges such as coinmarketcap.com

Account based blockchains have two types of transactions. The first transaction type involves a transfer of the used cryptocurrency, such as Ether on Ethereum, between two addresses. This can be modeled with a directed edge between the two addresses.

The second type, internal transactions, are created when smart contracts change states associated with addresses. In the most basic scenario, consider a sell order issued by address a_1 to a Smart Contract where the *to* parameter is a_2 and the *value* parameter is 2 token. The Smart Contract creates an internal transaction that transfers 2 tokens from a_1 to a_2. Internal transactions can be discovered in two ways: by parsing the transaction's message and updating states associated with a_1 and a_2 manually, or by running the transaction message through the smart contract code and observing the states and logs created during execution. The second option requires running a full Ethereum node and executing every contract transaction, which is costly in terms of time and resources. The parsing option is easier as it does not require code execution. However, the parsing method cannot discover transaction failures (due to reasons such as insufficient gas), and create internal transactions that do not actually exist.

IV. Blockchain Data Analytics Methods and Tools

Largely deriving from existing network analysis methodology, early research works analyze UTXO data by creating a graph that employs a single type of node only. These are transaction and address graph approaches.

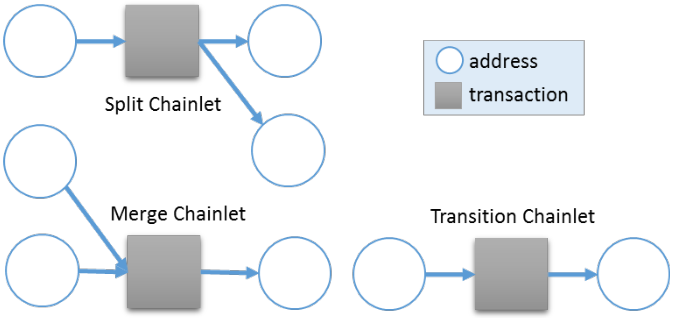

Fig. 2. Chainlets encode a transaction with its inputs and outputs. Chainlets can be aggregated and their occurrence information can then be used in machine learning tasks.

In the transaction graph approach, addresses are ignored and edges are created among transaction nodes [11], [12]. Naturally, the transaction graph is acyclic and a transaction node cannot have new edges in the future.

In the address graph approach [13], transactions are ignored and edges are created among address nodes. However, because of the Mapping Rule (see Sec. III), inputs of a transactions must be connected to all output addresses of a transaction, which may create large cliques if too many addresses are involved in a transaction.

Single node type approaches do not provide a faithful representation of the Blockchain data (See [14] for more on Blockchain data models). The loss of information about addresses or transactions seem to have an impact in predictive models [15].

K-chainlets [16] offer a lossless way to encode network subgraphs where nodes can be addresses or chainlets. The model utilizes local higher order structures of the Blockchain graph. Rather than using individual edges or nodes, subgraphs can be used as the building block in Blockchain analysis. The term *chainlet* refers to such subgraphs.

This choice is due to two reasons. First, the subgraph can be taken as a single data unit because inclusion of nodes and edges in it is based on a single decision. As a transaction is immutable, joint inclusion of input/output nodes in its subgraph cannot be changed afterwards. This is unlike the case on a social network where nodes can become closer on the graph because of actions of their neighbors. Second, as shown in Fig. 2, subgraphs have distinct shapes that reflect their role in the network, and these roles can be aggregated to analyze network dynamics.

As Bitcoin became popular, a number of studies aimed at using various network characteristics for price predictions. For instance, [17], [15] employ such network features as mean account balance, number of new edges and clustering coefficients. In turn, network flows and temporal behavior of the network have been used as alternative price predictors by [18] and [19], respectively.

Studies in network features show that since 2010 the Bitcoin network can be considered a scale-free network [20]. In- and out-degree distributions of the transactions graphs are highly heterogeneous and exhibit a disassortative behavior [21]. Active entities on the network change frequently, but there are consistently active entities [22]. The most central nodes in the network are coin exchange sites [23].

As all transactions are one-to-one, account based blockchains enable the usage of traditional graph analysis tools easily [24], [25]. However care must be taken to extract internal transactions from ordinary transactions, so that all relationships (i.e., token buy/sell) between addresses can be modeled on the graph.

As a second issue, the complete Ethereum graph have overlapping layers of token networks; each token can be represented with a separate graph on the Ethereum network where nodes are user/contract addresses. A token network is a directed, weighted multi-graph. Two token networks may share nodes but not edges. The complete Ethereum graph consists of layers of token networks. Multiple edges can exist between two nodes of the Ethereum graph, and each edge can transfer a different token. On the Ethereum blockchain, it is not rare to see hundreds of edges between two nodes.

A. *Tools*

A criticism of Blockchain is that data blocks are written into files (e.g., as levelDB files in Ethereum and .dat files in Bitcoin) on disk, which makes data querying time consuming. Recent years have witnessed development of Blockchain query languages [26] and analytics frameworks [27] but their adoption is still limited. Companies such as Santiment.com and Chainalysis.com have developed in-house data querying and analysis tools, but these are not yet open to public. Online explorers such as blockchain.com and etherscan.io provide limited analytics tools to the public.

A widely used tool in Bitcoin data analytics is the BlockSci project [28]. A similar tool is the Bitcoin Network Visual Analytics tool Biva.[1]

Besides transaction data that involve financial relations between addresses, the advent of Ethereum 2.0, which brought software code to blockchains, has propelled smart contract analysis [29] as an important data analytics direction. However, most research approaches in this direction are based on static code analysis for tasks such as contract classification [30], and do not analyze the decisions made by the studied smart contracts.

V. Applications of Blockchain Data Analytics

Since the seminal Bitcoin paper [2] in 2008, cryptocurrencies [7] have been the most prominent Blockchain application. Recently there has been an interest in analyzing Blockchain platform (e.g., Ethereum [25]) data but Bitcoin and a few other alt-coins have been the main focus of Blockchain Data Analytics. Broadly, studies address the capacity and limitations for coins to provide a robust and transparent economic system for all economic participants.

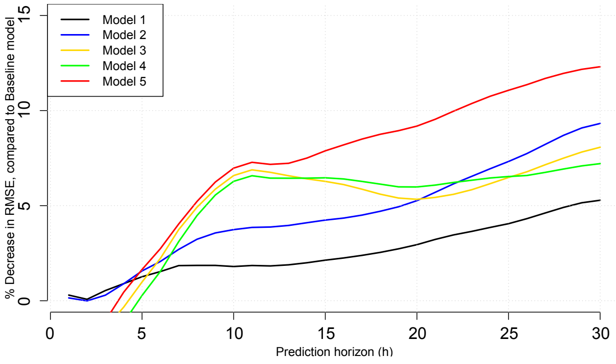

Fig. 3. Daily Bitcoin price prediction in 2016. Model performances for various chainlet models are shown. For three or more steps ahead forecasts, chainlets play an increasingly significant predictive role in Bitcoin price formation, even when other more conventional factors, such as historical price and number of transactions, are accounted for in the model.

a) Price Prediction: One central question is how Bitcoin fares as a financial asset class - in particular whether the transaction graph is linked to price formation and impact liquidity, or even a market crash. Analyzing the relationship between transactions and addresses and Bitcoin price, has therefore become an important analytics research direction [31]. In particular, there is a growing focus on building statistical models which can predict and attribute price movements to transactions and transaction graph properties. While simple Blockchain transactional features, such as average transaction amount, are shown to exhibit mixed performance for cryptocurrency price forecasting [15], a number of recent studies show the utility of global graph features to predict the price [32], [33], [19], [15]. For instance, [17] analyzed the predictive effects of average balance, clustering coefficient, and number of new edges on the Bitcoin price and [16] use Blockchain chainlets as predictors. Two network flow measures were recently proposed by [18] to quantify the dynamics of the Bitcoin transaction network and to assess the relationship between flow complexity and Bitcoin market variables.

The extent to which we can build predictive models from the chainlets has already led to some promising results [16]. In particular, we have been able to identify certain types and groups of chainlets that exhibit predictive influence on Bitcoin price and volatility. Fig. 3 shows the percent decrease in root mean squared error (RMSE) for some of these models relative to a simple baseline model, which uses Bitcoin prices and transaction volumes of previous days only [16]. Specifically, we evaluate $\psi_{M_i}(h)/\psi_{M_0}(h)$, $i = 1, \dots, 5$, for $h = 1, \dots, 30$ days ahead, where $\psi_{M_i}(h)$ and $\psi_{M_0}(h)$ are the RMSEs for the chainlet predictive model M_i and the baseline model M_0, respectively. We find that Model 5, which uses five different chainlets, yields the most competitive predictive performance.

In addition to price prediction, chainlets are a lead indicator for price risk - the relative daily loss distribution conditioned on the 'extreme chainlets' leads to more accurate outlier prediction [32]. Without extreme chainlets, risk models underestimate the extreme Bitcoin losses. These extreme chainlets represent transactions from a large number of accounts to fewer addresses or vice versa. Such transactions represent

[1] https://github.com/feog/Biva

systemic movements of coins to and from exchanges and other funds.

Which Blockchain representation? The core idea behind the aforementioned predictive analytics approaches is to extract certain global network features made available through Blockchain and employ them for predictions, with utility in financial markets. The best approach to represent the network is application driven and an open area of research, ranging from approaches which offer a defensible economic interpretation to purely attractive on theoretical grounds. However, there is mounting empirical evidence that Blockchain data augments conventional predictive studies (i.e., combining other data sources with Blockchain data is necessary).

b) Criminal Usage Detection: Since its early days, Bitcoin has been used in dark markets, such as SilkRoad.com to connect illegal vendors with buyers. By design, cryptocurrencies are pseudo-anonymous because users do not need identify themselves to enter the network, but all of their transactions on the Blockchain are public. Knowing this aspect, criminals devise schemes to separate their real life and online identities. Such schemes include connecting to the Blockchain network through privacy-enhancing distributed platforms such as Tor [34]. Furthermore, criminals aim at making their actions indistinguishable from the actions of ordinary users on the Blockchain. This involves creating transactions that look *normal* in terms of frequency, time and amount, with varying success.

Beyond online trade, cryptocurrencies are used in payments for human trafficking [35], ransomware [36], personal blackmails [37] and money laundering [38], among many others. Blockchain Data Analytics tools and algorithms can be used by law agencies [39] to detect and analyze such criminal activities.

Securities governance concerns have arisen around the stability of digital coins as a currency, its susceptibility to price manipulation and illegal usage for money laundering and blackmailing [38], [40], [41], [22], [42]. A key question is the extent to which Blockchain delivers the anonymity which financial criminals seek. The lack of explicit mapping (see the Mapping rule in Sec. III) in UTXO based cryptocurrencies, such as Litecoin and Bitcoin, hinders tracking flow of coins among addresses in time. Although some heuristics [43] have been used to track coins, it is possible to use a series of mixing [44] transactions to hide coin flows. Researchers have found empirical clues for this mixing behavior in the Bitcoin blockchain [45], [13]. Later crypto-currencies such as Zcash [46] and Monero [47] improve the mixing capability by introducing additional measures such as shielded pools. These cryptocurrencies give strong guarantees for anonymity.

The anonymity question can be addressed by first understanding the patterns of criminal usage. Moser et al. [38] analyzed the opportunities and limitations of anti-money laundering (AML) on Bitcoin by identifying how successive transactions are used to transfer money. Blockchain addresses can then be linked to identify suspects behind suspicious transaction patterns in cryptocurrencies [13]. The pattern is usually defined as a repeating transaction involving the movement of digital coins from a black (i.e., affiliated with criminal gains) address to an online exchange, where the coins can be cashed out without being confiscated by authorities. The black address that starts the transaction chain may be related to money laundering [38] and ransomware payment [36]. There is growing evidence on the existence of these illicit activities in Blockchain networks and the reader is referred to [48], [49].

In general, the relative scarcity of wallet addresses labelled as either malicious, fraudulent or the target of ransomware, motivates the application of unsupervised learning. For example, a few known ransom addresses can be used to discover other associated wallet addresses by observing the 'co-spending' behavior [43]. Other techniques such as oversampling, adaptive penalization and Bayesian networks have been used to address the class imbalance problem in detection of Bitcoin Ponzi schemes [50].

VI. Conclusion

Blockchain technology has recently witnessed a spark of consumer and industrial interest in a broad range of applications, from digital finance to food safety to health care to weapon tracking. As increasingly more new Blockchain applications appear everyday, the complexity and volume of the data stored by Blockchain also rapidly expand – thereby, constituting a new standalone research direction of *Blockchain Data Analytics*.[2] In this invited paper, we provide a brief overview of the current state of Blockchain Data Analytics, focusing both on methodological advancement and the emerging research challenges, as well as offering insight into some of the most important financial applications.

Acknowledgment

The work of Gel is partially supported by NSF IIS 1633331, NSF DMS 1736368 and NSF ECCS 1824710.

The work of Kantarcioglu was supported in part by NIH award 1R01HG006844, NSF awards CICI-1547324, IIS-1633331, CNS-1837627, OAC-1828467 and ARO award W911NF-17-1-0356.

References

[1] A. Marc, "Why bitcoins matters: https://dealbook.nytimes.com/2014/01/21/why-bitcoin-matters/," *New York Times*, vol. 21, 2014.

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

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

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

[5] 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.

[6] C. Long, "Icos were 45% of ipos in q2 2018, as cryptos disrupt investment banks," www.forbes.com/sites/caitlinlong/2018/07/22/icos-were-45-of-ipos-in-q2-2018-as-cryptos-disrupt-investment-banks.

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

[8] 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.

[2] BlockchainTutorial.Github.io

[9] H. A. Kalodner, M. Carlsten, P. Ellenbogen, J. Bonneau, and A. Narayanan, "An empirical study of namecoin and lessons for decentralized namespace design." in *WEIS*. Citeseer, 2015.

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

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

[12] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *International Conference on Financial Cryptography and Data Security*. Springer, 2013, pp. 6–24.

[13] 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.

[14] C. G. Akcora, Y. R. Gel, and M. Kantarcioglu, "Blockchain: A graph primer," *arXiv preprint arXiv:1708.08749*, pp. 1–17, 2017.

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

[16] 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.

[17] 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.

[18] S. Y. Yang and J. Kim, "Bitcoin market return and volatility forecasting using transaction network flow properties," in *IEEE SSCI*, 2015, pp. 1778–1785.

[19] D. Kondor, I. Csabai, J. Szüle, and G. Pósfai, M.and Vattay, "Inferring the interplay between network structure and market effects in bitcoin," *New J. of Phys.*, vol. 16, no. 12, p. 125003, 2014.

[20] M. Lischke and B. Fabian, "Analyzing the bitcoin network: The first four years," *Future Internet*, vol. 8, no. 1, p. 7, 2016.

[21] D. Kondor, M. Pósfai, I. Csabai, and G. Vattay, "Do the rich get richer? an empirical analysis of the bitcoin transaction network," *PloS one*, vol. 9, no. 2, p. e86197, 2014.

[22] 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.

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

[24] T. Chen, Y. Zhu, Z. Li, J. Chen, X. Li, X. Luo, X. Lin, and X. Zhange, "Understanding ethereum via graph analysis," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1484–1492.

[25] 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.

[26] S. Bragagnolo, H. Rocha, M. Denker, and S. Ducasse, "Ethereum query language," in *Proceedings of the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain*. ACM, 2018, pp. 1–8.

[27] M. Bartoletti, S. Lande, L. Pompianu, and A. Bracciali, "A general framework for blockchain analytics," in *Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*. ACM, 2017, p. 7.

[28] H. Kalodner, S. Goldfeder, A. Chator, M. Möser, and A. Narayanan, "Blocksci: Design and applications of a blockchain analysis platform," *arXiv preprint arXiv:1709.02489*, 2017.

[29] S. Ducasse, H. Rocha, S. Bragagnolo, M. Denker, and C. Francomme, "Smartanvil: Open-source tool suite for smart contract analysis," 2019.

[30] M. Bartoletti and L. Pompianu, "An empirical analysis of smart contracts: platforms, applications, and design patterns," in *International Conference on Financial Cryptography and Data Security*. Springer, 2017, pp. 494–509.

[31] P. Tasca, A. Hayes, and S. Liu, "The evolution of the bitcoin economy: extracting and analyzing the network of payment relationships," *The Journal of Risk Finance*, vol. 19, no. 2, pp. 94–126, 2018.

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

[33] S. Madan, I.and Saluja and A. Zhao, "Automated bitcoin trading via machine learning algorithms," 2015.

[34] P. Syverson, R. Dingledine, and N. Mathewson, "Tor: The secondgeneration onion router," in *Usenix Security*, 2004.

[35] R. S. Portnoff, D. Y. Huang, P. Doerfler, S. Afroz, and D. McCoy, "Backpage and bitcoin: Uncovering human traffickers," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 1595–1604.

[36] 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.

[37] A. D. S. Phetsouvanh, F. Oggier, "Egret: Extortion graph exploration techniques in the bitcoin network," in *IEEE ICDM Workshop on Data Mining in Networks (DaMNet)*, 2018.

[38] 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.

[39] E. U. A. for Law Enforcement Cooperation, "Internet organised crime threat assessment (iocta): https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2017," pp. 1–80, 2017.

[40] G. Di Battista, V. Di Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia, "Bitconeview: visualization of flows in the bitcoin transaction graph," in *IEEE VizSec*, 2015, pp. 1–8.

[41] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in bitcoin," in *IFCA*. Springer, 2013, pp. 34–51.

[42] J. M. Griffin and A. Shams, "Is bitcoin really un-tethered?" 2018.

[43] 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.

[44] T. Ruffing, P. Moreno-Sanchez, and A. Kate, "Coinshuffle: Practical decentralized coin mixing for bitcoin," in *European Symposium on Research in Computer Security*. Springer, 2014, pp. 345–364.

[45] D. McGinn, D.and Birch, D. Akroyd, M. Molina-Solana, Y. Guo, and W. J. Knottenbelt, "Visualizing dynamic bitcoin transaction patterns," *Big data*, vol. 4, no. 2, pp. 109–119, 2016.

[46] G. Kappos, H. Yousaf, M. Maller, and S. Meiklejohn, "An empirical analysis of anonymity in zcash," *arXiv preprint arXiv:1805.03180*, 2018.

[47] A. Kumar, C. Fischer, S. Tople, and P. Saxena, "A traceability analysis of moneros blockchain," in *European Symposium on Research in Computer Security*. Springer, 2017, pp. 153–173.

[48] A. Bogner, "Seeing is understanding: anomaly detection in blockchains with visualized features," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 2017, pp. 5–8.

[49] D. D. F. Maesa, A. Marino, and L. Ricci, "Detecting artificial behaviours in the bitcoin users graph," *Online Social Networks and Media*, vol. 3, pp. 63–74, 2017.

[50] M. Bartoletti, B. Pes, and S. Serusi, "Data mining for detecting bitcoin ponzi schemes," *arXiv preprint arXiv:1803.00646*, 2018.