IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2021, accepted February 24, 2021, date of publication February 26, 2021, date of current version March 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062652

Knowledge Discovery in Cryptocurrency Transactions: A Survey

XIAO FAN LIU[1], (Member, IEEE), XIN-JIAN JIANG[2], SI-HAO LIU[2], AND CHI KONG TSE[3], (Fellow, IEEE)

[1]Department of Media and Communication, City University of Hong Kong, Hong Kong, SAR, China
[2]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[3]Department of Electrical Engineering, City University of Hong Kong, Hong Kong, SAR, China

Corresponding author: Xiao Fan Liu (xf.liu@cityu.edu.hk)

This work was supported by City University of Hong Kong Strategic Research Grant 7005405.

ABSTRACT Cryptocurrencies gain trust in users by publicly disclosing the full creation and transaction history. In return, the transaction history faithfully records the whole spectrum of cryptocurrency user behaviors. This article analyzes and summarizes the existing research on knowledge discovery in the cryptocurrency transactions using data mining techniques. Specifically, we classify the existing research into three aspects, i.e., transaction tracings and blockchain address linking, the analyses of collective user behaviors, and the study of individual user behaviors. For each aspect, we present the problems, summarize the methodologies, and discuss major findings in the literature. Furthermore, an enumeration of transaction data parsing and visualization tools and services is also provided. Finally, we outline several gaps and trends for future investigation in this research area.

INDEX TERMS Bitcoin, complex network, cryptocurrency, data mining, Ethereum, transaction analysis.

I. INTRODUCTION

As of 2020, more than 7000 cryptocurrencies are actively trading in more than 20 000 online exchanges. Their total market capitalization has exceeded USD 300 billion [1]. Although these cryptocurrencies are not backed by any tangible assets, they gain trust from users by publicly disclosing the full creation and transaction history in peer-to-peer blockchain networks.

Each transaction in the blockchain consists of transferring a virtual value from a virtual identity, i.e., a blockchain address or a set of addresses, to another. The sizes of transaction records are quickly expanding. The total transaction volumes of Bitcoin and Ethereum (the top two cryptocurrencies by market capitalization) have exceeded 500 million [2] and 600 million [3], respectively, at the end of 2020. Although technically challenging in extracting, transforming, and analyzing, these transaction histories have given us an unprecedented opportunity to study the panorama of human behavior in a complex economic environment.

Reid and Harrigan [4] conducted the first study on the entire cryptocurrency transaction history (up to mid-2011), revealed emerging structure from the Bitcoin flow network, and demonstrated the transaction history's forensic capabilities. Since then, the data mining from cryptocurrency transactions has grown into a large body of research and been successfully applied in assisting multiple law enforcement actions, including ceasing the then-largest darknet market Silk Road in 2013 [5] and arresting suspects in a major theft from the then-largest cryptocurrency exchange Mt. Gox in 2017 [6]. To date, cryptocurrency transaction analysis, also called blockchain analysis, has become an essential means in fighting drug trafficking, computer network hacking, money laundering, and terrorism financing, as well as studying many other social-economical scenarios [7]–[9].

This paper will survey, categorize, and summarize the existing research in cryptocurrency transaction data mining. Specifically, we identify three distinct research directions, each with a series of research questions.

1) *Traceability and linkability issues.* This line of work addresses three research questions: (1) whether chains of transactions can be traced back and whether blockchain addresses can be associated, i.e., linked, to the same identity; (2) how to counter the traceable and link-able nature of cryptocurrency transactions and hide the traces and associations; and (3) how to resolve the counter-tracing measures.

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

2) *Collective user behaviors.* This line of work addresses the emerging user behavior in the cryptocurrency realm from a macroscopic perspective. Specific research questions include (1) the structure and dynamics of the complex networks formed by the transactions and the user behaviors reflected in the network evolution and (2) other collective transaction patterns, including the wealth accumulation of blockchain addresses, creation and usage of blockchain-issued smart contracts, and blockchain transaction fees.
3) *Individual user behaviors.* This line of work focuses on single users or particular types of users. Specific research questions include (1) descriptive analysis of the behaviors of a particular type of address and (2) using features engineered from transaction patterns in downstream machine learning tasks, e.g., classifying address holder identifies and detecting malicious addresses with signature behaviors.

We also provide a summary of transaction analytical and visualizing tools. Specifically, we categorize the tools into the extract, transform, and load (ETL) tools, visualization tools, and online intelligence platforms that provide environments for or results of real-time transaction data analysis.

Existing articles published before September 2020 are downloaded from the Web of Science database using search terms ''cryptocurrency'', ''cryptocurrencies'', ''transaction'', ''Bitcoin'', ''Ethereum'', and several notable altcoins and payment protocols such as ''Monero'', ''Zcash'', and ''Lightning Network''. Search results appeared across primary computer science and engineering [10]–[12], physics [13], and economic venues [14]. We further filter the articles by only retaining those focusing on analyzing cryptocurrency transaction records, and the most recent ones if authored by the same researchers and appeared in multiple venues. We have also included articles that are not in the selection but are commonly cited for complementation. It is worth noting that apart from the transaction data, other sources of blockchain-related data are also of interest in existing studies. For example, traffic analysis of the blockchain peer-to-peer networks is instrumental in understanding the systems' communication overhead and revealing user identities [15]. Modeling and analysis of the blockchain peer-to-peer networks also can help understand the system dynamics and user behaviors [16]. The analyses of cryptocurrency prices in the context of market efficiency are also commonly seen [17].

The rest of the survey is organized as follows. Section II provides basic concepts of the cryptocurrency economy and blockchain data models. Section III addresses the traceability and linkability nature in Bitcoin and altcoins and summarizes the counter-tracing measures as well as their resolutions. Section IV addresses the collective patterns in cryptocurrency user behaviors, emphasizing the emerging structural patterns in the transaction networks. Section V categorizes the transaction features into several classes, introduces the descriptive transaction pattern analysis of individual addresses, and summarizes the machine learning tasks such as address identity inference, cryptocurrency market price prediction, and anomaly detection using the transaction features. Section VI summarizes the extract, transform, and load (ETL) and visualization tools for cryptocurrency transactions as well as major online intelligence platforms. Section VII discusses several open problems in the field and Section VIII concludes.

This survey contributes to the literature by providing a complete spectrum of knowledge discovery from cryptocurrency transactions and also serves as a handbook for researchers and practitioners interested in harnessing the concurrency transaction data in their research. Nonetheless, we also recognize existing reviews that address individual subtopics covered in our survey, as follows. Technical introductions to current cryptocurrencies [18] and blockchain [19], [20] designs cover issues such as data models, consensus protocols of distributed ledgers, and system throughputs. Some works specifically address the anonymity, privacy, and security issues of Bitcoin and Bitcoin-like cryptocurrencies [21]–[23]. De-anonymization techniques using transaction records, mixing services, and designs of altcoins were surveyed and discussed. A survey of tools for smart contract code analysis can be found in [24].

II. PRELIMINARIES

A. THE CRYPTOCURRENCY ECONOMY

Early cryptocurrencies, such as Bitcoin and its derivatives, were merely used as payment media. Modern users also treat cryptocurrencies as investment or speculation targets in primary and secondary markets or as tokens in gambling and recreational games.

1) CRYPTOCURRENCY AS PAYMENT MEDIUM

a: MINING REWARDS

Mining, or minting, is the process of coin generation in proof-of-work (PoW)-like blockchain systems. The generation of cryptocurrencies requires solving a computationally heavy problem. The party who successfully solves a problem can get a certain amount of cryptocurrency as a payment for their resources spent. The mining process can be either an individual or a collective effort. Individual miners can contribute their computing resources to a mining pool and get rewards any time a peer miner solves the problem.

b: GENERAL FAUCET

Like mining, cryptocurrencies can also be rewarded to users who complete generic tasks, such as solving a captcha. This rewarding process is also called a faucet. In this case, the party to disseminate cryptocurrencies is a human envoy.

c: PURCHASING

Laszlo Hanyecz made the first documented offline purchase with Bitcoin—10 000 BTC for two pizzas—back in 2010 [25]. Nowadays, end-users of Bitcoin can use cryptocurrencies to make various purchases via online marketplaces or offline shops, e.g., multimedia content, electronics,

and clothes. However, the most common purchases made using cryptocurrencies are drugs in darknet markets.

d: RANSOM

Attributing to the anonymous feature of cryptocurrencies, extorsions such as computer malware, human kidnapping, sextortion, and blackmailing often ask cryptocurrencies for ransom. Similarly, phishing emails asking for cryptocurrency transactions are also commonly seen.

e: WALLETS

Online wallets are similar to banks, providing storage services for users' cryptocurrencies. Clients deposit to the wallet by transferring their cryptocurrencies to the wallet services' blockchain addresses and can make payments by sending from online wallet addresses directly. Instead of maintaining their offline public/private key pairs, users use a pair of traditional username and password to access their wallets.

f: MONEY LAUNDERING

The fast-moving nature of cryptocurrencies provides an ideal channel for money laundering. Users can conceal the origin of illegally obtained money by buying cryptocurrencies and later selling them to make the money "clean." Similar activities, such as bribing, also take advantage of the anonymity of cryptocurrencies. However, this process leaves traces in the cryptocurrency transactions and risks being exposed through transaction analysis.

2) CRYPTOCURRENCY AS VIRTUAL ASSETS

a: PRIMARY MARKET

Start-up projects and companies can issue their own cryptocurrencies to represent the equities of their projects. They sell their cryptocurrency for fiat money or other value-bearing media through crowdfunding activities, such as Initial Coin Offering (ICO) and Initial Equity Offering (IEO). Investors who buy these cryptocurrencies usually sell them in cryptocurrency exchanges for a profit later. Start-up projects can also use airdropping, i.e., sending out tokens to investors for free, to gain awareness of their projects.

b: SECONDARY MARKET

Bitcoin was first publicly traded in online cryptocurrency exchanges in 2010. Now, people can buy and sell cryptocurrencies with/for fiat money, cryptocurrencies, or other value-bearing media on these platforms. Exchanges can either execute users' selling and buying orders automatically or allow users to list and match their orders in a forum-like platform, i.e., in an over-the-counter (OTC) fashion. Exchanges also provide various cryptocurrency-related financial products, such as futures and options. Some exchanges even conduct "pump and dump" schemes to manipulate cryptocurrency prices.

c: PONZI SCHEME AND PYRAMID SELLINGS

Fraudulent activities are widely seen in the cryptocurrency economy. Ponzi schemes, or HYIP, are the most common types of fraud [26]. A Ponzi scheme lures investors to its program by awarding early investors with an unreasonably high yield. However, when a new investment slows down, or the scheme organizers see fit, they stop giving out yields and take away all the investments. Ponzi scheme often involves pyramid selling, i.e., by allowing investors to sell the cryptocurrencies to their peer investors for profit.

3) CRYPTOCURRENCY AS TOKEN

a: GAMBLING

Gambling games, such as dice games and roulette, use cryptocurrencies as chips. Gambling games are often the most active applications on blockchain networks.

b: MULTIPLAYER GAMES

The introduction of the smart contract further enables better flexibility to game design. Games such as Cryptokitties allow users to generate tokens with random features, list them for sale, and buy or rent tokens from other users.

B. DATA MODELS

Bitcoin and Ethereum are the two milestones, i.e., blockchain 1.0 and 2.0, in cryptocurrency and blockchain design. Most cryptocurrencies adopted and modified upon their transaction record data models [20]. Bitcoin and its derivatives use the unspent transaction output (UTXO)-based data models, while Ethereum and its derivatives use the account-based data models.

1) UTXO-BASED DATA MODEL

In a typical unspent transaction output (UTXO) data model (see Figure 1), addresses are the basic identities that hold virtual values. An address can be generated offline using Bitcoin's customized hash function to a public key generated by the Elliptic Curve Digital Signature Algorithm (ECDSA) to a user-specified random number as the private key. The transfers of values are settled by recording a transaction in the blockchain. In a typical UTXO transaction, e.g., transaction 1 in Figure 1, all the values stored in the input addresses, e.g., addresses 1 and 2, are transferred to output addresses 3, 4, and 5, with specific value allocations. The output addresses can be further used as the input addresses in the following transactions. Note that, theoretically, the maximum numbers of input and output addresses in a transaction are not limited, but since the size of a transaction record cannot be larger than the block size, the practical total number of input and output addresses has a limit.

Altcoins stand for alternatives for Bitcoin. For example, Litecoin and Dogecoin are typical early altcoins that replicate most of Bitcoin's technical designs. Later altcoins, such as Zerocash and Monero, also adopt the UTXO-data model but use extra cryptography techniques to enhance their anonymity. Each altcoin has its own running blockchain network, which stores the transactions of this particular altcoin exclusively.

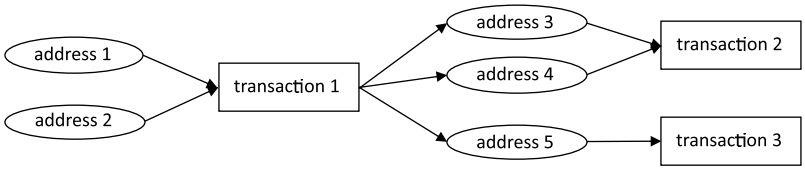

FIGURE 1. UTXO-based transaction data model. Transaction 1 takes two unspent transaction outputs (UTXOs) from address 1 and 2, respectively, and generates three new UTXO outputs to address 3, 4, and 5, which are later used in transactions 2 and 3.

2) ACCOUNT-BASED DATA MODEL

Blockchains can store not only transactions but also other formats of data, including text, image, and even computer codes. The code stored and executed in the blockchain database is also called a (smart) contract. Ethereum and its derivatives, such as Neo and EOS, use account-based transaction data models [27], where the accounts are still blockchain addresses but can be either an externally owned account (EOA) or a contract account (CoA). While an EOA can be created using a similar method as an address in the Bitcoin blockchain, CoAs must be created by a transaction: an EOA or a CoA sends a transaction to a "null" receiver with computer codes written in the auxiliary information. An address will then be generated by the blockchain system and assigned to the CoA. Both types of accounts can hold the blockchain's original cryptocurrency, and their current holdings are stored in the blockchain's running memory.

There are three possible types of transactions in an account-based data model. First, a transfer of the blockchain's original cryptocurrency from an EOA to an EOA or a CoA (transaction 1 in Fig. 2). Second, the creation of a contract or an invocation of the computer codes stored in a contract by transferring a zero-value original cryptocurrency to it with auxiliary information indicating the target function and a set of parameters (transaction 2 in Fig. 2). Third, a token transfer. Contract creators can install a virtual token in a smart contract, allow transfers of the token or part of it between blockchain identities through specific functions like `transfer()` or `distribute()` (transaction 3 in Fig. 2). The contract then records the changes to account balances. Note that in this case, only the amount of the original cryptocurrency or tokens held by EOAs or CoAs will be changed, but no actual transaction is explicitly logged in the blockchain.

3) OTHER TRANSACTION DATA SOURCES

Except for the transactions recorded in individual blockchain networks, transfers of cryptocurrency can also happen across different blockchains or even beyond blockchains.

The most popular blockchains, e.g., Bitcoin and Ethereum, occupy hundreds of gigabytes of space in computer storage and are ever-expanding on a daily basis. New technologies such as the lightning network, sharding, and cross-chain transactions have been proposed to ease the management overhead and reduce the resources required to maintain the blockchain database. A lightning network enables users to create "payment channels" and conduct transactions in the channels before reporting the clearings to the blockchain. Sharding enables the blockchain network to store the entire transaction history but only requires each blockchain nodes to store a proportion of the transaction history. Cross-chain protocols enable users to swap cryptocurrencies across different blockchains. In these cases, the shattered or entangled transaction records add to the difficulty of transaction network analysis.

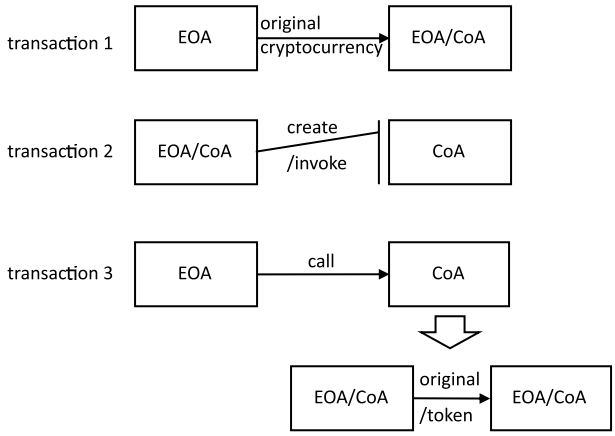

FIGURE 2. Account-based transaction data model. Transaction 1: an externally owned account (EOA) can initiate a transaction of the original cryptocurrency of the blockchain to another EOA or a contract account (CoA). Transaction 2: an EOA or CoA initiates a transaction (no original cryptocurrency is needed to be transferred) to create or invoke a smart contract. Transaction 3: a smart contract call can induce a transfer of original cryptocurrency or user-defined token between accounts.

Users trade cryptocurrencies with other users in online exchanges. Most of the transactions in centralized exchanges are not publicly available, except for rare exceptions, such as leaked datasets from hacking of an exchange's database [28]. However, some OTC exchanges, e.g., bitcoin alpha [29], on the contrary, disclose the transactions to the public. These transactions can be further related to blockchain transactions, providing an auxiliary information source of cryptocurrency flows. Besides, the peer-to-peer nature of OTC exposes users to counterparty risks. Therefore, OTC marketplaces, such as bitcoin-otc and bitcoin alpha, offer a scoring service allowing users to rate each other on their trading honesty. The trust relationships among users form a trust network, enabling various further studies [30], [31].

III. TRACEABILITY AND LINKABILITY ISSUES

A. TRACING CRYPTOCURRENCY TRANSACTIONS

The transparency of cryptocurrency transactions enables forensic analyses of various crimes, using statistical analysis

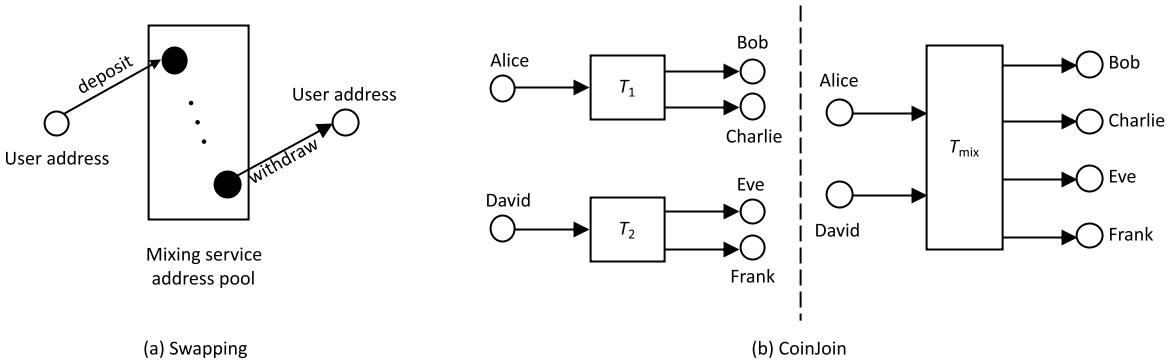

FIGURE 3. Schematics of swapping and CoinJoin mixing. (a) A swapping-based mixing service accepts deposits from users using one of the addresses in an address pool and allows withdraws from another. (b) A CoinJoin mixing allows two or more unrelated transactions (left) to be merged into a single transaction (right).

and graphical visualization techniques to the payment transactions of thefts, ransomware, sextortion, and illicit tradings. Several studies have demonstrated that forensic analyses can reveal the entire course of criminal transactions with surprisingly high accuracy.

The most massive theft in cryptocurrency history was the Mt. Gox exchange hacking in 2011 [32]. The hackers allegedly stole more than 850 000 Bitcoin and led directly to the bankruptcy of the world's once largest cryptocurrency exchange. As cryptocurrencies must be turned into fiat money for the thieves to profit, they must first transfer the stolen cryptocurrency into exchanges. Tracing analysis showed that the stolen bitcoins changed hands several times before landing in exchanges BTC-e, 0x, Bitcoinica, and CryptoXChange. Several interim addresses in the transaction flow, along with those from several other major thefts, passed through addresses that belonged to Alexander Vinnik, the founder and primary beneficiary of cryptocurrency exchange BTC-e. Vinnik was eventually arrested for alleged money laundering in 2017 [6].

The Silk Road market was shut down by the Federal Bureau of Investigation (FBI) in October 2013. The FBI seized 114 336 Bitcoins, i.e., transferred them into an FBI-created arrest custody address. 89% of the Bitcoins came from a set of 15 addresses at the end of a market escrow chain. However, among the distributions that peeled off the chain, more than 100 000 Bitcoins finally arrived at an address not relevant to the FBI [5]. It was believed that not all the darknet Bitcoins were seized, but some were detained by individual FBI agents who were later convicted for stealing in the Silk Road case [33].

Cryptocurrencies are ideal ransom payment mediums owing to the anonymity feature. Many recent computer hacks, such as Cryptolocker and WannaCry, asked for Bitcoin for ransom. However, the ransomware payments, including the infamous CryptoLocker, CryptoWall, DMA Locker, and WannaCry, are all traceable, mostly to a handful of responsible parties [34], [35]. The estimated economic impact of the ransomware from 2013 to mid-2017 was estimated to be USD 12 million minimum.

Moreover, cryptocurrency transactions can also reveal the monetary flows of human trafficking and sextortion. For example, Portnoff *et al.* [36] were able to uncover human traffickers by associating sex ads to specific Bitcoin transactions and addresses with 90% accuracy. Paquet-Clouston *et al.* [37] tracked and investigated monetary flows of a series of sextortion campaigns and found that one single entity was likely controlling the financial backbone worth a minimum of USD 1.3 million.

B. COUNTER-TRACING MEASURES

In the light that the original design of Bitcoin's transactions is easy to trace, smart thieves and extortioners may use counter-tracing solutions to cover the trace of their activities. Typical solutions include mixing services provided by third parties and altcoins with intrinsic privacy-enhancing designs.

1) MIXING SERVICES

Mixing services aim to solve cryptocurrencies' traceability issues by merging irrelevant transactions. The two typical types of mixing methods are swapping and CoinJoin. A swapping-based mixing service (Fig. 3 a) accepts deposits from users to one of the addresses in an address pool and withdraw from another. Hence, the linkage between the deposit and withdrawal addresses are disconnected. Mixing services using swapping include BitcoinFog, BitLaundry, and Helix. The CoinJoin (Fig. 3 b) mechanism allows two or more individual transactions (left) to be combined in a single CoinJoin transaction, which has the same presence as an ordinary multiple-input-multiple-output transaction (right) on the blockchain. Therefore, the relationship between real input-output pairs is obscured. CoinJoin-based services include JoinMarket, CoinShuffle, and Blockchain.info's SharedCoin (ceased service). CoinJoin-like mixing services can be realized by smart contracts on enabling blockchains, e.g., Tornado.cash on Ethereum [38].

Mixing services are designed to hide the relationship between addresses in consecutive transactions. Therefore, they are often used for money laundering purposes for illicit activities. Under today's virtual asset market regulations,

cryptocurrency exchanges can reject deposits from mixing service outflows under anti-money laundering (AML) regulation [39].

2) ALTCOINS WITH PRIVACY-ENHANCING DESIGN

This section covers the two major privacy-enhancing designs in altcoins, i.e., altcoins with decentralized mixing capabilities, such as Zerocoin, Zerocash, Zcash, and Dash; and CryptoNote design, including Monero, Bytecoin, and DigitalNote. We will use Zcash and Monero to illustrate the two lines of designs.

Zcash allows users to store and transact ZEC, i.e., the Zcash cryptocurrency, with two types of addresses (transparent and shielded) [40]. "Transparent" addresses transfer values to other addresses essentially the same way as Bitcoin, while "shielded" addresses make transactions in "shielded pools". In particular, when depositing into the pool, the recipient is specified using shielded addresses, i.e., *z*-address, which hides the recipient but still reveals the sender, and withdrawing from the pool hides the sender but reveals the recipient. The cryptographical basis for the shielded pool is practical zero-knowledge proofs called *zk*-SNARKs. From the perspective of data models, Zcash transactions resemble a swapping mixing pool pattern.

Though zero-knowledge proofs form virtual mixing pools, they suffer from the disadvantage of computational cost. CryptoNote-like cryptocurrencies, such as Monero, take another perspective, i.e., ring signature, to add complexity to the transaction records without causing much computational overhead [41]. A Monero transaction allows several outputs from previous transactions to be merged as its inputs, but only that some of the inputs can be "decoy" as their values are never transferred to the output.

C. TAINT ANALYSIS TECHNIQUES

Mixing services and altcoin designs add extra layers of complexity to the transaction record by hiding the association between blockchain addresses. Nevertheless, they can still leave sufficient information for "taint analysis", i.e., tracing monetary flow and linking related blockchain addresses. We categorize existing taint analysis techniques into several types of heuristics and the subset-sum matching approach.

1) HEURISTICS

Commonly used heuristics in address associating and linking include the multiple inputs rule, the coin change rule, the zero-mix rule for cryptocurrencies with ring-signature design, and temporal heuristics. We formulate the first three types using graph representations, as shown in Figure 4.

Multiple inputs is the most basic and widely adopted rule in associating UTXO addresses that potentially belong to the same user. When initiating a transaction, users have to sign the transaction with all of the input addresses' private keys. Therefore, all the input addresses in a transaction can be assumed to be owned by the same party [4], [10], [42]. For example, addresses 1 and 2 in Figure 4a can be considered belonging to the same user.

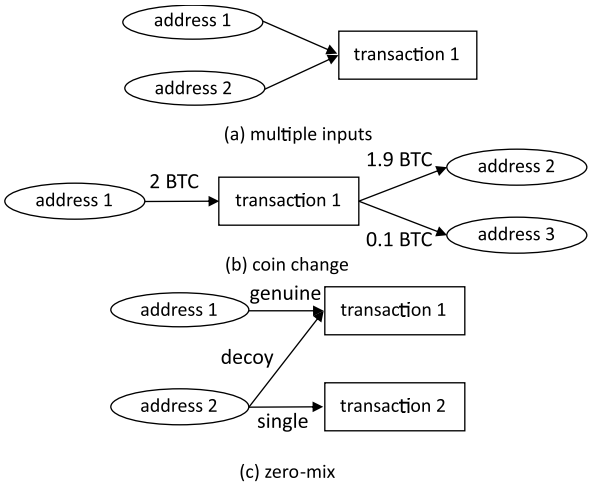

FIGURE 4. Commonly used heuristics in taint analysis. (a) Multiple inputs: all the input addresses, e.g., address 1 and 2, in a transaction are presumed to be owned by the same party. (b) Coin change: if the input value is larger than the designated transaction value, the residue must be returned to a (usually new) address held by the transaction initiator, e.g., address 3. (c) Zero-mix: the single UTXO input transaction (transaction 2) reveals that the input UTXO from address 2, which was also used in a previous transaction 1, is actually a decoy input before; hence the input from address 1 was the real input for transaction 1.

Coin Change: The input value and output value in a UTXO transaction must be equal. If the input value is larger than the designated transaction value, the residue must be returned to a (usually new) address held by the transaction initiator, i.e., a change address. For example, addresses 3 in Figure 4 b is potentially a change in this transaction, and hence, belongs to the same user who owned address 1.

The roll-out of the "coin selection" strategy in the official Bitcoin offline wallet, Bitcoin Core, in 2012 made the coin change rule even more apparent. When the user enters the amount of Bitcoin to be transferred to destination addresses, the client software automatically chooses the set of input addresses with an exact match to the value or a minimum change output. The change addresses usually hold only a small value and typically appear in transactions only once or twice. The coin change rule is usually used in conjunction with the multiple inputs rule [10], [42], [43].

Except for the most commonly used multiple inputs and coin change rules, other heuristics for Bitcoin transactions may also consider that all the output in a coinbase transaction belong to the same entity [44] or exploit specific transaction patterns, e.g., apparent self-transferring operations and those that resemble money laundering activities in a conventional banking system, to associate addresses [45].

Zero-mix, aka cascade attack or cascade effect, is a specific heuristic for CryptoNote cryptocurrencies. Let's assume the scenario as shown in Figure 4 c, where transaction 1 takes two unspent-inputs from addresses 1 and 2 at time t and transaction 2 takes the unspent-input from address 2 at time $t+\delta t$. In this case, transaction 2 takes no foreign outputs used as mix-ins for the associated ring-signature and therefore is a zero-mix transaction. Hence, address 2 must be a real input in transaction 2, and a decoy in transaction 1 [41].

Zero-mix is a special case in the more general "closed set" attack proposed in [46]. If the number of inputs equals the number of distinct public-keys included in a CryptoNote transaction, i.e., forming a *closed-set*, they are all real-spends in this transaction hence decoys in other transactions. A computationally efficient realization of this attack was also proposed in [46].

Temporal heuristics exploits the timing of deposit and withdraw transactions in mixing services and altcoins transactions. For example, temporally close deposit and withdraw transactions in Zcash's shielded pools have a high possibility of linkability [47]. Moreover, the ages of transaction outputs could also differentiate genuine from decoy inputs in CryptoNote transactions [41].

2) SUBSET SUM MATCHING

Mixing services and altcoin designs follow a similar basis: merge irrelevant transfers into one blockchain transaction record to decrease these transactions' traceability. However, the fingerprints of transferred cryptocurrency values could still reveal the actual input-output pairs.

Considering that users would deposit and withdraw the same amount of cryptocurrency to and from the mixing service, a widely adopted method to find input-output pairs is to find matched values or value combinations in the multiple-input-multiple-output transactions [41], [48], [49]. This problem can be related to the classical subset sum problem. Given a set of inputs $I = \{I_1, I_2, \ldots\}$ and a set of outputs $O = \{O_1, O_2, \ldots\}$, the subset matching tries to find the exact match or the most probable matches between subsets of $I' \subseteq I$ and $O' \subseteq O$. As shown in Figure 5, the input values of address 1 and 2 match the output to address 5, and therefore, addresses 1, 2, and 5 might belong to the same user.

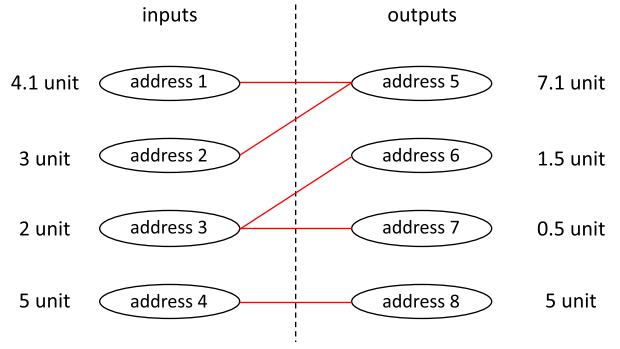

FIGURE 5. Illustration of the subset sum problem in tracing merged transactions. The sum of input values of address 1 and 2 matches the output to address 5. Therefore, addresses 1, 2, and 5 may belong to the same user. Other matches in this transaction are also linked by the red lines.

Note that the exact subset sum problem is NP-complete. However, transaction fees are generally charged in a transaction, resulting in the sum of the deposits slightly larger than that of the withdraws. Therefore, one can only try to find the most probable, rather than an exact match, between the subsets of input and output. The approximate subset sum problem becomes NP-hard. Simplification of the problem can impose a time constraint, e.g., only consider temporally close input and outputs, to narrowing down the search for possible matches [50]. Special cases of the matching method are sometimes referred to as *value fingerprinting heuristic* [40] or "round-trip" transactions [47], where single input and output values have the exact same values or multiple digits [51].

Despite not yet applied to the transaction matching, several classical heuristic algorithms can solve the approximate subset sum problem in polynomial time. For example, the ε-approximation algorithm [52] which finds an approximation sum $\hat{P}$ within ε of the desired optimal sum P^*, i.e. satisfying $(P^* - \hat{P})/P^* \leq \varepsilon$, where $\varepsilon = 1/(1+k)$ and k is a positive integer, can achieve time complexity $O(n^k)$, and n is the problem size. Moreover, the ε-approximation algorithm can be further optimized to achieve time complexity $O(n)$ [53].

D. FINDINGS

Overall, Bitcoin and its replicates showed strong traceability: 87.6% of transactions in Bitcoin, 88.5% in Bitcoin Cash, and 85% in Litecoin have a single output [54]. As for linkability, using the multiple inputs rule, the number of clusters of associated Bitcoin addresses is approximately 45%–70% of the total number of addresses [22]. When the coin change rule is applied with the multiple input rule, the number of associated Bitcoin address clusters further decreases to 22%–37% of the total number of addresses [22]. Associating addresses can reveal address identities if one of the associated addresses is revealed. The largest cluster of associated Bitcoin addresses up to 2012 (156 722 addresses) were identified as the cryptocurrency exchange Mt. Gox, followed by well-known wallets and mining pools [55]. Liao *et al.* [56] successfully associated 968 unknown addresses to two addresses belonging to CryptoLocker found in Reddit.

However, heuristic rules are also prone to error. For example, the multiple inputs rule's effectiveness depends on repeated address use by a single user and can lose its efficacy if no UTXO address is reused or with CoinJoin-like mixing and ring signature altcoins [57]. Nick [42] tested the multiple inputs and coin change rules on a dataset containing 30 000 sets of Bitcoin addresses from the leaked BitcoinJ wallets, where each set of addresses belongs to an end-user. Using the multiple inputs rule alone, the associated clusters of addresses achieved a 68.6% average recall ratio. With the coin change rule imposed, the average recall ratio only rose to 69.3%, which accounts for less than 1% of an increase in the accuracy. Ermilov *et al.* [58] showed that addresses clustered by mere heuristic rules could belong to several different entities, and therefore, proposed a probabilistic algorithm that utilizes known address tags to improve the heuristic rules.

Due to the lack of mixing services' ground truth data, reported precisions of mixing services' taint analysis are rare. Nonetheless, Hong *et al.* [50] used subset-sum matching and temporal heuristics and found that 99.1% of the input and

output transactions performed by the mixing service provider Helix can be associated.

Although providing decentralized mixing functions, most Zcash activities are in the transparent part of the blockchain. Mining pools play essential roles in Zcash ecology. Mining rewards can be linked to 87.5% addresses and 25.7% transactions [59], and that 95.5% of the total number of Zcash transactions are potentially linkable to public addresses by just observing the mining activity [60]. What is more, 31.5% of all coins sent to shielded addresses can be matched to public addresses using taint analysis techniques [47]. The primary users of the shielded pools are still mining pools: 65.6% of the value withdrawn from the shielded pool can be linked back to deposits made by either founders or miners [61].

The multiple inputs rule loses its efficacy against CryptoNote designs, as the decoy addresses are usually not owned by the holders of the real inputs. However, careless usage of transaction outputs together in a new transaction can still reveal the ownership of the original transaction outputs [41]. Using zero-mix rule only, Möser *et al.* [62] was able to identify 62% of the real inputs in Monero transactions with one or more mixes. The accuracy increased to 80% when temporal heuristics are considered. Kumar *et al.* [63] found that in 87% of cases, the real Monero output from a redeemed can be easily identified with certainty with temporal heuristics. Yu *et al.* [46] were able to identify the real coin being spent in 71% Monero inputs, 74% Bytecoin inputs, and in 92% DigitalNote inputs, using the zero-mix rule and their ''closed set'' attack.

IV. COLLECTIVE TRANSACTION PATTERNS

The cryptocurrency economy, whose activities are faithfully recorded in the transaction records, resembles a typical complex system. The emergence of collective patterns is commonly found in complex natural systems, e.g., the fractal patterns in snowflakes, and socio-techno-economic complex systems, e.g., the 20/80 wealth distribution rule in human society. These often counter-intuitive behaviors cannot be simply explained by an aggregate of agent behavior but have to be understood from a connective perspective [64], [65]. This section surveys the existing analysis on cryptocurrency transactions from a system perspective.

A. TRANSACTION NETWORK CONSTRUCTION

Network is an essential perspective of modeling complex systems. Networks are ubiquitous in physical, technical, social, and economic systems with interconnected components. The study of complex networks in the past 20 years has shown that real-world networks possess universal underlying structural properties, such as scale-free property and small-world phenomena, and similar network formation mechanisms, e.g., preferential attachment [66]. Therefore, network analysis is widely adopted in the study of cryptocurrency transactions and has proven particularly useful in characterizing cryptocurrency user activities by examining these networks' temporal and structural properties.

A network $G = (V, E)$ consists of two sets of entities, i.e., a collection of nodes, V, and a collection of edges, E, connecting the nodes together. The construction of transaction networks for UTXO, account-based, and other transaction data models have subtle differences. This section summarizes three types of networks constructed from UTXO transactions, three types from account-based transactions, and discusses transaction network construction from the Lightning payment channels.

1) UTXO TRANSACTION NETWORKS (FIG. 6)

Type I network, or address network, uses addresses as nodes in the network and the flow of virtual values as directed edges [67]–[69]. For example, input addresses 1 and 2 are connected to addresses 3, 4, and 5 with directed edges representing the value flow in transaction 1. The edges are usually unweighted because the actual transferred value between a specific pair of addresses is not explicitly given. Multiple-input-multiple-output transactions may create large cliques in the network.

Type II network, or transaction network, uses transactions as nodes in the network, in which transactions are connected by directed edges, which represent the sharing of output/input addresses [49], [70]. For example, transaction 1 is connected to transaction 2 with directed edges, as the output addresses 3 and 4 of transaction 1 are used as the input addresses in transaction 2. The edges in this type of network can be weighted, i.e., the value being held in the interim addresses at that time. In some cases, Type I and Type II networks are combined, i.e., all the addresses and transactions are nodes in the network, while the input and output relationship between the addresses and transactions are considered edges [71]–[73].

Type III network, or user network, considers the directed flow of cryptocurrencies between users [55], [74], [75]. However, since the blockchain addresses are anonymous, i.e., the addresses cannot be tied to user identities directly, association rules must be applied to associate addresses with potentially the same identity. Also, note that the Type II network is actually an interim state between the Type I and Type III networks in that it only considers all the inputs of a transaction being an entity but does not further merge all the entities with shared input addresses.

2) ACCOUNT-BASED TRANSACTION NETWORKS

Type I network, or the original cryptocurrency transfer network, uses EOAs and CoAs as nodes in the network and the flow of non-zero original cryptocurrency as directed edges [68], [76]. The edges are directed and weighted by the amount sent in the transactions.

Type II network, or token transfer network, also uses EOAs and CoAs as nodes in the network but uses the flow of user-customized tokens as directed edges [77], [78].

Type III network, or invocation network, considers the creations and non-transactional function calls to smart contracts [12]. This network uses EOAs and CoAs as nodes, with directed edges pointing from the invoker addresses to

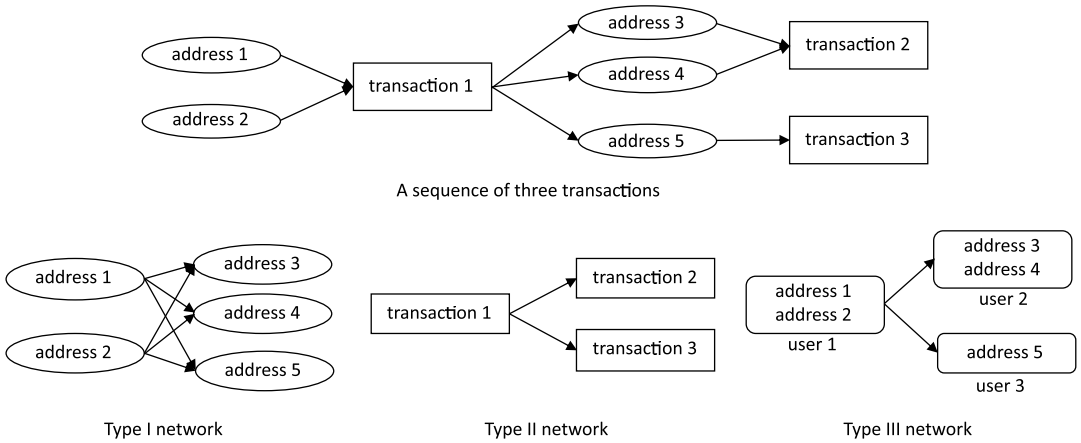

FIGURE 6. Network construction from UTXO-based transactions. Three types of networks can be constructed from the sequence of three transactions. Type I network uses addresses as nodes and the flow of virtual values as directed edges. Type II network uses transactions as nodes and the sharing of consecutive output/input addresses as directed edges. Type III network uses users, which may hold one or more addresses, as nodes and the flow between them as directed edges.

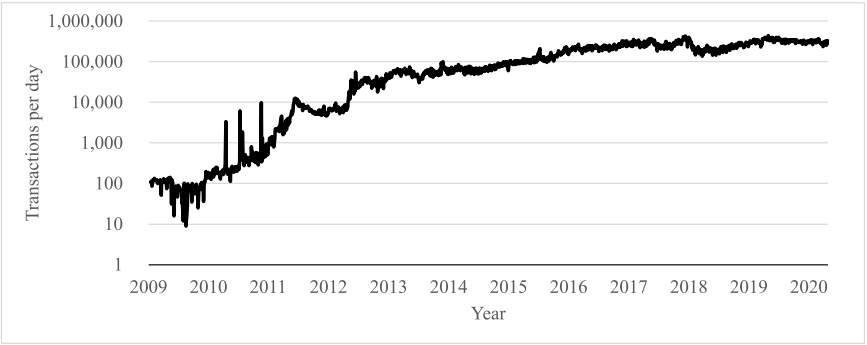

FIGURE 7. Number of transactions per day in the Bitcoin blockchain network from 2009 to 2020. The growth can be roughly divided into two stages: the initial stage (until the end of 2010) and the trading stage (since 2011).

the contract addresses, representing the creation or invoking relationship.

3) LIGHTNING NETWORK

Lightning Network (LN), launched in 2018, is the mainstream payment channel network (PCN). It attempts to relieve the pressure of the ever-expanding Bitcoin ledger by creating payment channels across which any two users could exchange off-chain payments without burdening the entire network. A user can open a payment channel with another by sending an amount of Bitcoin to a P2WSH address (at least 75% of all P2WSH transactions are Lightning transactions [79]). The two users then transfer coins between each other without writing the transactions in the blockchain. Upon payment channel closure, the P2WSH address transfers the cleared amount of coins back to one or two of the users. A user can open payment channels with multiple users at the same time. Therefore, the payment channel network is an undirected network of Bitcoin users connected by payment channels where off-chain payments are routed [80]. Information about live payment channels can be found on monitoring websites such as hashxp.org, 1ml.com, and lndexplorer.com [81].

B. EMERGING STRUCTURAL PROPERTIES

1) GROWTH OF THE NETWORKS

As of 2020, the total number of unique Bitcoin addresses appearing in the transactions was more than 600 million [82], and that of Ethereum over 100 million [3], with a daily increase of hundreds of thousands [2]. Therefore, new nodes and edges were introduced continuously into the transaction network, resulting in ever-expanding network size. As the first cryptocurrency, the growth of Bitcoin networks went through two stages: the initial stage (until the end of 2010) and the trading stage (since 2011) [68], [83], [84] (see Fig. 7). The networks of other cryptocurrencies showed similar trends.

The growth rate of the networks depends on the adoption of cryptocurrencies. In the initial stage, Bitcoin was still an experimental idea used only by a small group of early adopters. The transaction volume was low, and the network structure fluctuated severely. The growth in the numbers of nodes and edges, network diameter, and the average distance between nodes correlated positively with the Bitcoin price [85]. The rise of the Bitcoin price, in turn, attracted more attention to Bitcoin, which further expanded the user community and hence the network scale [86]. The average

balance of Bitcoin in the blockchain addresses decreased as the user community expanded [86]. In the trading stage, Bitcoin began to be accepted by a broader range of users and began to flow between addresses, resulting in a quickly expanding network size, and the structural properties, such as degree distribution and clustering coefficient, began to stabilize.

2) A CENTRALIZED NETWORK

The degree distribution is the most common characterization of networks. Many real-world networks, such as the World Wide Web, movie actor collaboration network, and power grid networks, exhibit long tail, and sometimes power-law, distributed node degrees [87]. In these networks, most nodes have a limited number of neighbors, but some nodes can have a massive amount of connections. The degree distribution of the Bitcoin network (Types I and III) converged to a power-law distribution gradually over time, resulting in a scale-free network around 2010 [69], [75], [85]. The Type I [68], [76] and Type III [88] Ethereum networks also exhibit power-law degree distributions with the power-law exponent $\gamma \approx 2$.

Preferential attachment, one plausible mechanism driving real-world networks' evolution, refers to new nodes joining the network tend to connect to existing nodes with higher degrees. The preferential attachment was observed in Type I Bitcoin network's growth: hub nodes grow faster than low-degree nodes [83]. However, preferential attachment to the higher degree or richer nodes may not be an accurate mechanism to the Type III Bitcoin network's growth. Biryukov *et al.* [84] proposed a fitness preferential attachment mechanism, where the fitness of a node v is its potential to create new connections, i.e.,

$$f_v(t) = \frac{k_v(t) - k_v(t-1)}{\sum_{u=1}^{m} (k_u(t) - k_u(t-1))}, \tag{1}$$

where t is the number of months starting from January 2019, k is the degree of a node, and m is the number of nodes in this month's network. There is no bounded interval for the value of the fitness of a node. The higher the value, the higher the ability of a node to attract new connections. Different types of users have different intrinsic fitness: cryptocurrency exchanges are more attractive to connections than active traders, who are then more attractive than a common adopter of Bitcoin.

Another possible outcome of preferential attachment is that nodes with small degrees connect disproportionally to those with large degrees. This connection pattern is also referred to as disassortative mixing, e.g., nodes tending to connect to those with different structural properties. Most early-stage cryptocurrency transaction networks, such as the Type I Bitcoin network from 2009 to 2013 [83], the Type I Litecoin network from 2009 to 2010 [69], and the Type I Ethereum network from 2015 to 2017 [76] all showed disassortative mixing. In the cryptocurrency economy, disassortative mixing reveals that most transactions happen between end-users and popular services such as cryptocurrency exchanges, wallets, gaming, and gambling services, yet less frequently among the end-users themselves. However, the assortativity of the transaction networks increased over time. For example, the Type I Bitcoin and Type I Namecoin networks' assortativity eventually converged to 0 [68]. This increment could have two causes: 1) high-degree nodes began to transfer cryptocurrencies among themselves, e.g., cryptocurrencies moving between multiple holding addresses with the same exchange, or 2) low-degree nodes could have started to have more interactions among themselves.

Not surprisingly, the Lightning Network also evolved into a centralized network. The early LN was disconnected, consisting of mostly small clusters and occasionally larger cliques [79]. On June 13, 2018, the LN's snapshot contained 1355 nodes in the weakly connected component and 889 nodes in the disconnected periphery, exhibiting a degree distribution with $\gamma \approx 2$ [89]. The LN gradually grew into a centralized network with a distinct core-periphery structure. The Bitcoins distribution in each channel has an unequal Gini coefficient of the node strengths 0.88 in 2019 [90]. The centralized structure of the payment network exposes it under targeted attacks, i.e., a DDoS attack targeting hub nodes can remarkably sabotage the LN's efficiency [91], [92].

Since most commonly seen complex networks evolve power law in their degree distributions, a transaction network that deviates significantly from these rules is usually induced by anomalous activities [93]. Although the in-degree and out-degree distributions of the Bitcoin network generally followed the power-law distribution, some severe fluctuations exist in the distributions [74], [94]. Maesa *et al.* [74] believed that these anomalies are caused by a deliberate transaction pattern called "pseudo-spam chain," i.e., a large number of tiny value transactions. Similar disruptions can also be found in the transaction time interval distribution in the Ethereum blockchain: instead of following a strict power law, the distribution has some spikes at certain time intervals [95]. This anomaly was conjectured to be due to deliberate individual activities such as trading bots.

3) A SMALL WORLD

Many complex networks show two characteristics when growing: densification, i.e., the increase in the number of edges in the network is super-linear to that of the nodes; and shrinking diameters, i.e., the average path length of the network, shortens [96]. These characteristics are the indicators of small-world networks, along with a large clustering coefficient [97].

Super-linearity between the increment of the number of edges, M, and the number of nodes, N, can be characterized by

$$M(t) \sim N(t)^{\alpha}, \tag{2}$$

where $M(t)$ and $N(t)$ are the number of edges and nodes in the network at time t, respectively. If $\alpha > 1$, super-linearity presents in the network evolution, and the average degree

of all nodes will also increase over time. Super-linearity appeared in the early stages of transaction network growth. For the Type I Bitcoin network, the average out-degree of nodes increased from less than two in 2009 to around six in 2012 [67]. The nodes' out-degrees in the Type III Bitcoin network increased from around 2.6 in 2013 to around 3.1 in 2015 [74]. The Type I Ethereum network showed a general linear growth between the number of edges to the number of nodes, i.e., $\alpha = 1.0$, from 2015 to 2017, but super-linearity emerged, i.e., $\alpha = 1.38$ in the last 1/3 of the transaction record [68].

However, the super-linear growth of the transaction networks did not persist [68]. Although $\alpha = 1.15$ for the Type I Bitcoin network constructed from transactions spanning from 2009 to 2017, it decreased to 0.86, i.e., demonstrating a sub-linearity, between 2014 and 2017. The situation was similar in the Type I Namecoin network, in which α decreased from 1.05 to 0.99. A decreasing α means that users tend not to reuse previously used blockchain addresses and create new addresses in transactions to preserve their anonymity better. Therefore, the network evolution enters a new stage in which nodes' growth rate surpasses that of edges, and the network becomes sparse. However, the final stage may not happen to account-based blockchains such as Ethereum because the account creation procedure on these blockchains is much more complicated than UTXO blockchains. A less user-friendly address creation procedure may discourage users from creating new accounts on the blockchain.

"Six degrees of separation" is a common metaphor for small-world networks, emphasizing a short average path length between each pair of nodes in the network. The average shortest path length in the Type III Bitcoin network's largest connected component decreased from around five to around four from 2013 to 2015 [74]. On the contrary, the average shortest path length in Ethereum increased. The average distance between nodes in weekly snapshots of the combination of Type I and Type III Ethereum networks increased from around four in 2015 to five in 2018 [98]. The shrinking of the average distance in a Bitcoin network may be attributable to the introduction of popular services such as exchanges and wallets during that period. The expanse of network distance in the Ethereum network may be because a great number of new users were adopting the blockchain network and creating a large number of new nodes, which were not yet densely connected.

A typical small-world network has a significantly larger clustering coefficient than a randomized network with the same size and density [97]. The average clustering coefficient of the Type III Bitcoin network was around 0.15 during 2011 and 2013 [75]. For the Type I Bitcoin network, the clustering coefficient was as high as 0.22 during 2010, but it decreased to 0.04 in 2014 [69] and stayed stationary around 0.05 afterward [68]. The clustering coefficient of the combination of Type I and III Ethereum networks was close to 0 initially but increased to and remained at around 0.01 after block height 3 000 [98]. The clustering coefficients in different transaction networks were all higher than those in the randomized networks.

4) A BOW-TIE STRUCTURE

Large directed networks can evolve into a visually "bow-tie"-like structure (Fig. 8). The bow-tie structure consists of four components: 1) a strongly connected component (SCC), which is the core of the network; 2) an in-component (IN); 3) an out-component (OUT), which are the sets of nodes reachable to and from the SCC, respectively; and 4) the tendrils, which are the sets of nodes unreachable to and from the SCC. Guo *et al.* [76] found that monthly snapshots of the Type I Ethereum network possess such a bow-tie structure. The SCC probably contains the hub nodes, e.g., exchange and wallets, and the IN and the OUT components are most likely the end-users. The tube component composed of tendrils can be regarded as extra bridges from the IN component to the OUT component.

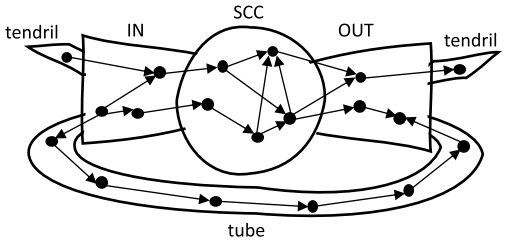

FIGURE 8. A schematic illustration of the bow-tie structure in a directed network. The bow-tie structure consists of four components: 1) a strongly connected component (SCC), which is the core of the network; 2) an in-component (IN); 3) an out-component (OUT), which are the sets of nodes reachable to and from the SCC, respectively; and 4) the tendrils, which are the sets of nodes unreachable to and from the SCC.

5) COMMUNITY STRUCTURE

A network contains a community structure if it can be easily grouped into densely-connected sub-networks. Community detection algorithms are used as an enhancement to the heuristics rules of blockchain address association. For example, Zheng *et al.* [99] conducted a two-step clustering process to Bitcoin addresses, i.e., first, use the multiple inputs and coin change rules to associate addresses, then use the Louvain algorithm to partition the transaction addresses into several communities further. They were able to find a set of CryptoLocker blackmail addresses using this process. Cazabet *et al.* [100] used multiple inputs and coin change heuristics to construct Type II Bitcoin network and used a community detection algorithm to partition the network into different activities further and improved the results obtained in [10].

C. OTHER COLLECTIVE PATTERNS

Preferential attachment results in not only a skewed degree distribution but also a centralized accumulation of *wealth of blockchain addresses*, e.g., wealthier addresses accumulate cryptocurrencies significantly faster than the less wealthy

ones [83]. The Bitcoin [55] and Ether [68], [76] wealth possession distribution in blockchain addresses both exhibit power-law. The highly wealthy addresses are not necessarily individual end-users but can also be exchanges or wallet services. Moreover, power-law distributions were also observed in the transaction value [85] and the time intervals between consecutive Bitcoin transactions initiated by the same addresses [83].

The increasing centralization has also shown in the *creation and usage of smart contracts*. Until 2018, smart contracts in Ethereum are only used to develop simple token-centric applications, e.g., ICOs and crowdsales. Eighty percents of the smart contracts use at most 211 instructions [101]. Smart contract code similarity reveals substantial code reuse, where less than 10% of user-created contracts are unique, and less than 1% of contract-created contracts are so [102]. Moreover, contracts are three times more likely to be created by other contracts than they are by users [102]. As a result, 0.05% of the smart contracts are the target of 80% of the transactions [101], and that over 60% of contracts have never been interacted with [102]. Pinna *et al.* [103] surveyed 10 000 smart contracts source codes and a dataset of meta-data from Etherscan.io. They found that the number of transactions and the balances of these contracts follow power-law and that the 20 smart contracts with the topmost number of transactions are all financial contracts.

Transaction fees are the small amount of money that a user pays to the miners, i.e., the blockchain ledger keepers, when initiating a transaction. Depending on the busyness of the blockchain networks, fees may vary. The Bitcoin transaction fee per transaction surged at the end of 2017, to over USD 50, due to intensive network activities at that time and stabilized at several US dollars in 2020 [2]. Notably, a non-negligible amount of Bitcoin addresses possess only a ''dust'' amount of values, i.e., they cost more in transaction fees to spend than the output value [54]. Transaction fees in the Ethereum blockchain are called gas, which is the cost necessary to perform a transaction by miners. A transaction involving complex smart contract execution can have higher prices than an ordinary Ether transaction [104]. Miners can set the minimum gas price and decline to process a transaction if it does not meet their price threshold. Pierro *et al.* [105] found that the number of pending transactions and the number of miners in the network significantly influence Ethereum gas fees.

V. INDIVIDUAL BEHAVIOR ANALYSIS

An extensive literature focuses on characterizing and differentiating the activities of a specific type of agent in the cryptocurrency economy. The identities of blockchain addresses can be obtained from public online venues or interacting with known cryptocurrency services. Then, transaction features are extracted for the exploratory study of user behaviors as well as downstream machine learning tasks to classify the identities of blockchain addresses or look for anomalies in the transaction records.

A. TAGGING ADDRESSES

Despite the anonymity nature of cryptocurrencies, service providers, such as exchanges, wallets, and gaming, choose to disclose their blockchain addresses publicly. Some end-users also post their addresses in online marketplaces or forums to collect payments. Addresses related to major theft cases [4] and Ponzi schemes [106] can also be found on Reddit and BitcoinTalk. Researchers can also proactively collect address identities using cryptocurrency services and tracing the transactions from their own addresses [10]. Furthermore, the Ethereum Naming Service assigns humanly readable names to complex hash addresses, and therefore, can be used to reveal blockchain addresses' identities [107].

The most commonly revealed addresses include cryptocurrency exchanges, merchants, escrow services, mining pools, gaming, gambling, and online wallets. Other less commonly found labels include mixing services, various scams, including Ponzi schemes, ransoms, stolen bitcoins, and attackers. Start-up projects also disclose their holding addresses when disseminating tokens in the primary market. Note that, using the addresses association techniques mentioned in Section III, when one address in a node is tagged with a label, the label can be automatically inherited by all the other associated addresses.

Today's online intelligence platforms such as Blockchain.info [2], Etherscan.io [3], and WalletExplorer [108] maintain lists of known blockchain addresses for user reference. Blockchain data analysis service providers, such as Elliptic [109] and Chainalysis [110], also provide address labels to collaborators for research and law enforcement purposes [111].

B. TRANSACTION FEATURES

We categorize the commonly considered transaction patterns for individual addresses into four categories: volume, temporal, network structural, and contract code features.

Volume features of an address include attributes such as the numbers of incoming/outgoing transactions, total volume of these transactions, balance, transaction fee paid, mining rewards revived, and aggregated figures such as the sum, average, mean, and standard deviation of the previous features.

Temporal features of an address include the activity period duration, activity intensity, and the average, mean, and variance of activity time intervals, as well as the skewness and kurtosis of the time interval distribution.

Transaction network structural features include node centrality, motif, network embedding, and neighbor identity information. Centrality is a measure of the importance of the nodes and edges in the network [112]. Generic node centrality measurements include degree centrality, Eigenvector centrality, PageRank centrality, betweenness centrality, and closeness centrality [113]. Transaction network-specific centralities include the generalized entropic centrality proposed in [114] and mint centrality proposed in [70]. The generalized entropic centrality of a node u in Type III UTXO networks

measures how likely a monetary flow goes from u to any other node w. First, the probability of flow starting at u and end with w is defined by

$$p_{uw} = \sum_{P \in \mathcal{P}s,w} \prod_{v \in Ps} \tau_{P_v(v)} \frac{f(v', v)}{|\mathcal{S}(P_v)|}, \tag{3}$$

where s is an auxiliary vertex that serves as the source of the flow and has a single edge pointing to i; the sum is over all paths P from s to w; $\tau_{P_v(v)}$ is the probability that the flow goes out of v on a particular subset of $\mathcal{N}(Pv)$, given the path P_v was used to arrive at v; and $f(v', v)$ is the amount of monetary flow from v' to v. The mint centrality in Type II UTXO network of an address A at a given block height h is the number of distinct block heights of coinbase transactions with which address A can be associated, through the transaction outputs it owned at any height prior to and including h. For example, in a Type II UTXO network as shown in Fig. 9, the squared boxes indicate different blocks; C_1 and C_2 are coinbase transactions; T_1, T_2, and T_3 are non-coinbase transactions. The mint centrality $mc(A, h)$ can be computed as follows:

$$mc(A, h) = \frac{1}{h} \sum_{j=1}^{h} rh_j, \tag{4}$$

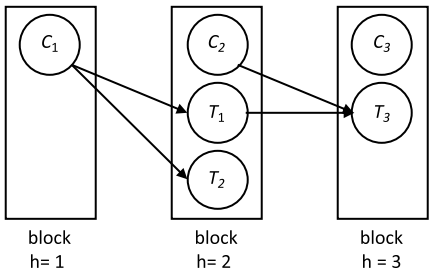

FIGURE 9. Schematic of mint centrality. The squared boxes depict data blocks; h is the given block height; C_1 and C_2 are coinbase transactions; T_1, T_2, and T_3 are noncoinbase transactions.

where rh_j is either 1 or 0, representing whether or not the coinbase at height j is linked to address A. Highly ranked Bitcoin addresses in mint centrality belong to SatoshiDICE as well as its associated addresses, faucet, and donation addresses.

Motifs are small building blocks that center around, start from, or end with a target node. In some networks, particular motifs appear more frequently than in the randomized network. These signature motifs in the network can often reveal the functional features of the real systems [115]. The smallest motif is a loop of two nodes connected by a pair of directed edges. A larger motif can consist of a particular connection pattern of three or four nodes. The clustering coefficient can also be categorized in this feature class because it calculates the number of triangles around a central node.

Network embedding encodes the structural features of a node into a low-dimensional space. Embeddings are usually constructed by network representation learning, i.e., an end-to-end training method that automatically transforms a network structure into a low-dimensional space. Early network representation methods include deepwalk [116], node2vec [117], or other customized biased random walks [118]. These methods capture the similarity between nodes as the overlap of neighbor nodes found by a random walk. By contrast, graph neural networks (GNN) take a network as the input and predefined node labels as the output and learn the nodes' feature vectors in an end-to-end learning scheme [119]. Since the transaction networks are temporal, i.e., the structure changes with time, temporal GNN models such as EvolveGCN are also used in transaction analysis [120].

Neighbor identity features are usually dummy coded features indicating the existence of a labeled sample in a node's neighbors.

Contract code features. Source code and compiled code features are commonly used for the analysis of smart contracts. Code stylometry refers to the quantification and measurement of unique styles, e.g., wording frequency and the use of specific variable and function names [121], [122]. Furthermore, the symbolic analysis of programs' control flow can also yield multiple useful information for identifying bugs or malicious contracts [123].

C. SIGNATURE BEHAVIORS OF CRYPTOCURRENCY ECONOMY AGENTS

1) MINERS

Blockchain networks reward the ledger keepers for their resources consumed. For example, a coinbase transaction is written as the first record in each block of the Bitcoin blockchain, containing a particular value of Bitcoin transferred from a ''no input'' address to one or more miner specified addresses. Bitcoin mining rewards were given to single miners in the early days when Bitcoin was only adopted by a small group of early players. However, when Bitcoin's price surged, the mining game changed. The difficulty of the mathematical problem that ledger keepers need to solve skyrocketed, and single miners possessed little chance to solve a problem alone. Therefore, small miners formed or joined mining pools to pool their computational power and share the mining reward based on the resources invested [124]. Ren and Ward [125] found that the percentage of pool-mined blocks was already 91.12% in Bitcoin and 92.2% in Ethereum in July 2018. Mining pool addresses are with the highest degree-based and betweenness centralities in the transaction networks [126].

The transaction network structure helps to reveal the reward distributions among pool members in different mining pools. The simplest reward distribution method for a mining pool is to record a coinbase transaction with multiple output addresses, which directly belong to the miners (Fig. 10 a). However, this distribution method has several downsides. First, when the mining pool grows big, e.g., to hundreds or more members, the mining reward cannot be distributed in a single transaction as the block size limits the maximum number of input and output addresses a transaction contains. Second, it requires a consensus of fixed distribution among

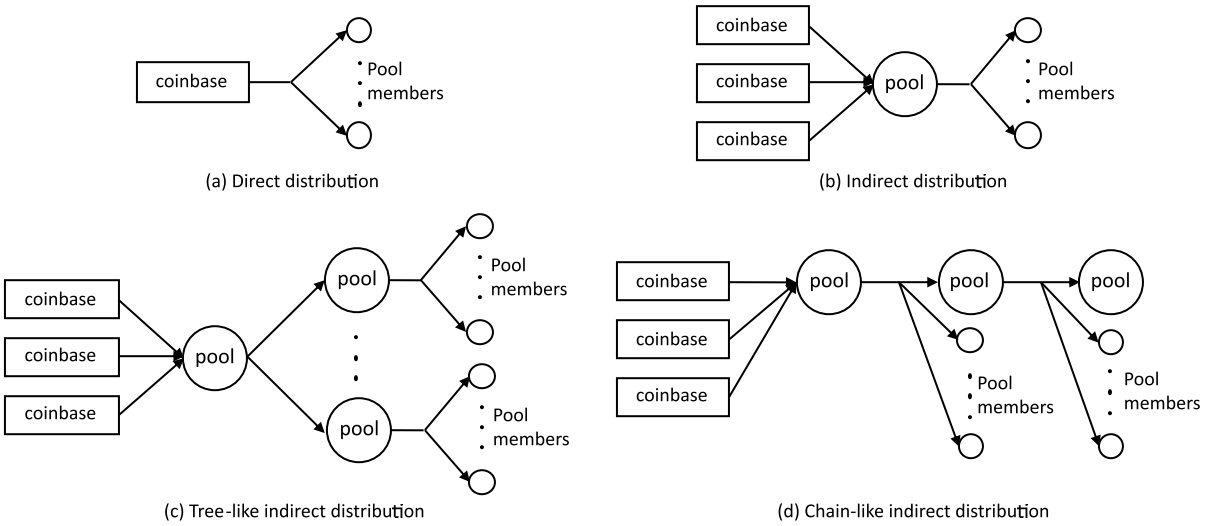

FIGURE 10. Distribution modes of mining pools in PoW UTXO cryptocurrencies. (a) Direct distribution: a coinbase transaction with multiple output addresses which directly belong to the miners. (b) Indirect distribution: use an interim reward holding address to aggregate all the output from coinbase transactions before distributing them among the pool members. (c) Tree-like indirect distribution: divide the reward into multiple holding addresses and use them to initiate multiple output transactions to the pool members. (d) Chain-like indirect distribution: initiate a multiple output transaction to the miners using the mining pool's holding address, but include another holding address in the outputs for future distribution.

members a priori to the actual mining task. Lazy miners could retain partial computational power in the actual task to receive a higher reward than deserved. Therefore, A natural improvement is that mining pools can use an interim reward holding address to aggregate all the output from coinbase transactions before distributing them among the pool members (Fig. 10 b), e.g., F2Poll in 2014 [127]. Rewards to pool members can also be distributed in batches. For example, the mining pool ViaBTC divides the reward into multiple holding addresses and uses them to initiate multiple output transactions to the pool members (Fig. 10 c). Another possibility is to initiate a multiple output transaction to the miners using the mining pool's holding address but include another holding address in the outputs. Hence, the mining pool can further distribute mining rewards and eventually form a chain-like distribution pattern (Fig. 10 d). Note that the output holding address in each chain transaction can be the input address (as used in BTC.com) or another new address (as used in AntPool) [128].

The competition between mining pools became severe over the years. PoW-based coin mining is a process of looking for a random number whose hashed value falls into a specific range. Some mining pools simultaneously mined several blockchain ledgers with the same design, e.g., Bitcoin, Litecoin, Namecoin, Dogecoin, Huntercoin, and Myriadcoin [129]. However, such "merged mining" has operated at the edge of, and even beyond, the security guarantees offered by the underlying Nakamoto consensus for extended periods. Some other mining pools tried to exploit the cryptocurrency system design for a larger profit. For example, the Ethereum blockchain not only rewards the winner of the mining mechanism but also rewards those who produced new but unused ledger updates (uncle blocks). Werner *et al.* [130] found that during May and July 2018, a swarm of 7500 miners with conspicuously small hash rates orchestrated by a single adversary managed to receive 19% of the total uncle block without competing directly for the primary mining rewards.

Miners might also hop from pool to pool to boost their reward [131]. An analysis of Kano and Slush mining pool members found that hopping miners' medium rewards were three times higher for those stuck to one pool [132]. However, no matter which mining pool they belonged, the miners all transferred the reward quickly into the same set of blockchain addresses, which belonged to exchanges, wallets, or gambling services [133]. The average interval between miners receiving the reward to such transfers shortened from 138 days in 2009 to 1.5 days in 2013 [127].

2) DARKNET MARKET ESCROW

Darknet markets are online trading platforms hosted on covert computer servers, which are only accessible through encrypted networks. Illicit merchandise, such as drugs, weapons, and private data, are the most common commodities on darknet markets. Cryptocurrencies became the major payment medium of darknet markets, attributed to their anonymity nature. Payments between vendors and buyers have two modes. One is that the buyers transfer cryptocurrencies directly into the vendors' addresses. However, vendors and buyers have to make extra effort to build prior trust before making such transfers. In this case, an escrow service is provided by the marketplaces. The escrow service first accepts the buyer's deposit using an interim address and later transfers the value to the vendor once the trade is confirmed.

Conceptually, buyers, sellers, and the escrow service form a star-like subgraph, with the escrow address in the middle and edges pointing from the buyers to the escrow and from the

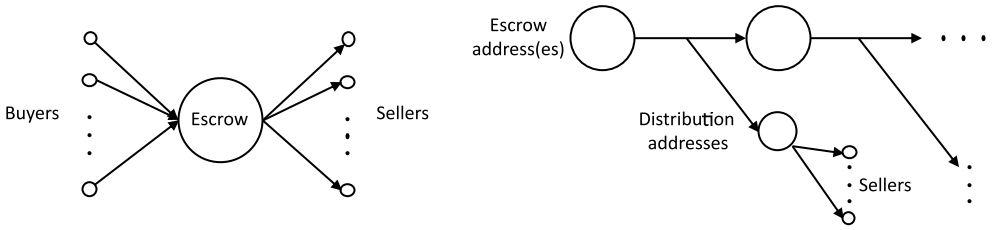

FIGURE 11. Market escrow distribution pattern. (a) Conceptual escrow transactions: the escrow address in the middle receive from buyers and transfer to the sellers. (b) The actual escrow distribution: a major chunk of the escrow is held in the escrow addresses, and payments to the vendors peel off in small amounts from a transaction chain.

escrow to the sellers (Fig. 11 a). The escrow nodes naturally occupy the center of the transaction network. Popular darknet markets, such as SilkRoad, Agora, Wikispeed, and Evolution Market, ranked high in transaction volume, degree, Eigenvector, PageRank [74], [134] and betweenness centrality in the Type III Bitcoin network [75]. A high betweenness showed the diversity of darknet marketplaces' users because the market escrow nodes can connect nodes in different communities. Identifying an escrow address chain helps reveal market sale volume [135] or the sale of goods with particular prices [136].

The tracing of escrow transactions has revealed the operation patterns of the market escrow services. Similar to the situation in mining reward distribution, the escrow address cannot transfer to an unlimited number of seller addresses in a single transaction. Therefore, a tree-like distribution or a chain-like distribution was observed from escrow transaction histories [10]. For example, A major chunk of the escrow is held in the escrow addresses, and payments to the vendors first arrive in distribution addresses by a smaller amount and are further forwarded to the sellers (see Fig. 11 b).

3) MIXING SERIVCES

Swapping-based mixing and CoinJoin-based mixing have distinct features in the transaction records. Swapping services have high daily transaction numbers [137] and transaction volumes [138], short active periods, and low balances [139]. In the short time of activation, swapping addresses tends to act as a transit node, i.e., the middle node of a directed path of length 2, and rarely receives more than one transaction from the same address [139]. Some swapping service providers, such as Darklaunder, repeatedly use a single receiving address in a short time [140], resulting in a larger degree of centrality than normal nodes [138]. Moreover, since users who belong to different communities may use the same mixing service, mixing service nodes can also act as bridges to nodes with few connections before mixings [141].

By contrast, CoinJoin-based transactions resemble ordinary multiple-input-multiple-output transactions. The average number of input addresses of Blockch-ain.info's SharedCoin mixing service transactions was 14.5 (between 4 and 40), and the average number of output addresses was 25.8 (between 4 and 42) [142]. Structurally, these transactions can be well disguised among regular transactions and lower the precision of address association rules in de-anonymizing Bitcoin addresses. Yanovich *et al.* [143] estimated that the volume of CoinJoin transactions was about 2.5% of all Bitcoin transactions in June 2016.

A mixing service can also counter the taint analysis by making withdraw transaction values have a lower variant than normal transactions, i.e., creating similar outputs to prevent mixing re-identification [138], [144]. However, doing so also makes the mixing transactions more recognizable in the transaction network. Phetsouvanh *et al.* [73] proposed an entropy measure to detect CoinJoin transactions in Bitcoin transaction history. For a multiple-input-multiple-output transaction t, its normalized input and output entropies

$$H_{\text{in,norm}(t)} = \frac{H_{\text{in}}(t)}{\log_2(k_{\text{in}}(t))}, \tag{5}$$

$$H_{\text{out,norm}(t)} = \frac{H_{\text{out}}(t)}{\log_2(k_{\text{out}}(t))}, \tag{6}$$

where $k(t)$ is the degree of t, and $H_{\text{in}}(t)$ is the absolute entropy of t, i.e.,

$$H_{\text{in}}(t) = -\sum_{e \in E_{t,\text{in}}} p_{e,\text{in}} \log_2 p_{e,\text{in}}, \tag{7}$$

$$H_{\text{out}}(t) = -\sum_{e \in E_{t,\text{out}}} p_{e,\text{out}} \log_2 p_{e,\text{out}}, \tag{8}$$

where $p_{e,\text{in}}$ and $p_{e,\text{out}}$ are the values transferred in and out, respectively, through e, which is an incoming (outgoing) edge of t, normalized by the total number of Bitcoin inputs (outputs). A high entropy corresponds to a uniform distribution of bitcoin amounts in inputs (outputs) and indicates a potential mixing service.

4) EXCHANGES AND WALLETS

Since the launch of BitcoinMarket.com on March 17, 2010 (now defunct), cryptocurrency exchange activities have occupied most of Bitcoin's blockchain space. The title of the largest exchange changed hands several times, e.g., Mt. Gox, Poloniex, and Binance. Overall, it was estimated that 18% of the addresses belonged to exchanges in Bitcoin transactions from 2009 to 2015 [72]. Network nodes that belong to popular exchanges are highly ranked in degree [75], [85], PageRank, betweenness, and closeness centralities in the transaction networks [12], [126], [134].

Exchange nodes have particular patterns of network connections, too. It was found that in the Type I Bitcoin network, the middle node of a directed path of length two that both starts and ends at an exchange address are highly likely to be another exchange address, indicating that many inter-exchange Bitcoin transactions occur [72]. However, those exchange addresses may not belong to the same exchange [145].

Online wallets are like banks, accepting numerous deposits from users and initiating a large amount of withdraws to user addresses. Wallet addresses ranked among the highest in degree, PageRank, and closeness centralities in transaction networks [75], [134].

5) GAMES AND GAMBLINGS

Blockchain technology perfectly suits the need for casino games with a transparent game logic, which can provide users with an additional source of trust in the games than in traditional online casino games. Therefore, casino games are the most popular applications on the blockchain.

SatoshiDICE, a simple fortune redistribution game, was the most popular gambling game during 2012 and 2013, generating more than half of the transaction volume in the Bitcoin network [10]. Gambling transactions commonly have identical volumes, e.g., USD 1 or 0.01 BTC, and consecutive transactions between gambling services and players usually have short time intervals and high-intensity [137], [146]. Transaction network nodes belonging to SatoshiDICE have a very high degree of centrality in Type III Bitcoin network [75], [134]. Other gambling services such as BTC Dice, BTCLucky, Clone Dice, and DiceOnCrac also have very high degrees of centrality in the Type III Bitcoin network [10].

Smart contract-based games have gained popularity in recent years. As the most popular smart contract-based game, cryptokitties occupied caused a large scale Ethereum traffic congestion and raised the volume of pending transactions in the blockchain network from less than 1 000 to more than 10 000 at the end of 2017 [147].

6) MARKET MANIPULATION BY WHALE ADDRESSES

The cryptocurrency market is notorious for its high fluctuation, though existing research has not been fully conclusive on its efficiency. Urquhart [17] reported that the Bitcoin market is inefficient from 2010 to 2016, but with a trend of being efficient towards the end of this period. Further research confirmed that the Bitcoin market efficiency has been increasing since 2014 [148]. Meanwhile, specific inefficient periods were identified during April to August 2013 and August to November 2016 [149]. Nadarajah and Chu [150] further found that though Bitcoin returns were concluded to be inefficient, the power transformation of the Bitcoin returns can be considered informational efficient. In addition to the Bitcoin market, a recent study showed that most of the hundreds of cryptocurrencies exhibit high informational efficiency [151].

Nonetheless, it is commonly believed that cryptocurrencies have no price baseline and that the volatility of the prices is mainly driven by the supply and demand of cryptocurrencies in the market [152]. Therefore, the change of majority in the numbers of buyers and sellers may lead to price changes; that is, if more cryptocurrencies are for sale, the prices drop, and vice versa [153]. Empirically, cryptocurrency market prices were also found to be interactive with various assets, including gold [154], the stock markets [155]–[157], and among different cryptocurrencies [158]–[160]. Prices were also found to be driven by social media [161]–[163] and government regulations [164].

It is believed that some users who have preliminary information in the cryptocurrency economy decide the market's future movement [17], [165]. Ante [166] traced hundreds of large Bitcoins transactions between 2018 and 2019 and found a positive abnormal trading volume for the 15-minute window before these transactions. Kondor *et al.* [13] constructed daily Type III Bitcoin networks among the long-living nodes and most active nodes during 2012 and 2013; they found that the singular vector weights of Principal Component Analysis (PCA) analysis can explain the fluctuation of Bitcoin prices to a large extent, e.g., 0.85 Pearson's correlation. Akcora *et al.* [14] constructed a combination of Type I and Type II Bitcoins, i.e., using both addresses and transactions as nodes in the transaction network. They considered each transaction a ''chainlet,'' which consisted of a central node, i.e., the transaction, and several incoming and outgoing edges connected to addresses. The number of particular chainlets, e.g., the number of input and output addresses > 20, showed Bitcoin prices' predictive power.

A cryptocurrency exchange can also deliberately manipulate the supply and demand inside itself. The leaked Mt. Gox internal transaction records showed that one account bought USD 112 million worth of Bitcoins in a short 60-day window during September and November 2013 and caused the Bitcoin price to surge from lower than USD 200 to higher than USD 1 000 in two months [28]. The former managerial personnel later confirmed that the exchange operated this account. Chen *et al.* [167] found that the transaction paths between the abnormal accounts that traded with a significantly higher or lower price than the market price formed many closed circles, i.e., self-loop, reciprocal edges, triangles, and polygons. These patterns resulted in a high face transaction volume, yet little actually changed hands. These patterns were also evidence of market manipulation. Among the cryptocurrencies, the ones with small market capitalization, a low traction volume, and trading in fewer exchanges were more prone to market manipulation [168].

7) PONZI SCHEME

Early Ponzi schemes in the cryptocurrency economy used Bitcoin as an investment target. Later, the invention of smart contracts boosted the schemes. Since the HYIPs were written in codes and investment yields are automatically distributed to the investors, investors would establish blind trust in these

TABLE 1. A summary of binary supervised learning tasks.

Research	Cryptocurrency	Labels (sample sizes)	Span of transactions	Features	Outperforming algorithm	Performance	Class imbalance solution
[72]	Bitcoin	Exchange addresses (2.4 million), non-exchange (72.7 million) *	Sep 2011 to Apr 2015	Volume, **network**	RF	>0.99 F1	Undersampling
[173]	Bitcoin	Ponzi scheme addresses (32), non-Ponzi (6,400)	All transactions of the sample addresses	**Volume**, temporal	RF	31 out of 32, 1% FPR	Cost-sensitive learning
[170]	Bitcoin	Ponzi scheme (2,026 addresses, 1,813 entities), non-Ponzi (26,967 addresses, 955 entities)	Jan 2009 to Feb 2017	**Volume**, temporal, **network**	RF	91% TPR, 10% FPR for addresses, 95% TPR, 4.9% FPR for entities	Undersampling
[120]	Bitcoin	Illicit transactions (4,545), licit (42,019)	Not disclosed	**Volume**, embedding	EvolveGCN	0.97 F1	Cost-sensitive learning
[174]	Bitcoin	Illicit transactions (956,000), licit (800,000) *	Not disclosed	**Volume**	RF	>0.90 F1	-
[138]	Bitcoin	Mixing transactions (7,461,895), regular (37,907,769) *	Late 2014	Embedding	Adaboost	0.94 F1	-
[175]	Ethereum	Fraudulent accounts (2,179), non-fraudulent (2,502)	All transactions (fraudulent), blocks 3,800,000 to 3,805,000 (non-fraudulent)	**Volume**, **temporal**	XGBoost	0.99 AUC	-
[176]	Ethereum	Fraudulent accounts (2,200), non-fraudulent (349,999)	All transactions (fraudulent), Jul 2015 to May 2019 (non-fraudulent)	**Volume**, temporal	RF	0.02% FPR at 0.37 F1	-
[177]	Ethereum	Ponzi scheme contracts (200), non-Ponzi (3,580)	Jul 2015 to Sep 2017	Contract code	RF	0.79 F1	-
[178]	Ethereum	Ponzi scheme contracts (172), non-Ponzi (3,203)	Aug 2015 to Aug 2017 (Ponzi), Feb 2016 to Jun 2018 (non-Ponzi)	Volume, temporal, contract code	J48	0.97 F1	-
[118]	Ethereum	Phishing accounts (1,259), non-Phishing (1,259)	Jul 2015 to Mar 2019	Embedding	One-class SVM	0.91 F1	-

F1: F1 score; TPR: true positive rate; FPR: false positive rate; AUC: area under curve.
Not all the samples are used in the classification tasks. They are further sampled randomly or with truncated transaction records. The resulting training and testing sample sizes may be significantly smaller than the reported numbers.

seemingly transparent investment programs and ignore the possibility of dealing with a Ponzi scheme [169].

Ponzi scheme addresses typically have lower transaction values, e.g., between USD 0.01 and 0.1; higher transaction frequencies than regular addresses [137], [170]; and shorter average time span (37 days, as of 2014) than traditional offline Ponzi schemes [169], [171]. However, frequent interactions between the scheme organizers and end-users could prolong the active time [172]. The Gini coefficient of yield distribution and the proportion of incoming transactions could also help distinguish Ponzi scheme addresses from normal addresses [169], [173].

D. LEARNING TASKS USING TRANSACTION FEATURES

1) ADDRESS IDENTITY INFERENCE

Address identity inference is a popular research topic in cryptocurrency transaction analysis. This task, with transaction tracing, is also collectively known as the de-anonymization of addresses. With address tags broadly available, supervised machine learning algorithms are applied to capture the difference between labeled samples with the transaction features mentioned above. Some works consider the identification of entities, which are a cluster of addresses associated with heuristic rules, such as multiple-input.

Considering that the extracted features can be both numerical and categorical, the most commonly used classification algorithms are decision tree-based algorithms, including plain decision tree (DT), classification and regression trees (CART), ensemble algorithms such as gradient boosting decision trees (GBDT) (e.g., LightGBM, XGBoost), gradient boosting regression trees (GBRT), random forests (RF), and isolation forests (IF). Other commonly used algorithms include logistic regression (LR), naïve Bayes (NB), Bayes network (BN), supporter vector machine (SVM). Learning algorithms that take the transaction network structure into account include graph convolution networks (GCN). However, neural networks generally perform worse than tree-based algorithms in these tasks.

Identity inference tasks can be categorized into binary classification, e.g., mostly to identify whether an address is an illicit one (Table 1), and multi-class classification for generic types of economic agents (Table 2). The most significant feature categories, best learning algorithms, and the best performances are also marked with a bold font. Categories with a lower identification accuracy are marked with bold font in the table.

Binary classification tasks generally achieve very high accuracy, showing that illicit activities are highly separable in their transaction patterns. Comparatively, the accuracy of multiple classification tasks is significantly lower. In particular, exchange nodes are commonly confused with gambling, the marketplace, and Ponzi schemes in Bitcoin networks. Nonetheless, according to the estimation by Sun Yin *et al.* [183], approximately 42% of the Bitcoin addresses belong to cryptocurrency exchanges, 23% to mining pools, 23% to personal wallets, and the rest 12% to scams, ransom, marketplace, and other games and services.

Feature-wise, transaction volume features play essential roles in differentiating entities. Network structure-based features, such as centrality [145], neighbors' identities [72], [183], and motifs [170] are proven to have prediction power. Network embedded feature selection methods, when used alone, can also achieve reasonable performance compared to hand-picked features [118], [138], [179]. As for learning algorithms, decision tree-based methods, especially random forests, achieved the highest performance in most of the tasks. GNN based methods, despite their novelty and popularity in recent literature, did not show superior prediction power to well-established methods.

TABLE 2. A summary of multi-class supervised learning tasks.

Research	Cryptocurrency	Labels (sample sizes)	Span of transactions	Features	Outperforming algorithm	Performance	Class imbalance solution
[179]	Bitcoin	Exchange addresses (144,135), gambling (15,584), service (378,200), and general (11,904,377) *	Nov 2018	Embedding	HDDT+ECOC **	0.91 F1	Undersampling
[180]	Bitcoin	Exchange entities (137), service entities (16), gambling entities (76), mining Pool entities (25), mixer entities (37), and marketplace entities (20)	Blocks 0 to 561,620	Volume, network	Cascading GBDT	>99% accuracy	-
[181]	Bitcoin	Mining pools addresses (89), miners (4,030), mixing services (800), gambling (911), exchanges (1,666) and others (1,312)	Blocks 520,850 to 520,950	**Volume**, temporal, embedding	RF	0.96 F1	-
[137]	Bitcoin	**Exchange/wallet (157 entities, 10,469 addresses)**, faucet offering (61 entities, 340 addresses), **gambling (89 entities, 6,734 addresses)**, **HYIP (956 entities, 2,026 addresses)**, **marketplace escrow (17 entities, 1,900 addresses)**, mining pool (38 entities, 1,645 addresses), mixer (32 entities, 3,199 addresses) *	Jan 2009 to Feb 2017	**Volume**, temporal, network	RF	70% accuracy for addresses, 72% accuracy for entities	-
[182]	Bitcoin	**Exchange addresses (10,466)**, faucet (340), **gambling (6,733)**, HYIP (2,026), **market (1,900)**, mixer (3,199), mining pool (1,644)	Jan 2009 to Jun 2018	**Volume**, temporal, network	LightGBM	0.86 Macro F1, 0.87 Micro F1	Cost-sensitive learning
[145]	Bitcoin	Exchange entities (108), service (68), gambling (65), **mining pool (19)**, and darknet marketplace (12)	Blocks 0 to 514,971	Volume, temporal, **network**	LightGBM	0.91 F1	-
[183]	Bitcoin	Exchange entities (306), hosted-wallet (11), personal-wallet (293), darknet marketplace (46), gambling (102), merchant-services (17), mining-pool (67), mixing (10), ransomware (21), scam (23), stolen-bitcoins (4), others (57)	Not disclosed	Volume, temporal, network	Extra Trees	96% accuracy	Oversampling
[122]	Ethereum	1071 smart contract authors	-	Stylometrics	RF	91% accuracy using source code, 80% accuracy using byte code	-

* Not all the samples are used in the classification tasks. They are further sampled randomly or with truncated transaction records. The resulting training and testing sample sizes may be significantly smaller than the reported numbers.
** HDDT: a binary decision tree for unbalanced data using Hellinger distance as the split criteria; ECOC: Error Correcting Output Codes.

The class imbalance problem is severe in the classification tasks, especially those involving illicit addresses. For example, fraudulent addresses are extremely difficult to obtain, while exchange and wallet addresses are abundant. Common solutions fall into two categories, i.e., sampling-based methods and adapted learning algorithms. Sampling-based methods, including the undersampling of the large class and oversampling of small class (using repeated sampling or sample synthetics such as synthetics minority oversampling (SMOTE)), deal with the class imbalance problem at the data sampling stage. Adopted learning algorithms, such as cost-sensitive learning, minimize the classification error by imposing extra punishment on wrongly classified small-class samples.

2) PRICES PREDICTION

Transaction features can also be used to predict cryptocurrency prices. As summarized in Table 3, two tasks are commonly seen across existing works: predicting the exact price and predicting the price direction, i.e., whether the price will go up or down next. Mean squared error (MSE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) are commonly used as precision indicators. The price prediction across different works may be incomparable as these measures are sensitive to the cryptocurrency price scales.

Transaction volume, mining difficulty, and market information, i.e., past prices, are commonly used as features in the prediction models. Network structural features, such as centralities and motifs [186], can also provide predictive power.

Interestingly, compared to blockchain address identity classification problems, neural network algorithms, such as deep neural networks (DNN), long short-term memory (LSTM), Bayesian neural networks (BNN), Stochastic neural networks (SNN), and Genetic Algorithm-based Selective Neural Network Ensemble (GASEN), outperform tree-based algorithms in the price prediction tasks. One possible explanation is that the price prediction tasks, when considering past market information as features, exhibit strong auto-correlation nature, and therefore, are more suitable to the neural network-based methods.

The prediction results for price directions concentrate around 50–60% (with a few exceptions), which is marginally higher than a random guess but already applicable to real applications. Note that due to the strong fluctuation of cryptocurrency market price in some periods, the predictions using the same algorithm can deviate largely at different times, i.e., [190], [194]. Moreover, when considering the linear correlation between predicted and actual prices, the R^2 can be as high as 0.99, attributed mainly to that the price trend is highly non-stationary.

3) ANOMALY DETECTION

Bot activities, malicious attacks, and rare anomalous activities can also leave marks in the transaction records. However, the number of instances is too small to guarantee good classification performance in supervised learning tasks. Therefore, rule-based or unsupervised learning algorithms are used to identify outlier addresses in search of anomalous activities. A summary of transaction network anomaly detection methods and results can be found in Table 4.

Volume, temporal, and network, especially motif, features [94], [196] are commonly chosen for anomaly detection. The learning algorithms used are k-means, kd-tree,

TABLE 3. A summary of cryptocurrency prices prediction tasks.

Research	Cryptocurrency	Target	Span of dataset	Features	Outperforming Method	Performance
[13]	Bitcoin	Price	2012 to 2013	Network	Linear correlation	0.85 between singular vector and price
[86]	Bitcoin	Price	2011 to 2014	Volume	Linear correlation	0.69–0.85 between different features and price
[184]	Bitcoin	Price directions in the following ten seconds, ten minutes and one day	2009 to 2014	Volume, mining difficulty, market information	LR (daily), RF (10-minute)	99% accuracy (daily), 55% accuracy (10-minutes), low accuracy (10-seconds)
[153]	Bitcoin	Price and hourly direction	Jan 2009 to Apr 2013	Volume (of the most influential nodes)	LR (price), NN (price direction)	lowest MSE (price), 55% accuracy (price direction)
[185]	Bitcoin	Daily price direction	Dec 2015 to Jul 2017	**Volume**, network, market information	RF	74% accuracy
[186]	Bitcoin, Litecoin	Daily Price	2009 to 2018 (Bitcoin), 2011 to 2018 (Litecoin)	Network, market information	RF	Motif features contribute to the prediction
[187]	Bitcoin	Daily price	May 2015 to Jun 2017	Volume, mining difficulty, market information	GASEN	64% accuracy
[188]	Bitcoin	Daily price	Sep 2011 to Aug 2017	Volume, mining difficulty, market information	BNN	lowest RMSE and MAPE
[189]	Bitcoin	Daily price and direction	Nov 2011 to Dec 2018	Volume, mining difficulty, market information	LSTM (price), DNN (price direction)	lowest MAPE (price), 53% accuracy (price direction)
[190]	Bitcoin	Daily price and direction	Aug 2013 to Jul 2016 (Interval 1), Apr 2013 to Apr 2017 (Interval 2)	Volume, mining difficulty, market information	SVM, RNN and k-Means clustering ensemble	lowest MAPE (price, SVM); 62.91% accuracy (price direction, Interval 1, ensemble), 59.45% accuracy (price direction, Interval 2, SVM)
[191]	Bitcoin, Ethereum	Daily price	Apr 2016 to May 2018	Volume, mining difficulty, market information	LR (Bitcoin), GBRT (Ethereum)	$R^2 > 0.99$ (both)
[192]	Bitcoin	Price direction in the following five minutes and one day	Feb 2017 to Feb 2019 (daily), Jul 2017 to Jan 2018 (5-min)	Volume, mining difficulty, market information	LR (daily), LSTM (5-minute)	66% accuracy (daily), 67% accuracy (5-min)
[193]	Bitcoin, Ethereum, Litecoin	Daily price	mid 2017 to late 2019	Volume, temporal, mining difficulty, market information	SNN	lowest MAPE
[194]	Bitcoin	Price and direction for next, seventh, 30th, and 90th days	Apr 2013 to Jul 2016 (Interval 1), Apr 2013 to Apr 2017 (Interval 2), Apr 2013 to Dec 2019 (Interval 3)	Volume, mining difficulty, market information	LSTM	lowest MAE, RMSE, and MAPE (prices), 62–65% accuracy (price direction)
[195]	Bitcoin	Daily price and direction	Jan 2017 to Dec 2017	Volume, network	PDE *	0.82 relative accuracy (price) *, >50% accuracy (price direction)

* Relative accuracy: 1-($|P_{real} - P_{forecast}| / P_{real}$), where P_{real} is the real price, $P_{forecast}$ is the forecasted price; PDE: partial differential equation.

TABLE 4. A summary of unsupervised learning-based anomaly detection.

Research	Cryptocurrency	Span of transactions	Features	Outperforming algorithm	Outcome
[196]	Bitcoin	Jan 2009 to Apr 2013	Volume, network	*kd*-tree	Found 7 out of 30 known fraudsters
[94]	Bitcoin	Jan 2009 to Apr 2013	Volume, network	Local Outlier Factor	Found 1 theft case out of 30 known cases
[197]	Bitcoin	Jan 2009 to May 2013	Volume	RolX	Conjectured mixing service clusters
[198]	Bitcoin	Jan 2009 to Apr 2013	Network	Visual inspection	Hidden communities about hundreds of users
[199]	Bitcoin	Nov 2011 to Aug 2012	Volume	PCA and visual analysis	Detect several remarkable events of Pirate@40's HYIP scheme
[12]	Ethereum	Jul 2015 to Jun 2017	Volume, temporal	Threshold	Malicious attacks detected
[200]	Ripple	Jan 2016 to Jun 2016	Volume, temporal	IF, One-Class SVM, GMM	Gateway accounts, arbitrage bots, and other suspectable accounts

unsupervised SVM, Isolation Forest (IF), Gaussian Mixture Models (GMM), and a Role eXtraction (RolX) algorithm, which factorizes the feature matrix into two non-negative matrices and categorizes the nodes into clusters where smaller ones are considered anomalous [197].

Addresses with anomalous transaction patterns provide very strong predictive power to illicit activities such as the Ponzi schemes, thefts, and malicious attacks on the blockchain or cryptocurrency services. Though not fully disclosed by all the papers, we observe that the recalls of the detection algorithms could be quite low, exhibiting the difficulty in exhausting all the illicit activities in the cryptocurrency economy.

VI. TOOLS FOR ANALYZING AND VISUALIZING TRANSACTIONS

Part of the literature reviewed in the main text posted relevant codes or tools online. This section provides non-exhaustive lists of the analytics and visualization tools, as well as blockchain data analysis websites and services.

A. ETL TOOLS

ETL tools extract information such as scripts, transactions, address balances, smart contract codes, and their current states from the blockchain and feed the information into SQL, Graph, or NoSQL databases. ETL is the essential step before conducting a cryptocurrency transaction network analysis. Some tools ship with additional analytic features. Table 5 provides a list of ETL tools collected from the literature and online.

B. VISUALIZATION TOOLS

Visual analysis is also a powerful tool in network analysis. Table 6 summarizes the visualization tools that can construct transaction networks and provide visual analytic functions.

C. ONLINE INTELLIGENCE PLATFORMS

Online intelligence platforms are websites that provide in-depth blockchain information. Some platforms also allow users to post crowd-sourced knowledge to their databases. These intelligence platforms include Blockchain.info (now called Blockchain.com) [2], Etherscan.io [3], WalletExplorer [108], and BlockCypher [201]. Technology companies such as Chainalysis [110] and Elliptic [109] also provide comprehensive services in cryptocurrency data analytics and malicious activity monitoring.

TABLE 5. List of blockchain data ETL tools.

Tool	Cryptocurrency	Computer language	Database	Additional features
BitIodine [202]	Bitcoin	Python	Neo4j	address clustering and classification
Blockchain2graph [203]	Bitcoin	Java/typescript	Neo4j	Graph query with Cipher language
BlockSci [204]	Bitcoin, Litecoin, Namecoin, Zcash	C++/Python	in memory	MapReduce computation
BTCSpark [205]	Bitcoin	Cython	SQL	
bitcoin-blockchain-parser [206]	Bitcoin	Python	Local files	Supports SegWit
BlockETL [207]	Bitcoin	Java	SQL	
Blockparser [208]	Bitcoin	C++	in memory	Simple blockchain statistics: the closure of an address, all the block rewards and fees, taint analysis, etc.
rusty-blockparser [209]	Bitcoin, Litecoin, Namecoin, Dogecoin, Myriadcoin, Unobtanium	Rust	csv files, MySQL	Simple blockchain statistics: average transactions per block, largest transactions, transaction types, etc.
btctrackr [210]	Bitcoin	C++	MySQL	Address clustering based on multiple inputs rule
BitcoinUses [211]	Bitcoin	Java/Javascript	Hadoop	MapReduce computation
Blockchain-etl [212]	Bitcoin, Litecoin, Ethereum, Zcash, Dash, Dogecoin, Bitcoin Cash	Python	csv files, BigQuery	
BlockchainVis [213]	Bitcoin	N/A	Accumulo	address clustering, mixing identification, and visualization
ether_sql [214]	Ethereum	Python	SQL	
BlockAPI [215]	Bitcoin, Ethereum	Scala	MongoDB, MySQL, PostgreSQL, Fuseki	External data: exchange rates, address tags, protocol identifiers, etc.
EtherQL [216]	Ethereum	Java	MongoDB	Support range query and top-k query
DataEther [217]	Ethereum	N/A	ElasticSearch	Ethereum account balance, transaction tracing, contract analysis
TokenScope [218]	Ethereum	N/A	N/A	Detecting inconsistent token behaviors with regard to ERC20 token standards
TEETHER [219]	Ethereum	N/A	N/A	creating exploits for smart contracts
Erays [220]	Ethereum	N/A	N/A	reverse engineering tool for smart contracts

TABLE 6. List of cryptocurrency transaction network visual analytic tools.

Tool	Cryptocurrency	Computer language	database
BitConeView [221]	Bitcoin	Python	N/A
BitExTract [222]	Bitcoin	Python	MangoDB
BiVA [223]	Bitcoin	Python	Neo4j
SuPoolVisor [224]	Bitcoin	D3.js	N/A
goBlockchainDataAnalysis [225]	Faircoin	Go, NodeJS, AngularJS	MongoDB

VII. OPEN PROBLEMS

Despite the fruitful findings in the existing works, knowledge discovery studies from cryptocurrency transactions can advance in both methodology and the research questions.

A. NETWORK REPRESENTATION LEARNING

To the best of the authors' knowledge, cryptocurrency transaction networks are the largest networks that could be built from public data sources, with evolving network structures and abundant labels on the nodes and edges, providing much information for the knowledge discovery in transaction records. The state-of-the-art methodology in network analysis is network representation learning, which transforms nodes' structural properties into lower-dimensional vectors for downstream machine learning tasks. Although this method has begun to gain its usage in cryptocurrency transaction analysis, it is still out-performed by theory-based structural feature measures, e.g., in simple binary classification tasks [120], so far. Nonetheless, we believe that there is still considerable room for further development and application of network representation learning techniques in cryptocurrency transaction analysis in the future.

B. TRACING ACROSS LEDGERS

From the cryptocurrency economic perspective, transactions across different ledgers are potentially intertwined, as users often change their holdings of one cryptocurrency to another. The linkage between different cryptocurrencies can be studied from several perspectives. For example, the transactions that happened inside cryptocurrency exchanges [226] and on exchange-like blockchain ledgers, such as Ripple [227], can both be utilized further to trace the flow of monetary flow between users regardless of the actual currency used. Recently proposed cryptocurrency blockchain systems, e.g., Polkadot, also provide "interoperability" across different ledgers. In such a design, the transaction records of an amount of cryptocurrency can be "locked" on one blockchain and recorded on another. The study of multiple blockchain ledger integrations is just starting. There is no doubt that current methods can partially solve these questions, but many open questions remain, e.g., how to integrate multiple pieces of transaction histories, trace transactions among different cryptocurrencies, and detect illicit activities from these transactions.

C. NEW TOKENS STANDARDS

More than 300 000 user-customized tokens have been issued on the Ethereum blockchain as of 2020, following the ERC20, ERC721, and ERC777 standards. ERC20 is the initial version of the token standard, ERC777 is an update to ERC20, and ERC721 is fundamentally different. ERC20 and ERC777 tokens are fungible, which means that each token can have a certain amount of distribution, and proportions of the whole amount can circulate among blockchain addresses independently. Comparatively, ERC721 tokens are designed to represent ownership over digital or physical assets. They are non-fungible, meaning that each token is of only one instance and cannot be further divided. Fungible and non-fungible tokens can have a distinct nature of circulation. What are the most distinct characteristics between the two types of networks? What can the distinction tell about user behaviors? Such questions are worth further investigation.

D. PRIMARY MARKET ACTIVITIES

Some user-customized tokens issued on the blockchains are classified as securities by government regulatory bodies [228]. Companies and start-up projects raise funds by selling tokens to institutional and individual investors. This funding method, which bears the names of ICO, IEO, and security token offering (STO), is much like conventional primary market activities. However, because of the lack of regulation to this funding method in most parts of the world, fraud and scams frequently happen and often lead to investors' huge losses. A study of economic agent behavior in the funding process could help reveal the financial misconduct of token issuers and distributors, hence better protecting individual investors' interests.

E. DE-FI

The development of smart contract-based financial instruments and the increasing adoption of blockchain technology in investors have brought a new wave of cryptocurrency innovation — De-Fi. Short for decentralized financing, an umbrella term for the whole spectrum of financial activities over blockchain, De-Fi aims to disrupt and automate the entire financial industry. For example, users can deposit a digital asset into a smart contract as collateral for a loan of another digital asset. Other decentralized financial services, such as trading, lending, investment, asset management, and insurance, are also being developed. With all the transaction and usage history recorded on the blockchain ledgers, a thorough study of human financial behaviors based on an unprecedented rich dataset can be anticipated.

F. DIGITAL FIAT MONEY

The knowledge revealed by the transaction data stored in the blockchain networks not only helps to demonstrate the validity of the claim that the transparency of blockchain could facilitate auditing and regulating user activities, but could also be applied to a broader area beyond blockchain. The year 2020 sees the pilot tests and promotions of digital fiat money, such as Sweden's E-Krona and China's digital currency electronic payment (DC/EP). Although not necessarily adopting blockchain systems, the transaction records of these electronic versions of fiat money will be fully archived, e.g., in a central database. The successful experience of cryptocurrency transaction network analysis can be further borrowed by the electronic fiat money system in the future to help governments fight against bribing, money laundering, and terrorism financing.

VIII. CONCLUSION

Since the invention of Bitcoin in 2008, cryptocurrency has received wide acceptance among millions of users worldwide. A complete trace of users' activities and behaviors has been faithfully recorded on the blockchain. Having begun to notice the richness of the blockchain database a few years ago, academia has since produced a large body of research regarding cryptocurrency transactions.

The most extensively studied cryptocurrencies are Bitcoin, Ethereum, user-customized tokens issued on the Ethereum blockchain, and altcoins that provide extra privacy-preserving functions. Bitcoin is the representative cryptocurrency using a UTXO data model to store their transactions, while the Ethereum blockchain is the representative account-based data model.

Attributed to the transparency of blockchain ledgers, most of the cryptocurrency transactions are traceable and linkable. Although various coin mixing services and privacy-enhancing alternative cryptocurrencies are proposed, Careless use still can largely reveal user identity and behaviors. The tracing of illicit money flows between Ponzi scheme organizers and their victims, thefts, laundering, and ransomware victims have also provided strong evidence for solving these crimes.

From a macroscopic view, cryptocurrency transactions among blockchain accounts form large and complex transaction networks. These networks are continually growing, with new blockchain addresses being created and used. Preferential attachment is the principal law governing the networks' growth: new nodes connect to existing nodes with higher connectivity. The networks eventually evolved into scale-free, e.g., with power-law degree distributions, and small-world networks, e.g., with short average path lengths and high clustering coefficients. The transaction networks can also show a bow-tie structure, with a large strongly connected component and obvious source and sink communities.

Agents in the cryptocurrency economy may have different economic behaviors and therefore form different transaction patterns. For example, mining pools send rewards to pool members in a tree-like or chain-like series of transactions. Marketplaces also use such distribution patterns to conduct their escrow services. Major agents such as cryptocurrency exchanges, online wallets, marketplaces, gambling games, and mixing services were all found to have the highest transaction volumes and possess transaction networks' central

positions. These structural features can be further utilized in machine learning algorithms to derive models that differentiate and identify economic agents in the transaction network.

With the cryptocurrency economy booming in recent years, we can foresee an abundance of new and disruptive innovations, especially blockchain-enabled financial services. We believe that cryptocurrency transactions will continuously provide new knowledge of various human social-economic behaviors in the future.

ACKNOWLEDGMENT

The authors would like to thank the valuable discussions from Prof. Ron G. Chen, Prof. Xiao-Ke Xu, and Prof. Ye Wu.

REFERENCES

[1] *Coinmarketcap*. Accessed: Nov. 1, 2020. [Online]. Available: https://coinmarketcap.com/
[2] *Blockchain Explorer*. Accessed: Nov. 1, 2020. [Online]. Available: https://www.blockchain.com/explorer
[3] *Ethereum Blockchain Explorer*. Accessed: Nov. 1, 2020. [Online]. Available: https://etherscan.io/
[4] F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 1318–1326.
[5] D. Ron and A. Shamir, "How did dread pirate roberts acquire and protect his bitcoin wealth?" in *Proc. 18th Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2014, pp. 3–15.
[6] K. Nilsson. (2017). *Breaking Open the MtGox Case, Part 1*. [Online]. Available: https://blog.wizsec.jp/2017/07/breaking-open-mtgox-1.html
[7] H. Kuzuno and C. Karam, "Blockchain explorer: An analytical process and investigation environment for bitcoin," in *Proc. APWG Symp. Electron. Crime Res. (eCrime)*, 2017, pp. 9–16.
[8] S. Naqvi, "Challenges of cryptocurrencies forensics—A case study of investigating, evidencing and prosecuting organised cybercriminals," in *Proc. 13th Int. Conf. Availability, Rel. Secur. (ARES)*, 2018, pp. 1–5.
[9] A. S. M. Irwin and G. Milad, "The use of crypto-currencies in funding violent jihad," *J. Money Laundering Control*, vol. 19, no. 4, pp. 407–425, Oct. 2016.
[10] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," *Commun. ACM*, vol. 59, no. 4, pp. 86–93, Apr. 2016.
[11] A. Biryukov and S. Tikhomirov, "Deanonymization and linkability of cryptocurrency transactions based on network analysis," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P)*, Jun. 2019, p. 2019.
[12] T. Chen, Y. Zhu, Z. Li, J. Chen, X. Li, X. Luo, X. Lin, and X. Zhange, "Understanding Ethereum via graph analysis," in *Proc. INFOCOM IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 1484–1492.
[13] D. Kondor, I. Csabai, J. Szüle, M. Pósfai, and G. Vattay, "Inferring the interplay between network structure and market effects in bitcoin," *New J. Phys.*, vol. 16, no. 12, Dec. 2014, Art. no. 125003.
[14] C. G. Akcora, M. F. Dixon, Y. R. Gel, and M. Kantarcioglu, "Bitcoin risk modeling with blockchain graphs," *Econ. Lett.*, vol. 173, pp. 138–142, Dec. 2018.
[15] S. Park, S. Im, Y. Seol, and J. Paek, "Nodes in the bitcoin network: Comparative measurement study and survey," *IEEE Access*, vol. 7, pp. 57009–57022, 2019.
[16] J. Mišić, V. B. Mišić, X. Chang, S. G. Motlagh, and M. Z. Ali, "Modeling of bitcoin's blockchain delivery network," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1368–1381, 2019.
[17] A. Urquhart, "The inefficiency of bitcoin," *Econ. Lett.*, vol. 148, pp. 80–82, Nov. 2016.
[18] F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2084–2123, 3rd Quart., 2016.
[19] J. Kolb, M. AbdelBaky, R. H. Katz, and D. E. Culler, "Core concepts, challenges, and future directions in blockchain: A centralized tutorial," *ACM Comput. Surv.*, vol. 53, no. 1, pp. 1–39, May 2020.
[20] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, "Untangling blockchain: A data processing view of blockchain systems," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 7, pp. 1366–1385, Jul. 2018.
[21] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *ACM Comput. Surv.*, vol. 52, no. 3, Jul. 2019.
[22] M. C. Kus Khalilov and A. Levi, "A survey on anonymity and privacy in bitcoin-like digital cash systems," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2543–2585, 3rd Quart., 2018.
[23] M. Conti, E. Sandeep Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3416–3452, 2018.
[24] M. di Angelo and G. Salzer, "A survey of tools for analyzing Ethereum smart contracts," in *Proc. IEEE Int. Conf. Decentralized Appl. Infrastruct. (DAPPCON)*, Apr. 2019, pp. 69–78.
[25] (2017). *Bitcoin Wiki: Laszlo Hanyecz*. [Online]. Available: https://en.bitcoin.it/wiki/Laszlo_Hanyecz
[26] *Ponzi Schemes Using Virtual Currencies*, U.S. Securities, Exchange Commission, Washington, DC, USA, 2013.
[27] V. Buterin. (2014). *Ethereum White Paper: A Next-Generation Smart Contract and Decentralized Application Platform*. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper
[28] N. Gandal, J. Hamrick, T. Moore, and T. Oberman, "Price manipulation in the bitcoin ecosystem," *J. Monetary Econ.*, vol. 95, pp. 86–96, May 2018.
[29] E. H. Aw, R. Gera, K. Hicks, N. Koeppen, and C. Teska, "Analyzing preferential attachment in peer-to-peer BITCOIN networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 1242–1249.
[30] P. D. Meo, "Trust prediction via matrix factorisation," *ACM Trans. Internet Technol.*, vol. 19, no. 4, pp. 1–20, Nov. 2019.
[31] S. Kumar, F. Spezzano, V. S. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 221–230.
[32] C. Dougherty and G. Huang. (2014). *Mt. Gox Seeks Bankruptcy After $480 Million Bitcoin Loss*. [Online]. Available: https://www.bloomberg.com/news/articles/2014-02-28/mt-gox-exchange-files-for-bankruptcy
[33] E. Perez. (2015). *2 Former Federal Agents Charged With Stealing Bitcoin During Silk Road Probe*. [Online]. Available: https://edition.cnn.com/2015/03/30/politics/federal-agents-charged-with-stealing-bitcoin/index.html
[34] M. Conti, A. Gangwal, and S. Ruj, "On the economic significance of ransomware campaigns: A bitcoin transactions perspective," *Comput. Secur.*, vol. 79, pp. 162–189, Nov. 2018.
[35] M. Paquet-Clouston, B. Haslhofer, and B. Dupont, "Ransomware payments in the bitcoin ecosystem," *J. Cybersecurity*, vol. 5, no. 1, p. tyz003, Jan. 2019.
[36] R. S. Portnoff, D. Y. Huang, P. Doerfler, S. Afroz, and D. McCoy, "Backpage and bitcoin: Uncovering human traffickers," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Min. (KDD)*, 2017, pp. 1595–1604.
[37] M. Paquet-Clouston, M. Romiti, B. Haslhofer, and T. Charvat, "Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem," in *Proc. 1st ACM Conf. Adv. Financial Technol.*, Oct. 2019, pp. 76–88.
[38] *Tornado.Cash*. Accessed: Nov. 1, 2020. [Online]. Available: https://tornado.cash/
[39] *Position Paper: Regulation of Virtual Asset Trading Platforms*, Hong Kong Securities, Futures Commission, Hong Kong, 2019.
[40] P. Peterson. (2017). *Transaction Linkability*. [Online]. Available: https://electriccoin.co/blog/transaction-linkability/
[41] A. Mackenzie, S. Noether, and Monero Core Team, "Improving obfuscation in the cryptonote protocol," Monero Res. Lab, Tech. Rep. MRL-0004, 2015. [Online]. Available: https://cryptochainuni.com/wp-content/uploads/Monero-Improving-Obfuscation-in-the-CryptoNote-Protocol.pdf
[42] J. D. Nick, "Data-driven de-anonymization in bitcoin," M.S. thesis, ETH-Zürich, Zürich, Switzerland, 2015, doi: 10.3929/ethz-a-010541254.
[43] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in bitcoin," in *Proc. 17th Int. Conf. Financial Cryptogr. Data Secur.* Berlin, Germany: Springer, 2013, pp. 34–51.
[44] C. Zhao and Y. Guan, "A graph-based investigation of bitcoin transactions," in *Proc. 11th IFIP Int. Conf. Digit. Forensics*. Cham, Switzerland: Springer, 2015, pp. 79–95.
[45] T.-H. Chang and D. Svetinovic, "Improving bitcoin ownership identification using transaction patterns analysis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 1, pp. 9–20, Jan. 2020.
[46] Z. Yu, M. H. Au, J. Yu, R. Yang, Q. Xu, and W. F. Lau, "New empirical traceability analysis of cryptonote-style blockchains," in *Proc. 23rd Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Cham, Switzerland: Springer, 2019, pp. 133–149.
[47] J. Quesnelle, "On the linkability of Zcash transactions," 2017, *arXiv:1712.01210*. [Online]. Available: http://arxiv.org/abs/1712.01210

[48] L. Chen, L. Xu, N. Shah, N. Diallo, Z. M. Gao, Y. Lu, and W. D. Shi, "Unraveling blockchain based crypto-currency system supporting oblivious transactions: A formalized approach," in *Proc. ACM Workshop Blockchain, Cryptocurrencies Contract*, 2017, pp. 23–28.
[49] M. Moser, R. Bohme, and D. Breuker, "An inquiry into money laundering tools in the bitcoin ecosystem," in *Proc. APWG eCrime Researchers Summit (eCRS)*, Sep. 2013, pp. 1–14.
[50] Y. Hong, H. Kwon, J. Lee, and J. Hur, "A practical de-mixing algorithm for bitcoin mixing services," in *Proc. 2nd ACM Workshop Blockchains, Cryptocurrencies, Contracts (BCC)*, 2018, pp. 15–20.
[51] A. Biryukov, D. Feher, and G. Vitto, "Privacy aspects and subliminal channels in zcash," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, Nov. 2019, pp. 1813–1830.
[52] D. S. Johnson, "Approximation algorithms for combinatorial problems," *J. Comput. Syst. Sci.*, vol. 9, no. 3, pp. 256–278, 1974.
[53] O. H. Ibarra and C. E. Kim, "Fast approximation algorithms for the knapsack and sum of subset problems," *J. ACM*, vol. 22, no. 4, pp. 463–468, Oct. 1975.
[54] C. Pérez-Solà, S. Delgado-Segura, G. Navarro-Arribas, and J. Herrera-Joancomartí, "Another coin bites the dust: An analysis of dust in UTXO-based cryptocurrencies," *Roy. Soc. Open Sci.*, vol. 6, no. 1, Jan. 2019, Art. no. 180817.
[55] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *Proc. 17th Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2013, pp. 6–24.
[56] K. Liao, Z. Zhao, A. Doupe, and G.-J. Ahn, "Behind closed doors: Measurement and analysis of CryptoLocker ransoms in bitcoin," in *Proc. APWG Symp. Electron. Crime Res. (eCrime)*, Jun. 2016, pp. 1–13.
[57] M. Harrigan and C. Fretter, "The unreasonable effectiveness of address clustering," in *Proc. Int. IEEE Conferences Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People, Smart World Congr. (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, Jul. 2016, pp. 368–373.
[58] D. Ermilov, M. Panov, and Y. Yanovich, "Automatic bitcoin address clustering," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 461–466.
[59] Z. Zhang, W. Li, H. Liu, and J. Liu, "A refined analysis of Zcash anonymity," *IEEE Access*, vol. 8, pp. 31845–31853, 2020.
[60] A. Biryukov and D. Feher, "Privacy and linkability of mining in Zcash," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2019, pp. 118–123.
[61] G. Kappos, H. Yousaf, M. Maller, and S. Meiklejohn, "An empirical analysis of anonymity in Zcash," in *Proc. 27th USENIX Secur. Symp.*, 2018, pp. 463–477.
[62] M. Möser, K. Soska, E. Heilman, K. Lee, H. Heffan, S. Srivastava, K. Hogan, J. Hennessey, A. Miller, A. Narayanan, and N. Christin, "An empirical analysis of traceability in the Monero blockchain," in *Proc. 18th Privacy Enhancing Technol. (PETS)*, 2018, vol. 2018, no. 3, pp. 143–163.
[63] A. Kumar, C. Fischer, S. Tople, and P. Saxena, "A traceability analysis of Monero's blockchain," in *Proc. 22nd Eur. Symp. Res. Comput. Secur.* Cham, Switzerland: Springer, 2017, pp. 153–173.
[64] S. Johnson, *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. Uttar Pradesh, India: Simon and Schuster, 2002.
[65] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schutte, and M. Perc, "Saving human lives: What complexity science and information systems can contribute," *J. Stat. Phys.*, vol. 158, no. 3, pp. 735–781, Feb. 2015.
[66] M. Newman, *Networks*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2018.
[67] B. Holtz, J. Fortuna, and J. Neff. (2013). *Evolutionary Structural Analysis of the Bitcoin Network*. [Online]. Available: http://snap.stanford.edu/class/cs224w-2013/projects2013/cs224w-029-final.pdf
[68] J. Liang, L. Li, and D. Zeng, "Evolutionary dynamics of cryptocurrency transaction networks: An empirical study," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0202202.
[69] M. K. Popuri and M. H. Gunes, "Empirical analysis of crypto currencies," in *Proc. 7th Workshop Complex Netw.* Berlin, Germany: Springer, 2016, pp. 281–292.
[70] B. B. F. Pontiveros, M. Steichen, and R. State, "Mint centrality: A centrality measure for the bitcoin transaction graph," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2019, pp. 159–162.
[71] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu, "Forecasting bitcoin price with graph chainlets," in *Proc. 22nd Pacific–Asia Conf. Adv. Knowl. Discov. Data Min.* Cham, Switzerland: Springer, 2018, pp. 765–776.
[72] S. Ranshous, C. A. Joslyn, S. Kreyling, K. Nowak, N. F. Samatova, C. L. West, and S. Winters, "Exchange pattern mining in the bitcoin transaction directed hypergraph," in *Proc. 21st Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Cham, Switzerland: Springer, 2017, pp. 248–263.
[73] S. Phetsouvanh, A. Datta, and F. Oggier, "Analysis of multi-input multi-output transactions in the bitcoin network," *Concurrency Comput. Pract. Exper.*, vol. 33, no. 1, p. e5629, Dec. 2019.
[74] D. Di Francesco Maesa, A. Marino, and L. Ricci, "Data-driven analysis of bitcoin properties: Exploiting the users graph," *Int. J. Data Sci. Analytics*, vol. 6, no. 1, pp. 63–80, Aug. 2018.
[75] M. Lischke and B. Fabian, "Analyzing the bitcoin network: The first four years," *Future Internet*, vol. 8, no. 4, p. 7, Mar. 2016.
[76] D. Guo, J. Dong, and K. Wang, "Graph structure and statistical properties of Ethereum transaction relationships," *Inf. Sci.*, vol. 492, pp. 58–71, Aug. 2019.
[77] S. Somin, G. Gordon, and Y. Altshuler, "Network analysis of ERC20 tokens trading on Ethereum blockchain," in *Proc. 9th Int. Conf. Complex Syst. (ICCS)*. Cham, Switzerland: Springer, 2018, pp. 439–450.
[78] F. Victor and B. K. Lüders, "Measuring Ethereum-based ERC20 token networks," in *Proc. 23rd Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Cham, Switzerland: Springer, 2019, pp. 113–129.
[79] M. Nowostawski and J. Tøn, "Evaluating methods for the identification of off-chain transactions in the lightning network," *Appl. Sci.*, vol. 9, no. 12, p. 2519, Jun. 2019.
[80] M. Conoscenti, A. Vetro, and J. C. De Martin, "Hubs, rebalancing and service providers in the lightning network," *IEEE Access*, vol. 7, pp. 132828–132840, 2019.
[81] Y. Guo, J. Tong, and C. Feng, "A measurement study of bitcoin lightning network," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 202–211.
[82] *Glassnode Studio*. [Online]. Available: https://studio.glassnode.com/metrics?a=BTC&m=addresses.Count
[83] 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, Feb. 2014, Art. no. e86197.
[84] A. Aspembitova, L. Feng, V. Melnikov, and L. Y. Chew, "Fitness preferential attachment as a driving mechanism in bitcoin transaction network," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0219346.
[85] A. Baumann, B. Fabian, and M. Lischke, "Exploring the bitcoin network," in *Proc. 10th Int. Conf. Web Inf. Syst. Technol. (WEBIST)*, 2014, pp. 369–374.
[86] M. Sorgente and C. Cibils. (2014). *The Reaction of a Network: Exploring the Relationship Between the Bitcoin Network Structure and the Bitcoin Price*. [Online]. Available: http://snap.stanford.edu/class/cs224w-2014/projects/cs224w-27-final.pdf
[87] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[88] S. Somin, Y. Altshuler, G. Gordon, A. Pentland, and E. Shmueli, "Network dynamics of a financial ecosystem," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Dec. 2020.
[89] A. H. Jun Ren, L. Feng, S. A. Cheong, and R. S. Mong Goh, "Optimal fee structure for efficient lightning networks," in *Proc. IEEE 24th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2018, pp. 980–985.
[90] J.-H. Lin, K. Primicerio, T. Squartini, C. Decker, and C. J. Tessone, "Lightning network: A second path towards centralisation of the bitcoin economy," *New J. Phys.*, vol. 22, no. 8, Aug. 2020, Art. no. 083022.
[91] S. Lee and H. Kim, "On the robustness of lightning network in bitcoin," *Pervas. Mobile Comput.*, vol. 61, Jan. 2020, Art. no. 101108.
[92] S. Martinazzi and A. Flori, "The evolving topology of the lightning network: Centralization, efficiency, robustness, synchronization, and anonymity," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0225966.
[93] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Trans. Knowl. Discovery from Data*, vol. 1, no. 1, p. 2, Mar. 2007.
[94] T. Pham and S. Lee, "Anomaly detection in the bitcoin system—A network perspective," 2016, *arXiv:1611.03942*. [Online]. Available: http://arxiv.org/abs/1611.03942
[95] M. Zwang, S. Somin, A. Pentland, and Y. Altshuler, "Detecting Bot activity in the Ethereum blockchain network," 2018, *arXiv:1810.01591*. [Online]. Available: http://arxiv.org/abs/1810.01591
[96] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD)*, 2005, pp. 177–187.
[97] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[98] S. Ferretti and G. D'Angelo, "On the Ethereum blockchain structure: A complex networks theory perspective," *Concurrency Comput. Pract. Exper.*, vol. 32, no. 12, p. e5493, Jun. 2020.
[99] B. Zheng, L. Zhu, M. Shen, X. Du, and M. Guizani, "Identifying the vulnerabilities of bitcoin anonymous mechanism based on address clustering," *Sci. China Inf. Sci.*, vol. 63, no. 3, pp. 1–5, Mar. 2020.
[100] C. Remy, B. Rym, and L. Matthieu, "Tracking bitcoin users activity using community detection on a network of weak signals," in *Proc. 6th Int. Conf. Complex Netw. Appl.* Cham, Switzerland: Springer, 2018, pp. 166–177.
[101] G. A. Oliva, A. E. Hassan, and Z. M. Jiang, "An exploratory study of smart contracts in the Ethereum blockchain platform," *Empirical Softw. Eng.*, vol. 25, no. 3, pp. 1864–1904, May 2020.
[102] L. Kiffer, D. Levin, and A. Mislove, "Analyzing Ethereum's contract topology," in *Proc. Internet Meas. Conf. (IMC)*, Oct. 2018, pp. 494–499.
[103] A. Pinna, S. Ibba, G. Baralla, R. Tonelli, and M. Marchesi, "A massive analysis of Ethereum smart contracts empirical study and code metrics," *IEEE Access*, vol. 7, pp. 78194–78213, 2019.
[104] N. Ajienka, P. Vangorp, and A. Capiluppi, "An empirical analysis of source code metrics and smart contract resource consumption," *J. Softw. Evol. Process*, vol. 32, no. 10, p. e2267, Oct. 2020.
[105] G. A. Pierro and H. Rocha, "The influence factors on Ethereum transaction fees," in *Proc. IEEE/ACM 2nd Int. Workshop Emerg. Trends Softw. Eng. Blockchain (WETSEB)*, May 2019, pp. 24–31.
[106] K. Toyoda, T. Ohtsuki, and P. T. Mathiopoulos, "Identification of high yielding investment programs in bitcoin via transactions pattern analysis," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
[107] S. F. Dyson, W. J. Buchanan, and L. Bell, "Scenario-based creation and digital investigation of Ethereum ERC20 tokens," *Forensic Sci. Int. Digit. Invest.*, vol. 32, Mar. 2020, Art. no. 200894.
[108] *Walletexplorer*. Accessed: Nov. 1, 2020. [Online]. Available: https://www.walletexplorer.com/
[109] *Elliptic*. Accessed: Nov. 1, 2020. [Online]. Available: https://www.elliptic.co/
[110] *Chainalysis*. Accessed: Nov. 1, 2020. [Online]. Available: https://www.chainalysis.com/
[111] M. A. Harlev, H. Sun Yin, K. C. Langenheldt, R. Mukkamala, and R. Vatrapu, "Breaking bad: De-anonymising entity types on the bitcoin blockchain using supervised machine learning," in *Proc. 51st Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2018, pp. 3497–3506.
[112] M. Jalili and M. Perc, "Information cascades in complex networks," *J. Complex Netw.*, vol. 5, no. 5, pp. 665–693, 2017.
[113] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Phys. Rep.*, vol. 689, pp. 1–54, May 2017.
[114] F. Oggier, S. Phetsouvanh, and A. Datta, "Entropic centrality for non-atomic flow networks," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, 2018, pp. 50–54.
[115] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, Oct. 2002.
[116] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD)*, 2014, pp. 701–710.
[117] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, 2016, pp. 855–864.
[118] J. Wu, Q. Yuan, D. Lin, W. You, W. Chen, C. Chen, and Z. Zheng, "Who are the phishers? Phishing scam detection on Ethereum via network embedding," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Sep. 2, 2020, doi: 10.1109/TSMC.2020.3016821.
[119] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Mar. 2020.
[120] M. Weber, G. Domeniconi, J. Chen, D. Karl I. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson, "Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics," 2019, *arXiv:1908.02591*. [Online]. Available: http://arxiv.org/abs/1908.02591
[121] R. Norvill, B. B. Fiz Pontiveros, R. State, I. Awan, and A. Cullen, "Automated labeling of unknown contracts in Ethereum," in *Proc. 26th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2017, pp. 1–6.
[122] S. Linoy, N. Stakhanova, and S. Ray, "De-anonymizing Ethereum blockchain smart contracts through code attribution," *Int. J. Netw. Manag.*, vol. 31, no. 1, p. e2130, 2021.
[123] C. F. Torres, M. Steichen, and R. State, "The art of the scam: Demystifying honeypots in Ethereum smart contracts," in *Proc. 28th USENIX Secur. Symp.*, 2019, pp. 1591–1607.
[124] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, "Bitcoin mining pools: A cooperative game theoretic analysis," in *Proc. 14th Int. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 2015, pp. 919–927.
[125] L. Ren and P. A. S. Ward, "Pooled mining is driving blockchains toward centralized systems," in *Proc. 38th Int. Symp. Reliable Distrib. Syst. Workshops (SRDSW)*, Oct. 2019, pp. 43–48.
[126] A. Anoaica and H. Levard, "Quantitative description of internal activity on the Ethereum public blockchain," in *Proc. 9th IFIP Int. Conf. New Technol., Mobility Secur. (NTMS)*, Feb. 2018, pp. 1–5.
[127] L. Wang and Y. Liu, "Exploring miner evolution in bitcoin network," in *Proc. 16th Int. Conf. Passive Act. Meas. (PAM)*. Berlin, Germany: Springer, 2015, pp. 290–302.
[128] M. Romiti, A. Judmayer, A. Zamyatin, and B. Haslhofer, "A deep dive into bitcoin mining pools: An empirical analysis of mining shares," 2019, *arXiv:1905.05999*. [Online]. Available: http://arxiv.org/abs/1905.05999
[129] A. Judmayer, A. Zamyatin, N. Stifter, A. G. Voyiatzis, and E. Weippl, "Merged mining: Curse or cure?" in *Proc. Int. Workshop. Data Privacy Manage, Cryptocurrencies, Blockchain Technol.*, vol. 10436, 2017, pp. 316–333.
[130] S. M. Werner, P. J. Pritz, A. Zamyatin, and W. J. Knottenbelt, "Uncle traps: Harvesting rewards in a queue-based Ethereum mining pool," in *Proc. 12th EAI Int. Conf. Perform. Eval. Methodologies Tools*, 2019, pp. 127–134.
[131] M. Rosenfeld, "Analysis of bitcoin pooled mining reward systems," 2011, *arXiv:1112.4980*. [Online]. Available: https://arxiv.org/abs/1112.4980
[132] M. Belotti, S. Kirati, and S. Secci, "Bitcoin pool-hopping detection," in *Proc. 4th IEEE Int. Forum Res. Technol. Soc. Ind. (RTSI)*, 2018, pp. 1–6.
[133] Z.-X. Lin and X. F. Liu, "Tracking the circulation routes of fresh coins in bitcoin: A way to identify coinminers based on transaction network structural properties," *J. Nanjing. Univ. Inf. Sci. Technol.*, vol. 10, no. 4, pp. 450–455, Apr. 2018.
[134] M. Fleder, M. S. Kester, and S. Pillai, "Bitcoin transaction graph analysis," 2015, *arXiv:1502.01657*. [Online]. Available: http://arxiv.org/abs/1502.01657
[135] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proc. 22nd Int. Conf. World Wide Web. (WWW)*, 2013, pp. 213–224.
[136] Y. Yannikos, A. Schäfer, and M. Steinebach, "Monitoring product sales in darknet shops," in *Proc. 13th Int. Conf. Availability, Rel. Secur. (ARES)*, Aug. 2018, pp. 1–7.
[137] K. Toyoda, T. Ohtsuki, and P. T. Mathiopoulos, "Multi-class bitcoin-enabled service identification based on transaction history summarization," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2018, pp. 1153–1160.
[138] Y. Hu, S. Seneviratne, K. Thilakarathna, K. Fukuda, and A. Seneviratne, "Characterizing and detecting money laundering activities on the bitcoin network," 2019, *arXiv:1912.12060*. [Online]. Available: http://arxiv.org/abs/1912.12060
[139] J. Wu, J. Liu, W. Chen, H. Huang, Z. Zheng, and Y. Zhang, "Detecting mixing services via mining bitcoin transaction network with hybrid motifs," 2020, *arXiv:2001.05233*. [Online]. Available: http://arxiv.org/abs/2001.05233
[140] T. de Balthasar and J. Hernandez-Castro, "An analysis of bitcoin laundry services," in *Proc. 22nd Nordic Conf. Secur. IT Syst. (NordSec)*. Cham, Switzerland: Springer, 2017, pp. 297–312.
[141] L. Nan and D. Tao, "Bitcoin mixing detection using deep autoencoder," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2018, pp. 280–287.
[142] S. Meiklejohn and C. Orlandi, "Privacy-enhancing overlays in bitcoin," in *Proc. 19th Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2015, pp. 127–141.
[143] Y. Yanovich, P. Mischenko, and A. Ostrovskiy. (2016). *Shared Send Untangling in Bitcoin*. [Online]. Available: http://cryptochainuni.com/wp-content/ploads/bitfury-whitepaper-shared-send-untangling-in-bitcoin-8-24-2016.pdf
[144] M. Möser and R. Böhme, "Anonymous alone? measuring bitcoin's second-generation anonymization techniques," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Apr. 2017, pp. 32–41.

[145] M. Jourdan, S. Blandin, L. Wynter, and P. Deshpande, "Characterizing entities in the bitcoin blockchain," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2018, pp. 55–62.
[146] P. Tasca and S. Liu, "The evolution of the bitcoin economy: Extracting and analyzing the network of payment relationships," *SSRN Electron. J.*, vol. 19, no. 2, pp. 94–126, 2018.
[147] K. Sedgwick. (2017). *The Ethereum Blockchain is Congested by Cats*. [Online]. Available: https://news.bitcoin.com/ethereum-blockchain-congested-cats/
[148] A. F. Bariviera, "The inefficiency of bitcoin revisited: A dynamic approach," *Econ. Lett.*, vol. 161, pp. 1–4, Dec. 2017.
[149] A. K. Tiwari, R. Jana, D. Das, and D. Roubaud, "Informational efficiency of bitcoin—An extension," *Econ. Lett*, vol. 163, pp. 106–109, Feb. 2018.
[150] S. Nadarajah and J. Chu, "On the inefficiency of bitcoin," *Econ. Lett.*, vol. 150, pp. 6–9, Jan. 2017.
[151] H. Y. D. Sigaki, M. Perc, and H. V. Ribeiro, "Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.
[152] M. Buchholz, J. Delaney, J. Warren, and J. Parker. (2012). *Bits and Bets, Information, Price Volatility, and Demand for Bitcoin*. [Online]. Available: https://www.reed.edu/economics/parker/s12/312/finalproj/Bitcoin.pdf
[153] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.
[154] S. S. Adebola, L. A. Gil-Alana, and G. Madigu, "Gold prices and the cryptocurrencies: Evidence of convergence and cointegration," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 1227–1236, Jun. 2019.
[155] W. Zhang, P. Wang, X. Li, and D. Shen, "The inefficiency of cryptocurrency and its cross-correlation with dow jones industrial average," *Phys. A, Stat. Mech. Appl.*, vol. 510, pp. 658–670, Nov. 2018.
[156] W. Fang, S. Tian, and J. Wang, "Multiscale fluctuations and complexity synchronization of bitcoin in China and US markets," *Phys. A, Stat. Mech. Appl.*, vol. 512, pp. 109–120, Dec. 2018.
[157] A. H. Dyhrberg, "Bitcoin, gold and the dollar—A GARCH volatility analysis," *Finance Res. Lett.*, vol. 16, pp. 85–92, Feb. 2016.
[158] A. ElBahrawy, L. Alessandretti, A. Kandler, R. Pastor-Satorras, and A. Baronchelli, "Evolutionary dynamics of the cryptocurrency market," *Roy. Soc. Open Sci.*, vol. 4, no. 11, Nov. 2017, Art. no. 170623.
[159] C. Beneki, A. Koulis, N. A. Kyriazis, and S. Papadamou, "Investigating volatility transmission and hedging properties between bitcoin and Ethereum," *Res. Int. Bus. Finance*, vol. 48, pp. 219–227, Apr. 2019.
[160] X. Fan Liu, Z.-X. Lin, and X.-P. Han, "Homogeneity and heterogeneity of cryptocurrencies," 2019, *arXiv:1910.01330*. [Online]. Available: http://arxiv.org/abs/1910.01330
[161] A. Burnie and E. Yilmaz, "An analysis of the change in discussions on social media with bitcoin price," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, Jul. 2019, pp. 889–892.
[162] D. Garcia and F. Schweitzer, "Social signals and algorithmic trading of bitcoin," *Roy. Soc. Open Sci.*, vol. 2, no. 9, Sep. 2015, Art. no. 150288.
[163] L. Kristoufek, "BitCoin meets Google trends and wikipedia: Quantifying the relationship between phenomena of the Internet era," *Sci. Rep.*, vol. 3, no. 1, Dec. 2013.
[164] M. Al Mamun, G. S. Uddin, M. T. Suleman, and S. H. Kang, "Geopolitical risk, uncertainty and bitcoin investment," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123107.
[165] N. A. Kyriazis, "A survey on efficiency and profitable trading opportunities in cryptocurrency markets," *J. Risk Financial Manage.*, vol. 12, no. 2, p. 67, Apr. 2019.
[166] L. Ante, "Bitcoin transactions, information asymmetry and trading volumes," *Quant. Finance Econ.*, vol. 4, no. 3, pp. 365–381, 2020.
[167] W. Chen, J. Wu, Z. Zheng, C. Chen, and Y. Zhou, "Market manipulation of bitcoin: Evidence from mining the Mt. Gox transaction network," in *Proc. INFOCOM IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 964–972.
[168] J. Hamrick, F. Rouhi, A. Mukherjee, A. Feder, N. Gandal, T. Moore, and M. Vasek. (2018). *An Examination of the Cryptocurrency Pump and Dump Ecosystem*. [Online]. Available: https://ssrn.com/abstract=3303365
[169] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, "Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact," *Future Gener. Comput. Syst.*, vol. 102, pp. 259–277, Jan. 2020.
[170] K. Toyoda, P. T. Mathiopoulos, and T. Ohtsuki, "A novel methodology for HYIP operators' bitcoin addresses identification," *IEEE Access*, vol. 7, pp. 74835–74848, Jun. 2019.
[171] M. Vasek and T. Moore, "There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams," in *Proc. 19th Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2015, pp. 44–61.
[172] M. Vasek and T. Moore, "Analyzing the bitcoin Ponzi scheme ecosystem," in *Proc. 22nd Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2018, pp. 101–112.
[173] M. Bartoletti, B. Pes, and S. Serusi, "Data mining for detecting bitcoin Ponzi schemes," in *Proc. Crypto Valley Conf. Blockchain Technol. (CVCBT)*, Jun. 2018, pp. 75–84.
[174] C. Lee, S. Maharjan, K. Ko, and J. W.-K. Hong, "Toward detecting illegal transactions on bitcoin using machine-learning methods," in *Proc. Int. Conf. Blockchain. Trustworthy. Syst.* Singapore: Springer, 2019, pp. 520–533.
[175] S. Farrugia, J. Ellul, and G. Azzopardi, "Detection of illicit accounts over the Ethereum blockchain," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113318.
[176] M. Ostapowicz and K. Żbikowski, "Detecting fraudulent accounts on blockchain: A supervised approach," in *Proc. 20th Int. Conf. Web Inf. Syst. Eng. (WISE)*. Cham, Switzerland: Springer, 2019, pp. 18–31.
[177] W. Chen, Z. Zheng, E. C.-H. Ngai, P. Zheng, and Y. Zhou, "Exploiting blockchain data to detect smart Ponzi schemes on Ethereum," *IEEE Access*, vol. 7, pp. 37575–37586, 2019.
[178] E. Jung, M. Le Tilly, A. Gehani, and Y. Ge, "Data mining-based Ethereum fraud detection," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 266–273.
[179] J. Liang, L. Li, W. Chen, and D. Zeng, "Targeted addresses identification for bitcoin with network representation learning," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Jul. 2019, pp. 158–160.
[180] F. Zola, M. Eguimendia, J. L. Bruse, and R. Orduna Urrutia, "Cascading machine learning to attack bitcoin anonymity," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 10–17.
[181] R. Michalski, D. Dziubaltowska, and P. Macek, "Revealing the character of nodes in a blockchain with supervised learning," *IEEE Access*, vol. 8, pp. 109639–109647, 2020.
[182] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, and S.-W. Liao, "An evaluation of bitcoin address classification based on transaction history summarization," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2019, pp. 302–310.
[183] H. H. Sun Yin, K. Langenheldt, M. Harlev, R. R. Mukkamala, and R. Vatrapu, "Regulating cryptocurrencies: A supervised machine learning approach to de-anonymizing the bitcoin blockchain," *J. Manage. Inf. Syst.*, vol. 36, no. 1, pp. 37–73, Jan. 2019.
[184] I. Madan, S. Saluja, and A. Zhao, "Automated bitcoin trading via machine learning algorithms," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.
[185] A. van Schetsen, "Impact of graph-based features on bitcoin prices," M.S. thesis, Dept. Elect. Eng., Delft Univ. Technol., Delft, The Netherlands, 2019. [Online]. Available: http://resolver.tudelft.nl/uuid:363d443c-64f6-4c35-9671-4092aa334923
[186] A. K. Dey, C. G. Akcora, Y. R. Gel, and M. Kantarcioglu, "On the role of local blockchain network features in cryptocurrency price formation," *Can. J. Statist.*, vol. 48, no. 3, pp. 561–581, Mar. 2020.
[187] E. Sin and L. Wang, "Bitcoin price prediction using ensembles of neural networks," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2017, pp. 666–671.
[188] H. Jang and J. Lee, "An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information," *IEEE Access*, vol. 6, pp. 5427–5437, 2018.
[189] S. Ji, J. Kim, and H. Im, "A comparative study of bitcoin price prediction using deep learning," *Mathematics*, vol. 7, no. 10, p. 898, Sep. 2019.
[190] D. C. A. Mallqui and R. A. S. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques," *Appl. Soft Comput.*, vol. 75, pp. 596–606, Feb. 2019.
[191] M. Saad, J. Choi, D. Nyang, J. Kim, and A. Mohaisen, "Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions," *IEEE Syst. J.*, vol. 14, no. 1, pp. 321–332, Mar. 2020.
[192] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, Feb. 2020, Art. no. 112395.
[193] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, and M. Alazab, "Stochastic neural networks for cryptocurrency price prediction," *IEEE Access*, vol. 8, pp. 82804–82818, 2020.

[194] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of bitcoin prices using high-dimensional features: A machine learning approach," *Neural Comput. Appl.*, pp. 1–15, Jul. 2020, doi: 10.1007/s00521-020-05129-6.
[195] Y. Wang and H. Wang, "Using networks and partial differential equations to forecast bitcoin price movement," *Chaos: Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 7, Jul. 2020, Art. no. 073127.
[196] P. M. Monamo, V. Marivate, and B. Twala, "A multifaceted approach to bitcoin fraud detection: Global and local outliers," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 188–194.
[197] J. Hirshman, Y. Huang, and S. Macke, "Unsupervised approaches to detecting anomalous behavior in the bitcoin transaction network," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2013.
[198] L. Ermann, K. M. Frahm, and D. L. Shepelyansky, "Google matrix of bitcoin network," *Eur. Phys. J. B*, vol. 91, no. 6, pp. 1–13, Jun. 2018.
[199] K. Toyoda, T. Ohtsuki, and P. T. Mathiopoulos, "Time series analysis for bitcoin transactions: The case of Pirate@40's HYIP scheme," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2018, pp. 151–155.
[200] R. D. Camino, R. State, L. Montero, and P. Valtchev, "Finding suspicious activities in financial transactions and distributed ledgers," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 787–796.
[201] *Blockcypher*. Accessed: Nov. 1, 2020. [Online]. Available: https://live.blockcypher.com/
[202] M. Spagnuolo, F. Maggi, and S. Zanero, "Bitiodine: Extracting intelligence from the bitcoin network," in *Proc. 18th Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2014, pp. 457–468.
[203] *Blockchain2graph*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/straumat/blockchain2graph
[204] *Blocksci*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/citp/BlockSci
[205] *Btcspark*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/JeremyRubin/BTCSpark
[206] *Bitcoin-Blockchain-Parser*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/alecalve/python-bitcoin-blockchain-parser
[207] *Blocketl*. Accessed: Nov. 1, 2020. [Online]. Available: http://sc.hubwiz.com/codebag/blocketl-java/
[208] *Blockparser*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/znort987/blockparser
[209] *Rusty-Blockparser*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/gcarq/rusty-blockparser
[210] *Btctrackr*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/adoll/btctrackr
[211] *Bitcoinuses*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/qdm12/BitcoinUses
[212] *Blockchain ETL*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/blockchain-etl
[213] S. Bistarelli, I. Mercanti, and F. Santini, "A suite of tools for the forensic analysis of bitcoin transactions: Preliminary report," in *Proc. Eur. Conf. Parallel Process.* Cham, Switzerland: Springer, 2018, pp. 329–341.
[214] *Ether_Sql*. [Online]. Available: https://github.com/analyseether/ether_sql
[215] *Blockapi: Blockchain Analytics API*. [Online]. Available: https://github.com/blockchain-unica/blockapi
[216] Y. Li, K. Zheng, Y. Yan, Q. Liu, and X. Zhou, "EtherQL: A query layer for blockchain system," in *Proc. 22nd Int. Conf. Database Syst. Adv. Appl. (DASFAA)*. Cham, Switzerland: Springer, 2017, pp. 556–567.
[217] T. Chen, T. Hu, J. Chen, X. Zhang, Z. Li, Y. Zhang, X. Luo, A. Chen, K. Yang, B. Hu, T. Zhu, and S. Deng, "DataEther: Data exploration framework for Ethereum," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2019, pp. 1369–1380.
[218] T. Chen, X. Luo, Y. Zhang, T. Wang, Z. Li, R. Cao, X. Xiao, and X. Zhang, "Tokenscope: Automatically detecting inconsistent behaviors of cryptocurrency tokens in Ethereum," in *Proc. 26th ACM Conf Comput. Commun Secur. (CCS)*, 2019, pp. 1503–1520.
[219] J. Krupp and C. Rossow, "TeeTher: Gnawing at Ethereum to automatically exploit smart contracts," in *Proc. 27th USENIX Secur. Symp.*, 2018, pp. 1317–1333.
[220] Y. Zhou, D. Kumar, S. Bakshi, J. Mason, A. Miller, and M. Bailey, "Erays: Reverse engineering Ethereum's opaque smart contracts," in *Proc. 27th USENIX Secur. Symp.*, 2018, pp. 1371–1385.
[221] 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 *Proc. IEEE Symp. Vis. Cyber Secur. (VizSec)*, Oct. 2015, pp. 1–8.
[222] X. Yue, X. Shu, X. Zhu, X. Du, Z. Yu, D. Papadopoulos, and S. Liu, "BitExTract: Interactive visualization for extracting bitcoin exchange intelligence," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 162–171, Jan. 2019.
[223] F. Oggier, S. Phetsouvanh, and A. Datta, "BiVA: Bitcoin network visualization & analysis," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2018, pp. 1469–1474.
[224] J.-Z. Xia, Y.-H. Zhang, H. Ye, Y. Wang, G. Jiang, Y. Zhao, C. Xie, X.-Y. Kui, S.-H. Liao, and W.-P. Wang, "SuPoolVisor: A visual analytics system for mining pool surveillance," *Frontiers Inf. Technol. Electron. Eng.*, vol. 21, no. 4, pp. 507–523, Apr. 2020.
[225] *goBlockchaindataanalysis*. Accessed: Nov. 1, 2020. [Online]. Available: https://github.com/arnaucube/goBlockchainDataAnalysis
[226] H. Yousaf, G. Kappos, and S. Meiklejohn, "Tracing transactions across cryptocurrency ledgers," in *Proc. 28th USENIX Secur. Symp.*, 2019, pp. 837–850.
[227] A. Di Luzio, A. Mei, and J. Stefa, "Consensus robustness and transaction de-anonymization in the ripple currency exchange system," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 140–150.
[228] *Investor Bulletin: Initial Coin Offerings*, U.S. Securities, Exchange Commission, Washington, DC, USA, 2017.

XIAO FAN LIU (Member, IEEE) received the B.Sc. degree (Hons.) in electronic and information engineering and the Ph.D. degree from The Hong Kong Polytechnic University, in 2008 and 2012, respectively. He is currently an Assistant Professor with the City University of Hong Kong, Hong Kong, SAR, China. His research interests include cryptocurrency, blockchain, and social network analysis. He is a member of the China Computer Federation (CCF) Block Chain Technical Committee.

XIN-JIAN JIANG received the B.Eng. degree in computer science and technology from Nanjing Agricultural University, Nanjing, China, in 2018. He is currently pursuing the M.Eng. degree in computer science and technology with Southeast University. He is currently working as a Research Assistant with City University of Hong Kong. His research interests include networks and data analysis for cryptocurrencies.

SI-HAO LIU received the B.Eng. degree in computer science and technology from Southeast University, Nanjing, China, in 2018, where she is currently pursuing the M.Eng. degree in computer science and technology. She is currently working as a Research Assistant with the City University of Hong Kong. Her research interests include networks and data analysis for cryptocurrencies.

CHI KONG TSE (Fellow, IEEE) received the B.Eng. degree (Hons.) in electrical engineering and the Ph.D. degree from the University of Melbourne, Melbourne, VIC, Australia, in 1987 and 1991, respectively. He is currently a Chair Professor of Electrical Engineering with the City University of Hong Kong, Hong Kong. His research interests include power electronics, nonlinear systems, and complex network applications.