International Journal of Robotics and Control Systems
Vol. 3, No. 3, 2023, pp. 609-626
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Exploring Blockchain Data Analysis and Its Communications Architecture: Achievements, Challenges, and Future Directions: A Review Article

Hamzah M. Marhoon [a,1,*], Noorulden Basil [b,2], Alfian Ma'arif [c,3]

[a] Department of Systems Engineering, College of Information Engineering, Al-Nahrain University, Jadriya, Baghdad, Iraq
[b] Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq
[c] Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[1] hamzah.marhoon@nahrainuniv.edu.iq; [2] noorulden@uomustansiriyah.edu.iq; [3] alfianmaarif@ee.uad.ac.id
* Corresponding Author

ARTICLE INFO

Article history
Received July 13, 2023
Revised August 19, 2023
Accepted September 06, 2023

Keywords
Blockchain;
Data Analysis;
Bitcoin;
Cryptographic

ABSTRACT

Blockchain technology is relatively young but has the potential to disrupt several industries. Since the emergence of Bitcoin, also known as Blockchain 1.0, there has been significant interest in this technology. The introduction of Ethereum, or Blockchain 2.0, has expanded the types of data that can be stored on blockchain networks. The increasing popularity of blockchain technology has given rise to new challenges, such as user privacy and illicit financial activities, but has also facilitated technical advancements. Blockchain technology utilizes cryptographic hashes of user input to record transactions. The public availability of blockchain data presents a unique opportunity for academics to analyze it and gain a better understanding of the challenges in blockchain communications. Researchers have never had access to such an opportunity before. Therefore, it is crucial to highlight the research problems, accomplishments, and potential trends and challenges in blockchain network data analysis and communications. This article aims to examine and summarize the field of blockchain data analysis and communications. The review encompasses the fundamental data types, analytical techniques, architecture, and operations related to blockchain networks. Seven research challenges are addressed: entity recognition, privacy, risk analysis, network visualization, network structure, market impact, and transaction pattern recognition. The latter half of this section discusses future research directions, opportunities, and challenges based on previous research limitations.

1. Introduction

A revolutionary form of distributed ledger technology, enabled by wireless communications, facilitates trustworthy transactions in an environment characterized by mutual mistrust, all without the need for intermediaries. Its anti-counterfeiting and non-tampering capabilities, coupled with the seamless deployment of smart contracts, differentiate blockchain as a disruptive technology in comparison to traditional database systems [1], [2]. The secure transfer of goods, data, and services among companies gains paramount importance in the business cycle, where a vulnerable supply chain

can have far-reaching consequences across industries. Notably, the healthcare sector faces an additional risk that directly affects patient health. Throughout the years 2017 to 2020, blockchain technology consistently secured a spot among the top 10 strategic technologies, as recognized by Gartner, a prominent IT research and analysis company [3]-[5].

Blockchain, as the underlying technology for digital currencies like Bitcoin, employs cryptographic procedures to ensure user anonymity while the distributed network verifies and records all transaction data. This decentralization is crucial for credible transactions, and there are three main types of blockchains: public chains (e.g., Bitcoin, Ethereum), consortium chains, and private chains. Public chains allow nodes to freely join or leave the network, providing easy access to blockchain data. This transparency and immutability make them ideal for applications where trust is paramount, such as cryptocurrencies. Consortium chains, in contrast, are semi-decentralized and restrict access to trusted nodes or organizations. They strike a balance between decentralization and control, making them attractive for use in industries like supply chain management. Private chains, on the other hand, are fully centralized and only accessible to authorized entities. They prioritize security and efficiency, making them suitable for sensitive applications within organizations. The availability of public blockchain data offers a unique opportunity for data analysts to explore the system's functionality. With substantial user bases and significant transaction volumes on public networks like Bitcoin and Ethereum, studying blockchain-based data analysis becomes crucial and valuable. However, analyzing blockchain data presents challenges due to the decentralized and anonymous nature of the system. Developing data analysis techniques based on anonymous networks becomes essential for effective blockchain data analysis. As blockchain technology is still in its early stages, and mature blockchain application projects are limited, data analysis based on blockchain remains exploratory. Nonetheless, comprehending the distinct characteristics and differences among public, consortium, and private chains is vital in determining their suitability for various use cases and industries [6]-[16].

Existing literature [17] primarily focuses on mature blockchains with substantial data, such as Bitcoin and Ethereum, which have garnered significant attention due to their long-standing presence. While numerous reviews on blockchain technology exist, covering technical architecture, consensus mechanisms, security and privacy concerns, attack issues, and application status, there is a dearth of relevant reports on the progress of blockchain data analysis [18]. This paper aims to compare and analyze the existing literature [19] on data analysis of Bitcoin and Ethereum, summarizing the two typical types of blockchain data and the corresponding analysis methods. Furthermore, it presents seven major research problems and progress in blockchain data analysis. The insights provided can serve as references and contribute to ongoing research on blockchain technology.

The integration of Information Science and Blockchain technology assumes a pivotal role in accurately identifying the most suitable indicators in the given context. Notably, information scientists have made significant advancements by creating an automated trading system within the MetaTrader5 platform, leveraging automated operators. This system significantly enhances the speed of decision-making and forecasting processes, as it effectively follows trading indicators that are meticulously programmed using the MQL code through MetaTrader Editors. The transition towards automated systems arises from the well-established observation that human traders are vulnerable to emotional biases, while machines demonstrate superior control and objectivity in executing trading strategies. The utilization of automated systems represents a groundbreaking development in the domain of algorithmic trading, supplanting traditional human-based approaches with more reliable and accurate methodologies. By relying on machine-driven decisions, this cutting-edge technology exhibits the potential to revolutionize financial markets and bolster trading efficiency. It is worth emphasizing that the continuous evolution of such systems promises even more refined and sophisticated solutions, contributing to the advancement of trading practices and yielding tangible benefits for traders and investors alike [20], [21].

According to the researchers in [22], blockchain and cryptocurrency offer both economic potential and risks on the internet. The authors have classified various cryptocurrencies and investigated the current and future economies of Bitcoin and blockchain. The researchers have also

explored the differences between cryptocurrencies and traditional cash, while examining statistics related to cryptocurrency research publication ratings and capitalization. Their study thoroughly examines the uses of cryptocurrencies and assesses their risks and benefits. Additionally, it discusses the global legal impact of cryptocurrencies. According to the study's authors, these technologies pose threats to financial markets and cryptocurrency exchanges but also hold the potential for digital economy growth. The report recommends mitigating the hazards associated with Bitcoin and blockchain on the 4.0 Industrial Platform in the Digital World. It presents data on cryptocurrency capitalization movements, rankings, and statistical analysis of scientific papers. The study focuses on various aspects of cryptocurrencies, including online and financial crypto assets. Furthermore, it addresses the challenges of global cryptocurrency law.

Blockchain technology has the potential to disrupt industries and has gained significant interest since the emergence of Bitcoin (blockchain 1.0) and Ethereum (blockchain 2.0). It presents opportunities for academic analysis but also raises challenges like user privacy and illicit activities. This article reviews blockchain data analysis, covering fundamentals and addressing seven research challenges. Future research directions and challenges are also discussed.

2. Blockchain Basics

In the present, we are overwhelmed by an abundance of data, but the authenticity and reliability of this data remain uncertain. However, blockchain technology possesses a unique mechanism that endows its data with the crucial attribute of “trust,” making information derived from blockchain data analysis highly valuable. To comprehend the reasons behind the trustworthiness of blockchain data, this chapter aims to outline the fundamental architecture of blockchain and delve into an analysis of the key technologies that contribute to the credibility of blockchain data.

2.1. Blockchain Architecture

In 2008, a researcher known by the pseudonym “Satoshi Nakamoto” published a groundbreaking paper on Bitcoin in the cryptography mailing group [23]. In January 2009, the initial version of Bitcoin was successfully implemented. Over time, Bitcoin gained widespread attention and popularity, owing to its exceptional qualities, making it a favored asset in the financial market. Blockchain technology, the underlying framework supporting cryptocurrencies like Bitcoin, has become a subject of interest [24]. Although there is no universally accepted definition for blockchain, in October 2016, China's Ministry of Industry and Information Technology released the China Blockchain Technology and Application Development White Paper, which defined blockchain as a new form of computing that encompasses a distributed ledger, a consensus process, an encryption algorithm, and a point-to-point transmission protocol [25], [26].

The white paper further divided the smart contract architecture into various layers, namely the data layer, transport layer, main body, verification layer, execution layer, and application layer. The data layer serves as the storage for blockchain data and communicates with the transport layer through an API to facilitate data transfer to the smart contract body. In certain commercial applications, the presence of coins or incentives may not be necessary, leading to the absence of an incentive layer. In terms of privacy, the design of blockchain incorporates three layers: network, transaction, and application [26], [27]. From a data analysis perspective, this study describes the structure of blockchain as a three-horizontal and one-vertical arrangement, as exhibited in Fig. 1.

The three horizontal sections in the diagram not only represent the abstraction of blockchain data types but also signify the three developmental stages of blockchain. The bottom layer, known as the transaction layer, corresponds to the blockchain 1.0 stage exemplified by Bitcoin. Transactions serve as the mechanism for updating existing blocks of information and are the fundamental building blocks of the blockchain. Blockchain 1.0 primarily focuses on recording transactions and ensuring the global uniqueness and data integrity of the Bitcoin blockchain itself. Above the transaction layer is the cryptographic protocol layer, which acts as the intermediate layer [28].

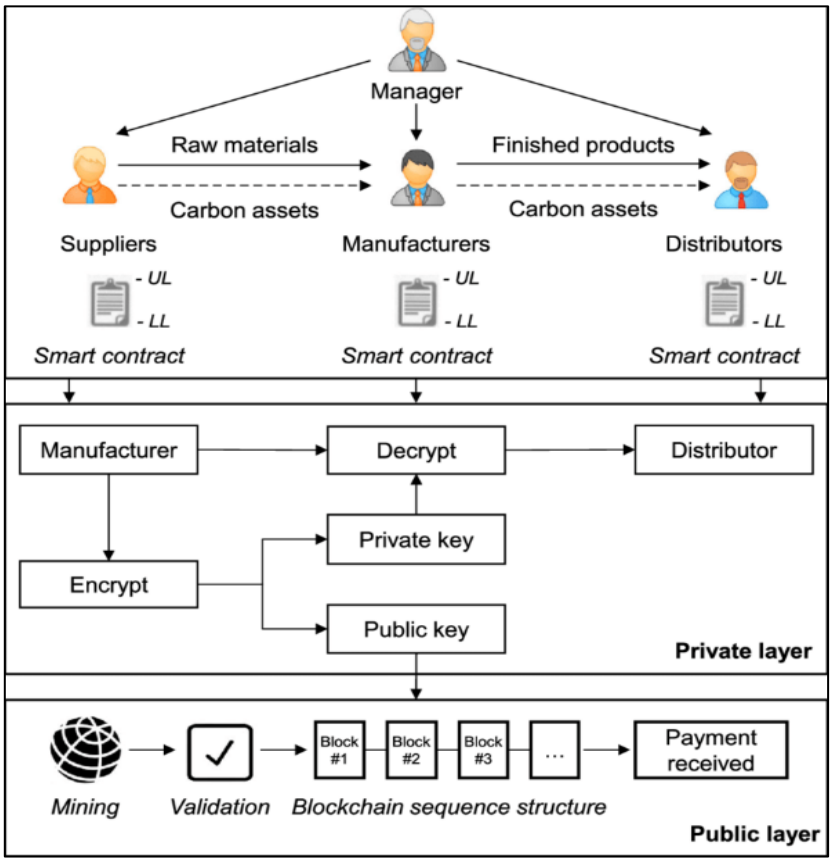

Fig. 1. The blockchain framework

Digitizing contractual terms led to the emergence of "smart contracts." A smart contract is essentially a computer program that becomes operational when predefined conditions are met. Leveraging the essential characteristics of blockchain data, such as immutability, integrity, and trustworthiness, smart contracts based on blockchain data have become a natural choice [29]. The integration of blockchains and smart contracts is often referred to as blockchain 2.0. Another type of data included in blockchain 2.0 is contract data, which encompasses smart contracts. It is worth noting that smart contracts are closely tied to transactions, as the implementation and execution of smart contracts generate transaction data. At the topmost layer of the stack lies the application layer, representing the apps of Blockchain 3.0. The contract layer is depicted as a dashed line in the diagram, indicating that apps can be designed either on top of smart contracts or without relying on smart contracts. An example of this would be applications built on the Bitcoin blockchain. The scarcity of authentic use cases and application data is due to the relative immaturity of blockchain technology, which is still in its early stages [30], [31].

The presence of a single vertical line indicates that blockchain is utilized in a decentralized environment. Each of the three levels of the blockchain is distributed since the distributed nature of the blockchain spans across all three levels. For instance, smart contract data is stored on every node within a distributed system. When a node receives a transaction initiating a contract process, it is responsible for executing the necessary contract based on the data stored in its own copy of the blockchain and transmitting the resulting operation. Due to the consensus mechanism, the data from the remote network remain synchronized with the data from the local network. In a typical implementation, the distributed network architecture supporting a blockchain comprises a significant number of nodes. Each node may have distinct responsibilities within the network, such as wallet management, mining (competition for accounting rights), full block data storage, and routing. Nodes often perform tasks unrelated to their primary function. Even if a node's main purpose is wallet implementation, it still needs to be connected to other nodes for broadcasting and validating

transactions. This aspect of the process is referred to as the routing function. Owing to resource limitations, it is usually sufficient to maintain only the blockchain headers rather than the entire block's worth of data. A node that performs all four functions is called a complete node, and a mining node typically becomes fully functional after fulfilling all assigned tasks and responsibilities. Full nodes play a critical role in the network's support and maintenance infrastructure due to their ability to fulfill all functions. In the subsequent discussion, unless specified otherwise, the term "node" refers specifically to this type of full node [32].

2.2. Key Technologies of Blockchain

The blockchain, in a distributed ledger database, can be viewed as a digital accounting system, distinguished by its central concept of "non-tampering." The immutability of data stored in a blockchain establishes its inherent trustworthiness. The main challenge in establishing an immutable ledger within a decentralized environment is determining how to organize the data in a way that prevents tampering and achieves consensus among all participants regarding the state of the ledger. These challenges can be addressed through the use of data structures and consensus mechanisms. For instance, the Bitcoin network demonstrates two essential technologies. The data structure of the blockchain determines its ability to track accounts and transactions, often utilizing Merkle trees to structure the listing of accounts and transactions [33]. Each node in a Merkle tree, or hash tree, possesses hash values. Fig. 2 illustrates the Merkle tree data structure employed in Bitcoin. When a Bitcoin transaction occurs, the node arranges the transactions based on their receiving order or transaction fees and computes a hash value for each transaction using a hash operation. This value serves as the hash of a Merkle leaf node. By repeatedly combining these values, fresh hash values can be generated. The root node (Merkle root) of the Merkle tree represents transactions for a specific period. Any modification to a transaction (such as address or amount) or changes in its order would result in a different root node. In Ethereum, the transaction is replaced with the account for smart contract facilitation. When the account status changes, the hash value also changes, consequently altering the root node of the Merkle tree. Recalculating the Merkle tree node in the corresponding block after a transactional or account status change should yield the same result as before. To protect blockchain data from modification, the initial step involves creating a "snapshot" at the root of the Merkle tree [33], [34].

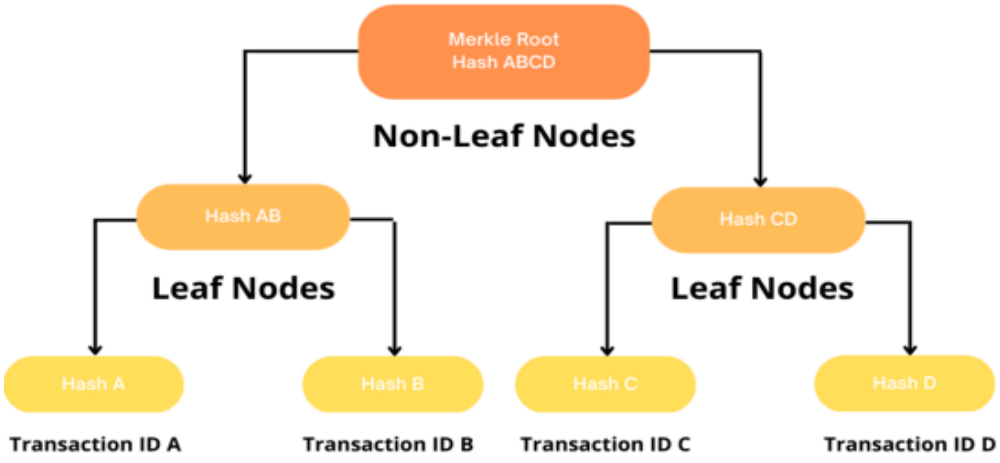

Fig. 2. Merkle tree structure

Merkle Trees, widely utilized in various computer science applications, including blockchain technology, offer several advantages, including the Simplified Payment Verification (SPV) feature. They enhance the effectiveness and security of data within networks like Bitcoin, providing a reliable and secure method for running and verifying blockchains. By utilizing a Merkle Tree database, block contents are securely divided, ensuring integrity and protection against loss, damage, or tampering. This data management technique allows for the validation of specific transactions without the need to

download the entire terabytes-sized blockchain. The Merkle tree is a mathematical data structure based on hashing that serves as a summary of all transactions within a block, as shown in Fig. 3. It offers a decentralized and efficient approach to verify data accuracy. Given their unique characteristics, Merkle trees play a crucial role in enhancing the security and efficiency of encrypting blockchain data, particularly in Peer-to-Peer (P2P) networks where information exchange and independent evaluation are necessary. Merkle trees have a binary tree structure, with the top hash known as the "Root" and the transactional data hashes in the bottom row referred to as "Leaf Nodes." While most hash tree implementations are binary, with each node having two child nodes, Merkle trees have the flexibility to accommodate any number of child nodes without limitations [35].

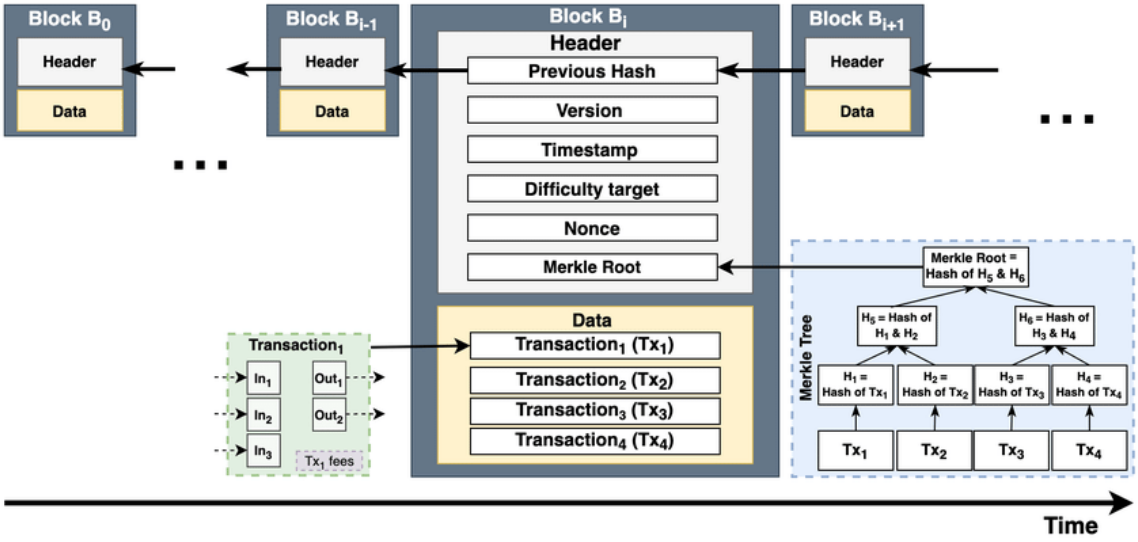

Fig. 3. Abbreviated Bitcoin blockchain data structure

As mentioned previously, the Merkle tree plays a pivotal role in enabling the efficient verification of transactions and significantly enhances the security of blockchain data. By employing a hierarchical structure of cryptographic hash functions, the Merkle tree condenses vast amounts of transactional data into a single root hash, representing the entire set of transactions within a given block. This condensed representation not only reduces the storage requirements but also expedites the verification process. When a new transaction is added to the blockchain, only the relevant branches of the Merkle tree need to be recalculated, rather than recomputing the entire tree. As a result, the verification process becomes much faster and resource-efficient. Moreover, the Merkle tree enhances the security of blockchain data through its inherent immutability. Any alteration or tampering of a single transaction within the block would lead to a mismatch in the root hash, causing the entire Merkle tree to be invalid. Since each block's hash includes the hash of the previous block, any modification in a block would create a domino effect, altering subsequent block hashes, thus rendering the entire blockchain inconsistent. This property ensures the integrity of the blockchain, making it highly resistant to tampering and unauthorized modifications. Merkle tree's implementation in blockchain technology optimizes the verification process, reduces resource consumption, and enhances data security. Its efficient structure contributes to the scalability and reliability of blockchain networks, making it a fundamental component for building robust and secure decentralized systems [36].

3. Blockchain Data Types

The foundation of blockchain data gathering is an appreciation for the various forms that blockchain data might take. Considering this, this section provides an overview of the features and formats of data found in blockchains. The three tiers of blockchain technology are distinct from one another and are evolving independently of one another. At the moment, the market is concerned with and has matured implementations of both Bitcoin's blockchain 1.0 and Ethereum's blockchain 2.0.

Therefore, the following will provide an overview of the most crucial forms of information in the present phase of the game: transaction data or contract data and the other types.

In the realm of blockchain technology, a variety of data types are employed to represent and store information on the decentralized ledger. These distinct blockchain data types include [37]-[40]:

a) Transaction Data: This data type is instrumental in encapsulating the particulars of a specific transaction within the blockchain. It typically encompasses essential information such as the sender's address, the recipient's address, the amount of cryptocurrency transferred, a unique transaction ID, timestamp, and the status of the transaction.

b) Block Data: Blocks are fundamental units that house groups of transactions. Block data incorporates a distinctive block header, a reference to the preceding block (establishing the backbone of the blockchain), a root hash derived from the Merkle tree, and a nonce value employed in proof-of-work consensus mechanisms.

c) Smart Contract Data: Smart contracts, constituting self-executing agreements with predefined rules encoded in code, rely on this data type. It comprises the contract code, its current state, the contract address, and relevant execution details. Smart contracts are integral to Blockchain 2.0, representing pieces of code that execute based on pre-defined conditions. Ethereum, the leading smart contract platform, implements smart contract data through its Turing-complete virtual machine (EVM) and offers high-level languages like Solidity for easier contract writing. The data associated with smart contracts includes code data (source code and bytecode) and transaction data. However, due to anonymity requirements, only bytecode is typically provided when deploying a smart contract, limiting access to the source code. Researchers face challenges in analyzing bytecode data, but techniques like decompilation into virtual machine opcodes can help study contract functionality and potential vulnerabilities. Research on smart contracts is still in its early stages, with many aspects, including contract functionalities and interrelationships, yet to be explored.

d) Address Data: In the context of blockchain, addresses are deployed to represent the identities of participants, encompassing users or nodes. Address data typically comprises a combination of alphanumeric characters and functions as a reference point for the initiation and reception of transactions.

e) Consensus Data: This data type encompasses information pertinent to the consensus mechanism deployed to ascertain the validity of transactions and blocks. For instance, in proof-of-work blockchains, consensus data would include the difficulty level and the nonce value.

f) Metadata: Blockchain systems frequently incorporate additional metadata to enhance data traceability and verification. Timestamps, digital signatures, or data pertaining to the network state are instances of such metadata.

The employment of these diverse blockchain data types plays a pivotal role in safeguarding the integrity, security, and overall functionality of the blockchain network. By facilitating the decentralized and immutable organization and management of data, blockchain technology establishes an environment that enables trustless and transparent transactions among participants.

4. Research Status and Progress

In this section, we will discuss the most significant concerns and recent advancements in analyzing blockchain data. We will review the relevant literature on this subject and present a summary of seven areas of study in blockchain data analysis: entity recognition, assessing the potential breach of personal privacy, mapping the operational dynamics of a network, identifying suspicious payments, evaluating market impact, and identifying and analyzing illegal activities. These seven areas of study are interconnected, as illustrated in Fig. 4. Solid arrows in the figure indicate that the research question represented at the originating end of the arrow supports the investigation of the

problem represented at the end of the arrow, or the originating end serves as the foundation for the end. The expertise of the research question represented at a certain angle end is the issue represented at the end, indicating the relationship between the whole and its parts. Additionally, the double-headed arrows in the figure demonstrate that the research of the whole supports the research of the parts [41], [42].

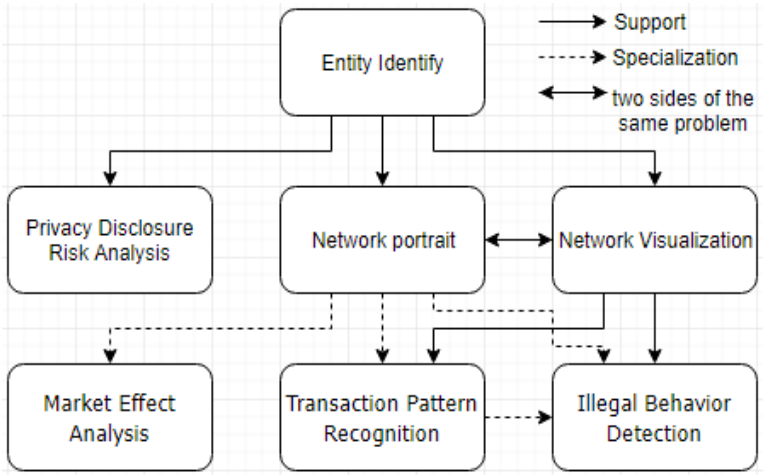

Fig. 4. Research problems and their relationship

4.1. Entity Recognition

Given that Bitcoin transactions are anonymous and can involve multiple inputs and outputs, it is natural to question whether it is possible to determine which addresses belong to the same user by examining their transaction history. The literature [43] commonly assumes that the identified entity is an entity rather than a user due to the absence of a method to confirm the distinction. Entities can include users, institutions, and others. However, it is important to note that a single user or organization may have control over multiple distinct entities. Common input methods and dynamic address methods are two examples of heuristic approaches used in published research [44] to discover probable entities. When many input addresses are used in a single transaction, the "common input" technique can be used to determine which input addresses share a "common input." Bitcoin requires you first to obtain the private key associated with a specific address before allowing you to transmit money to that address. Because users rarely reveal their private keys, the addresses used in a transaction's inputs are almost certainly associated with the same people or organisation. However, the literature [45] takes a different approach by investigating Bitcoin transactions using the Petri net theory rather than the method mentioned above. The addresses and transactions are converted into a matrix representation of the Petri net so that matrix operations can be used to examine a number of Bitcoin concerns such as entity recognition. Because of its simplicity, this method can be used to efficiently analyse a wide range of situations in a short period of time. As the number of transactions increases, the matrix dimensions become unmanageably large, which is a problem of this method when implemented; this is a negative of this approach. This technique has several problems, the most significant of which is that each transaction is represented by a column in the matrix, whereas each address is represented by a row.

Regardless of the method used to identify entities, it becomes evident that as the number of transactions in the Bitcoin system increases, the number of entities and corresponding addresses also grows. The size of entities can somewhat reflect the nature of the entities, and the number of entities can indicate the number of users in the system. For instance, an entity corresponding to a common user tends to be small, while an entity linked to an online wallet may consist of multiple addresses. It is important to note that this heuristic method is not foolproof, and its use poses risks to Bitcoin's anonymity. To enhance anonymity, solutions such as CoinJoin, MixCoin, and BlindCoin have been proposed. Additionally, literature [46] explores the effectiveness of this approach. The author first

employs the method to identify numerous super entities containing at least one thousand addresses. Through analysis, it is discovered that these super entities are associated with specific services, such as wallet services, which commonly reuse addresses to serve customers. Address reuse not only generates super entities but also compromises the privacy of service providers and users. However, the author notes that certain services, like the well-known bitcoin wallet and transaction provider Coinbase, completely avoid this issue, and no corresponding super entity exists. This highlights that the correspondence between entities and users cannot be solely determined based on the common input method, as a user may have multiple corresponding entities. Building upon the common input and change address heuristics, literature [47] proposes a community discovery method to further aggregate entities. The author conducts experiments to compare the effectiveness of different heuristic methods in entity recognition. The experimental results indicate that using the common input method alone achieves better recognition accuracy but has a low recall rate. Conversely, relying solely on the change address method results in the lowest accuracy. However, when combining all heuristics, although the accuracy decreases significantly, the recall reaches the highest level. This demonstrates that the communities discovered through the community discovery method are unlikely to belong to the same entity. Nevertheless, this method can uncover potential entities that ordinary heuristic methods may fail to identify. In order to improve the degree of entity aggregation established using the common input method, identifying change addresses among multiple output addresses of a transaction is crucial. Several heuristics exist for change address identification. The most straightforward approach is to consider the sole new address among the two output addresses of a transaction as the change address [48].

4.2. Risk Analysis of Privacy Leakage

The purpose of entity recognition is to identify all addresses in the system owned by the same entity. While this presents a security concern, it is impossible to link an address to a real person as it is merely a meaningless alias. However, it is reasonable to assume that if we have access to one user's address information, we have access to all of their addresses, along with their account activity, balance, and other private data. The goal of leakage risk analysis is to address questions such as “How can we link scheme entities to real entities?” and “What can we learn about a user's email when we have additional information about them?”. In terms of obtaining truthful entity information, the researchers in [49] employs heuristic methods to aggregate 12,056,684 addresses into 3,383,904 entities and then labels 2,197 of them by purchasing products and services from online merchants that accept Bitcoin payments. Labeling addresses through transactions can be precise but expensive. Some users voluntarily disclose their bitcoin addresses on forums, and organizations publish their bitcoin donation addresses, unintentionally divulging their address information. Labels for addresses are obtained through various communities and networks, which is a cost-effective method. Regarding the extent to which address information can be obtained based on additional information, the current analysis method is as follows: Firstly, based on the obtained entity information, transaction data is represented by a 5-tuple (S, R, M, V, T). A 5-tuple signifies that the transaction sender S sent a total amount of M coins to the receiver R at time T, with the corresponding value in legal currency being V. Secondly, it is assumed that the sender in the 5-tuple cannot be known, but the other four pieces of information may be exposed. Therefore, by querying blockchain data with the help of the other four pieces of information, the address information of the transaction sender can be obtained. This assumption is reasonable since users naturally protect their account information during payments, but other elements of a transaction may be exposed. Subsequently, latter customers can easily learn information about previous customers, such as payment amounts and times. Moreover, merchants may publicly display their address information for accepting payments. Based on these assumptions, the researchers in [50] analyzed the Bitcoin ledger and revealed that if a user inadvertently leaks payment amount information, an eavesdropper can analyze and find all transactions with similar amounts within the corresponding time window, thereby obtaining the corresponding address. Before 2012, around ten addresses could be found. Of course, if combined with additional information, the likelihood of finding a specific address would increase.

Various techniques exist for assessing user privacy in the Bitcoin blockchain system, and many Bitcoin-based businesses offer different methods to preserve user privacy and wallet security. However, according to research involving 990 Bitcoin users, only 46% utilize web-based privacy and security-preserving services, with half of the users having access to only one such service. Additionally, many people are unaware of the privacy issues associated with Bitcoin and only utilize a subset of available services. Furthermore, 22% of users reported losing bitcoins due to security and operational concerns. This scenario highlights the limited understanding among most Bitcoin users regarding this new phenomenon, which can easily contribute to unlawful and criminal activities [51].

4.3. Network Portrait

The Bitcoin main chain has mined over 500,000 blocks, encompassing more than 150GB of transaction data in recent years. When faced with such a significant volume of transaction data, it is natural to question the number of users involved, their characteristics, and whether this extensive payment network exhibits the traits of a complex network as a whole. Additionally, how is Bitcoin distributed among users as an "asset," and does it adhere to general economic regulations? This line of research, which explores aspects of the entire network, is known as network profiling. The following sections outline some recent developments in Bitcoin network profiling. Activity Profile: In order to understand the transformation of the Bitcoin community from an early coder's "experiment" to a vast cryptocurrency empire, numerous studies have examined the network's activities and dynamic changes. One study analyzed all transaction data in Bitcoin before May 2012 using the common input heuristic method, and it identified 1,851,544 entities with more than two addresses from a total of 3,730,218 addresses [52]-[53].

Service Portrait: The anonymous and expansive nature of the Bitcoin network often leaves people perplexed about the types of services and transactions that exist within it. To address this question, researchers have obtained information on many addresses corresponding to Bitcoin-based services such as mining pools, wallets, currency exchange services, and gambling by purchasing items and services. Network Characteristics: Early research on the Bitcoin network revealed power-law characteristics in the transaction network [54].

In complex network analysis, the preference attachment theory is often employed to explain why various network properties exhibit power-law distributions. In economics, preference attachment is also known as the "Matthew effect" or the phenomenon of the rich getting richer. As Bitcoin can be viewed as an asset, examining whether this phenomenon exists is an intriguing topic. Some literature has analyzed the Matthew effect in different periods within the Bitcoin system, and the findings indicate a clear wealth aggregation effect within the currency network—essentially, the rich tend to get richer [55].

4.4. Network Visualization

Keeping pace with the increasing number of transactions recorded on the blockchain signifies the growing acceptance of this distributed ledger technology. Researching methods for visualizing transaction networks is crucial in the face of a massive and continually expanding network. Numerous studies have been conducted on this topic. The literature serves as a foundation for blockchain graph mining by analyzing the distinctive graph structure embedded in each blockchain transaction and introducing properties specific to blockchains like Bitcoin and Ethereum [56].

One reference introduces Bit-Coin View, a network visualization system that enables straightforward, real-time monitoring of Bitcoin transactions. This framework introduces the concept of "purity," making it easier to identify mixed currency transactions and conduct real-time monitoring of potential money laundering activities within the Bitcoin network.

Graph-Sense, when applied to the blockchain, allows users to perform tasks such as monitoring monetary flows, implementing automatic entity recognition, and investigating unusual patterns in network transactions through graphical analysis. Reference describes a visual monitoring system designed to detect abnormal human or algorithmic activity in Bitcoin transactions. This technology

enables the identification of abnormal transaction patterns, such as money laundering within the Bitcoin system, and various offensive behaviors within the blockchain, including parasite transaction attacks. While most research has focused on Bitcoin's network, a group has developed a Scala-based open-source framework capable of examining both Bitcoin and Ethereum, the two largest blockchains, simultaneously [57]-[59].

4.5. Market Effect Analysis

The cryptocurrency represented by Bitcoin not only records the details of system transactions on a distributed ledger known as the blockchain, but it also: The eye-catching data is the exchange price between the cryptocurrency and the fiat currency, commonly referred to as the price of the cryptocurrency. As of the time of writing, coinmarketcap.com records 19,490 cryptocurrencies with 526 corresponding exchanges. The total market value of the cryptocurrency market exceeds $1.2 trillion, with Bitcoin being the dominant cryptocurrency valued at nearly $700 billion. The top three cryptocurrencies by market capitalization are Bitcoin, Ethereum, and Tether Coin, collectively accounting for over 60% of the entire cryptocurrency market. Fig. 5 illustrates the significant market growth, differentiating between Bitcoin and other cryptocurrencies. At the start of the year, Bitcoin represented 88% of the market share. However, by December, its market share had dropped to below 40% [60]. A general review of the widely used price influencing factors, which can be divided into six categories, is presented in Table 1. While the evolution of blockchain: key aspects and developments from 2017 to 2022 is presented in Table 2.

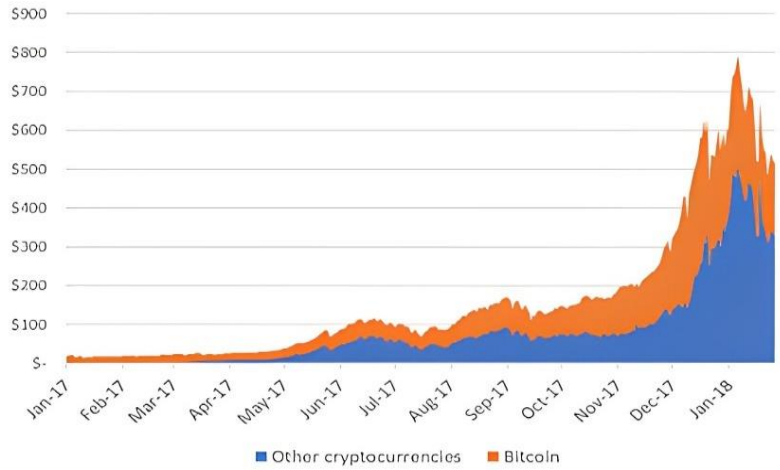

Fig. 5. Synoptic cryptocurrency market capitalization

4.6. Transaction Pattern Recognition

Bitcoin is decentralized and anonymous compared to more traditional payment mechanisms, such as banks. The emergence of certain characteristics of human payment behaviour in anonymous situations is an intriguing subject. However, due to Bitcoin's anonymous features facilitating unlawful activities like money laundering and fraud, it is important to identify unusual trends in the blockchain's transaction logs to uncover criminal activity. In the following sections, we describe various relevant studies that focus on identifying and analyzing Bitcoin transaction trends to find solutions to these issues. These studies encompass activities such as gambling and smuggling, which are known to be unlawful, but it is likely that many more illegal activities have yet to be uncovered. Therefore, it is crucial not only for the development of blockchain technology but also for the legislation and regulation of the blockchain industry to be able to identify illegal activities using blockchain data. The next part of this article will discuss the studies conducted on combating money laundering and fraud using blockchain technology. If M does not accept and return bitcoins from the same address, the flow of funds relationship is broken. M uses four addresses to accept mixed currency input, but in reality, he could use one address to accept all inputs in the same transaction. Although there are multiple ways to mix coins, after the coins are mixed, not only does the heuristic method based on common input

become invalid, but the attribution of the output address cannot be determined. In the example shown in Fig. 6, even if we know that a certain user (such as A) participated in the mixed currency transaction, we cannot determine which one of A', B', C', and D' belongs to the user because the four addresses appear to be identical [67], [68].

Table 1. Widely used price influencing factors can be divided into 6 categories [61]-[63]

Factors	Influencing
Miner factors	Are the main force in maintaining cryptocurrencies, and they are also the initial holders of cryptocurrencies. Their behavior will inevitably affect the price. There are many indicators that can reflect the behavior of absenteeism, such as hash ratio (representing miners' computing power input), transaction Fees, mining rewards, etc.
System factors	The setup of a cryptocurrency system is an important factor that affects the price of a currency. Taking Bitcoin as an example, its coin supply has been preset to be close to 1.2 trillion coins, its mining difficulty is adjusted every 2016 blocks, and the mining reward is reduced by half every 210,000 blocks. Settings and dynamics can affect the supply of a cryptocurrency and thus ultimately its price
User factors	Since Bitcoin has entered a mature stage from the early miner era and the illegal era dominated by gambling and black-market transactions [30], its participants are becoming more and more extensive and diverse. Therefore, the degree of user participation and participation methods must also be price-dependent. influencing factors. The main data indicators in this part come from blockchain data: such as the number of addresses, the number of entities, transaction volume, and transaction value. In addition, after constructing a transaction network with entities, many indicators can be obtained from the perspective of complex networks that reflect the activity and transaction patterns of users participating in transactions in the network
Policy and event factors	Due to the particularity of cryptocurrencies, they are greatly affected by policies. In addition, typical events in the cryptocurrency field, such as vulnerability attacks (DAO attacks), platform failures, etc. will inevitably have a huge impact on prices
Network factors	Internet search popularity and other factors reflect the degree of "pursuit" of ordinary netizens for cryptocurrencies, as well as the potential user scale and market sentiment of cryptocurrencies, which are also potential indicators
Competition and substitution factors	Competition and substitution factors. As mentioned earlier, there are currently 19,490 altcoins in the cryptocurrency market. Among the thousands of cryptocurrencies, some have completely different philosophies, but many more may just be improvements to the classics to some extent. Therefore, the prices between coins must also have a relationship of mutual influence and substitution. In addition, as an asset, cryptocurrencies may also compete or substitute with traditional assets such as gold and oil.

Table 2. Evolution of blockchain: key aspects and developments from 2017 to 2022 [64]-[66]

Year	Key Aspect	2017	2018	2019	2020	2021	2022
2017	Publication of Bitcoin Whitepaper by Satoshi Nakamoto	Acknowledgment of the groundbreaking Bitcoin whitepaper	-	-	-	-	-
	Blockchain Development	Bitcoin gains widespread attention as an asset	-	-	-	-	-
	Definition of Blockchain Technology	No industry-wide definition of blockchain	-	-	-	-	-
	China's White Paper on Blockchain Technology	China Blockchain Technology and Application Development White Paper released, providing a definition of blockchain	-	-	-	-	-
2018	Emergence of Smart Contracts	Smart contracts gain prominence as digitized contractual terms	-	-	-	-	-
	Concept of Blockchain 2.0	Integration of blockchains and smart contracts discussed	-	-	-	-	-
2019	Three Horizontal Structure of Blockchain	Introduction of the three-horizontal structure of blockchain	-	-	-	-	-
	Development Stages of Blockchain	Blockchain 1.0 (transaction layer) and Blockchain 2.0 (smart contracts) discussed	-	-	-	-	-
2020	Privacy Layers in Blockchain Design	Privacy layers in blockchain design explored	-	-	-	-	-
	Scarcity of Authentic Use Cases	Immaturity of blockchain technology leads to a lack of authentic use cases	-	-	-	-	-
2021	Focus on Data Analysis and Types	Emphasis on data analysis in blockchain and description of data types	-	-	-	-	-
2022	Distributed Nature of Blockchain	The distributed nature of blockchain across all levels discussed	-	-	-	-	-
	Roles and Responsibilities of Nodes	Different functions of nodes in a blockchain network examined	-	-	-	-	-

To find anti-money laundering measures for Bitcoin, numerous experiments were conducted in the literature using blockchain data by participating in mixing services provided by Bitcoin Fog, Bit

Laundry, and Blockchain.info. While transactions can still be tracked, mixing services based on Bitcoin Fog and Blockchain.info are no longer traceable. Furthermore, based on the characteristics of the mixed coins discovered, the authors propose several relevant anti-money laundering measures. Apart from money laundering through currency mixing services, the research paper [69] summarizes three typical transaction modes suspected of being involved in money laundering: aggregations, folding, and splits. Aggregations refer to transferring the balances of multiple related bitcoins to the same address; folding involves mixing illegal addresses with other normal addresses for aggregation transactions; splitting refers to transferring bitcoins from one address to multiple different addresses. Fraud is another significant issue due to the natural attributes of blockchain, such as participant anonymity, lack of border restrictions, and financial payment, combined with the relatively lagging laws and regulations. Various illegal activities utilizing blockchain platforms like Bitcoin and Ethereum, including transactions in black markets and Ponzi schemes, have become rampant. These activities pose a substantial challenge to national market supervision and result in significant losses for investors participating in the blockchain. In May 2017, the U.S. Securities and Exchange Commission (SEC) filed fraud charges against a Ponzi scheme operator suspected of digital currency and imposed a $74 million fine. On September 4, 2017, the People's Bank of China and seven financial regulators jointly issued a ban on Initial Coin Offerings (ICOs), mainly due to the potential involvement of financial fraud. These examples illustrate the prevalence of illegal activities based on blockchain technology. Hence, it is a pressing and practical research problem to study how to identify potential scams through blockchain data analysis. Ponzi schemes exploit new investors' funds to repay existing investors, and scammers generate income from fees and other sources. The popularity of smart contracts has led to an increase in Ponzi schemes. These scams are more deceptive due to the reliability and automated execution offered by blockchain-based contracts. By examining the source code of several free software contracts and scanning the network, over 100 smart contract-based Pyramid schemes based on Ethereum were uncovered. The normalized Levenshtein proximity method was used to identify similarities and partially hidden Ponzi schemes. SAD-Ponzi, a semantic-aware detection approach for Ethereum smart contract Ponzi frauds, was developed. The researchers presented a heuristically assisted symbolic execution technique to uncover investor-related transfer behaviours or distribution strategies. In a controlled experiment on a well-labelled baseline, SAD-Ponzi achieved flawless precision and recall, surpassing all other computer algorithms. Using SAD-Ponzi, all 3.4 million Ethereum smart contracts distributed by externally-owned accounts (EOAs) were examined, revealing 835 Ponzi schemes with $17 million in victim investments. This research highlights the importance of identifying and eliminating blockchain-based Ponzi schemes [67]-[71].

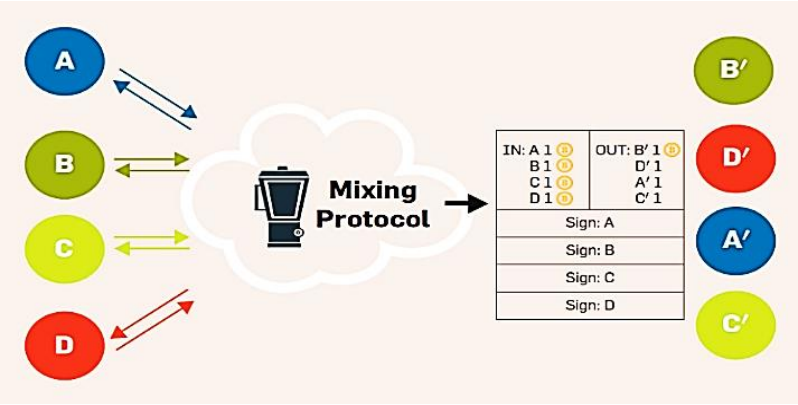

Fig. 6. Coin mixing example

5. Discussion

The technology known as blockchain is still relatively new, yet it has already shown the potential to revolutionize a wide variety of sectors. While the primary applications of blockchain technology, such as Bitcoin and smart contracts, are currently the most prominent, the future of blockchain holds

immense promise in various fields. Consequently, data analysis based on blockchain will also exhibit the diversity of project goals and the uniqueness of the technology. In this section, we will present three potential avenues for future research, along with the challenges associated with each direction. The Limitations of Blockchain can be clarified as Blockchain-based businesses and applications have significant potential. The implementation of blockchain technology, however, presents a number of obstacles. The utilisation of blockchain technology in Bitcoin presents numerous advantages as well, whereas, one of these involves eliminating the necessity for transactions to undergo authorization for the individuals who possess an interest in investing in cryptocurrencies have the capability to do so independently, without relying on the guidance or expertise of traditional financial advisors. Furthermore, the utilisation of a debit or credit card does not necessitate the involvement of a banking institution, it is feasible to engage in borrowing and lending monetary resources even in the absence of banks. In instances where expeditious completion of a transaction is required, the involvement of upper-level management inside the banking institution is not deemed essential. This advantage proves to be quite advantageous for public blockchains.

6. Conclusion

Despite being a relatively young technology, blockchain has demonstrated its potential to disrupt various industries. The introduction of blockchain 1.0 with Bitcoin and subsequent developments like blockchain 2.0 with Ethereum have expanded the scope of data that can be stored on blockchain networks. While the growing popularity of blockchain has raised concerns regarding user privacy and illegal financial activities, it has also facilitated significant technological advancements. The unique characteristic of blockchain technology, which utilizes cryptographic hashes to record transactions, presents an unprecedented opportunity for academic researchers to analyse publicly available blockchain data. This invaluable resource enables researchers to address challenges in ways that were previously unattainable. Hence, it is crucial to highlight the research problems, achievements, and potential avenues for future exploration and challenges in the field of blockchain data analysis. This article provides a comprehensive overview of blockchain data analysis by examining fundamental data types, analytical techniques, architecture, and operations within the blockchain ecosystem. The analysis of blockchain data identifies seven primary research challenges: entity recognition, privacy risk analysis, network representation, network visualization, market impact, and transaction pattern recognition. These challenges serve as focal points for further investigation and development. Looking ahead, future research directions, opportunities, and challenges in blockchain data analysis are discussed based on the identified gaps in previous studies. By addressing these research gaps and harnessing the vast potential of blockchain data, researchers can contribute to advancing the field and gaining deeper insights into the complexities of blockchain technology. In conclusion, this review underscores the importance of blockchain data analysis and its implications in academia. Through the exploration of research challenges, achievements, and prospects, researchers can enhance their understanding of and address the intricacies associated with blockchain technology, ultimately fostering innovation and progress in this dynamic field.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. Widayanti, E. P. Harahap, N. Lutfiani, F. P. Oganda, and I. S. P. Manik, “The impact of blockchain technology in higher education quality improvement,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 2, pp. 207-216, 2021, https://doi.org/10.26555/jiteki.v7i2.20684.

[2] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," *2017 IEEE International Congress on Big Data (BigData Congress)*, pp. 557-564, 2017, https://doi.org/10.1109/BigDataCongress.2017.85.

[3] J. Sidhu, "Syscoin: A Peer-to-Peer Electronic Cash System with Blockchain-Based Services for E-Business," *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1-6, 2017, https://doi.org/10.1109/ICCCN.2017.8038518.

[4] A. Rejeb, K. Rejeb, and J. G. Keogh, "Cryptocurrencies in modern finance: a literature review," *Etikonomi*, vol. 20, no. 1, pp. 93-118, 2021, https://doi.org/10.15408/etk.v20i1.16911.

[5] J. Chevallier, D. Guégan, and S. Goutte, "Is it possible to forecast the price of bitcoin?," *Forecasting*, vol. 3, no. 2, pp. 377-420, 2021, https://doi.org/10.3390/forecast3020024.

[6] A. A. Monrat, O. Schelén, and K. Andersson, "A Survey of Blockchain from the Perspectives of Applications, Challenges, and Opportunities," in *IEEE Access*, vol. 7, pp. 117134-117151, 2019, https://doi.org/10.1109/ACCESS.2019.2936094.

[7] M. Niranjanamurthy, B. N. Nithya, and S. J. C. C. Jagannatha, "Analysis of Blockchain technology: pros, cons and SWOT," *Cluster Computing*, vol. 22, pp. 14743-14757, 2019, https://doi.org/10.1007/s10586-018-2387-5.

[8] X. Xu *et al*., "A Taxonomy of Blockchain-Based Systems for Architecture Design," *2017 IEEE International Conference on Software Architecture (ICSA)*, pp. 243-252, 2017, https://doi.org/10.1109/ICSA.2017.33.

[9] R. Lai and D. L. K. Chuen, "Blockchain–from public to private," in *Handbook of Blockchain, Digital Finance, and Inclusion*, vol. 2, pp. 145-177, 2018, https://doi.org/10.1016/B978-0-12-812282-2.00007-3.

[10] B. S. Rawal, G. Manogaran, and M. Hamdi, "Multi-tier stack of blockchain with proxy re-encryption method scheme on the Internet of things platform," *ACM Transactions on Internet Technology (TOIT)*, vol. 22, no. 2, pp. 1-20, 2021, https://doi.org/10.1145/3421508.

[11] M. Salimitari, M. Chatterjee, and Y. P. Fallah, "A survey on consensus methods in blockchain for resource-constrained IoT networks," *Internet of Things*, vol. 11, p. 100212, 2020, https://doi.org/10.1016/j.iot.2020.100212.

[12] A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives," *Cryptography*, vol. 3, no. 1, p. 3, 2019, https://doi.org/10.3390/cryptography3010003.

[13] B. Sundarakani, A. Ajaykumar, and A. Gunasekaran, "Big data driven supply chain design and applications for blockchain: Action research using case study approach," *Omega*, vol. 102, p.102452, 2021, https://doi.org/10.1016/j.omega.2021.102452.

[14] R. Cole, M. Stevenson, and, J. Aitken, "Blockchain technology: implications for operations and supply chain management," *Supply Chain Management*, vol. 24, no. 4, pp. 469-483, 2019, https://doi.org/10.1108/SCM-09-2018-0309.

[15] R. L. Rana, C. Tricaseand, and L. De Cesare, "Blockchain technology for a sustainable agri-food supply chain," *British Food Journal*, vol. 123 no. 11, pp. 3471-3485, 2021, https://doi.org/10.1108/BFJ-09-2020-0832.

[16] T. Clohessy and T. Acton, "Investigating the influence of organizational factors on blockchain adoption: An innovation theory perspective," *Industrial Management & Data Systems*, vol. 119, no. 7, pp. 1457-1491, 2019, https://doi.org/10.1108/IMDS-08-2018-0365.

[17] S. K. Kim and J. H. Huh, "Artificial neural network blockchain techniques for healthcare system: Focusing on the personal health records," *Electronics*, vol. 9, no. 5, p. 763, 2020, https://doi.org/10.3390/electronics9050763.

[18] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics and Informatics*, vol. 36, pp. 55-81, 2019, https://doi.org/10.1016/j.tele.2018.11.006.

[19] W. Song *et al*., "Blockchain Data Analysis from the Perspective of Complex Networks: Overview," in *Tsinghua Science and Technology*, vol. 28, no. 1, pp. 176-206, 2023, https://doi.org/10.26599/TST.2021.9010080.

[20] H. Taherdoost, "Blockchain technology and artificial intelligence together: a critical review on applications," *Applied Sciences*, vol. 12, no. 24, p. 12948, 2022, https://doi.org/10.3390/app122412948.

[21] R. Skiba, "Blockchain Technology as a Health and Safety Contributor in the Transport and Logistics Industry – Human Resource Requirements," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 4, pp. 544-550, 2020, https://doi.org/10.38124/IJISRT20APR685.

[22] A. G. Aliyev, "Study of Development Trends and Application Risks of Cryptocurrency and Blockchain Technologies in the Digital Environment," *Informatica Economica*, vol. 26, no. 3, pp. 37-49, 2022, https://doi.org/10.24818/issn14531305/26.3.2022.04.

[23] F. Corradi and P. Höfner, "The disenchantment of Bitcoin: unveiling the myth of a digital currency," *International Review of Sociology*, vol. 28, no. 1, pp. 193-207, 2018, https://doi.org/10.1080/03906701.2018.1430067.

[24] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten, "SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies," *2015 IEEE Symposium on Security and Privacy*, pp. 104-121, 2015, https://doi.org/10.1109/SP.2015.14.

[25] X. Yu, C. Tang, P. Palensky, and A. W. Colombo, "Blockchain: What Does It Mean to Industrial Electronics? Technologies, Challenges, and Opportunities," in *IEEE Industrial Electronics Magazine*, vol. 16, no. 2, pp. 4-14, 2022, https://doi.org/10.1109/MIE.2021.3066332.

[26] A. G. Gad, D. T. Mosa, L. Abualigah, and A. A. Abohany, "Emerging trends in blockchain technology and applications: A review and outlook," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6719-6742, 2022, https://doi.org/10.1016/j.jksuci.2022.03.007.

[27] R. Huo *et al*., "A Comprehensive Survey on Blockchain in Industrial Internet of Things: Motivations, Research Progresses, and Future Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 88-122, 2022, https://doi.org/10.1109/COMST.2022.3141490.

[28] B. Bhushan, C. Sahoo, P. Sinha, and A. Khamparia, "Unification of Blockchain and Internet of Things (BIoT): requirements, working model, challenges and future directions," *Wireless Networks*, vol. 27, pp. 55-90, 2021, https://doi.org/10.1007/s11276-020-02445-6.

[29] A. Shahid, A. Almogren, N. Javaid, F. A. Al-Zahrani, M. Zuair, and M. Alam, "Blockchain-Based Agri-Food Supply Chain: A Complete Solution," in *IEEE Access*, vol. 8, pp. 69230-69243, 2020, https://doi.org/10.1109/ACCESS.2020.2986257.

[30] K. Peng, M. Li, H. Huang, C. Wang, S. Wan, and K. -K. R. Choo, "Security Challenges and Opportunities for Smart Contracts in Internet of Things: A Survey," in *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12004-12020, 2021, https://doi.org/10.1109/JIOT.2021.3074544.

[31] Z. Wenhua, F. Qamar, T. A. N. Abdali, R. Hassan, S. T. A. Jafri, and Q. N. Nguyen, "Blockchain technology: security issues, healthcare applications, challenges and future trends," *Electronics*, vol. 12, no. 3, pp. 546, 2023, https://doi.org/10.3390/electronics12030546.

[32] S. Dustdar, P. Fernández, J. M. García, and A. Ruiz-Cortés, "Elastic Smart Contracts in Blockchains," in *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 12, pp. 1901-1912, 2021, https://doi.org/10.1109/JAS.2021.1004222.

[33] W. Yang, E. Aghasian, S. Garg, D. Herbert, L. Disiuta, and B. Kang, "A Survey on Blockchain-Based Internet Service Architecture: Requirements, Challenges, Trends, and Future," in *IEEE Access*, vol. 7, pp. 75845-75872, 2019, https://doi.org/10.1109/ACCESS.2019.2917562.

[34] Z. Zhu, G. Qi, M. Zheng, J. Sun, and Y. Chai, "Blockchain-based consensus checking in decentralized cloud storage," *Simulation Modelling Practice and Theory*, vol. 102, pp. 101987, 2020, https://doi.org/10.1016/j.simpat.2019.101987.

[35] R. Paulavičius, S. Grigaitis, and E. Filatovas, "A Systematic Review and Empirical Analysis of Blockchain Simulators," in *IEEE Access*, vol. 9, pp. 38010-38028, 2021, https://doi.org/10.1109/ACCESS.2021.3063324.

[36] H. Liu, X. Luo, H. Liu and X. Xia, "Merkle Tree: A Fundamental Component of Blockchains," *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*, pp. 556-561, 2021, https://doi.org/10.1109/EIECS53707.2021.9588047.

[37] H. Taherdoost, "Smart Contracts in Blockchain Technology: A Critical Review," *Information*, vol. 14, no. 2, p. 117, 2023, https://doi.org/10.3390/info14020117.

[38] P. Zheng, Z. Zheng, J. Wu, and H. -N. Dai, "XBlock-ETH: Extracting and Exploring Blockchain Data from Ethereum," in *IEEE Open Journal of the Computer Society*, vol. 1, pp. 95-106, 2020, https://doi.org/10.1109/OJCS.2020.2990458.

[39] S. Linoy, N. Stakhanova, and A. Matyukhina, "Exploring Ethereum's Blockchain Anonymity Using Smart Contract Code Attribution," *2019 15th International Conference on Network and Service Management (CNSM)*, pp. 1-9, 2019, https://doi.org/10.23919/CNSM46954.2019.9012681.

[40] F. Contro, M. Crosara, M. Ceccato, and M. D. Preda, "EtherSolve: Computing an Accurate Control-Flow Graph from Ethereum Bytecode," *2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC)*, pp. 127-137, 2021, https://doi.org/10.1109/ICPC52881.2021.00021.

[41] X. Chao, Q. Ran, J. Chen, T. Li, Q. Qian, and D. Ergu, "Regulatory technology (Reg-Tech) in financial stability supervision: Taxonomy, key methods, applications and future directions," *International Review of Financial Analysis*, vol. 80, p. 102023, 2022, https://doi.org/10.1016/j.irfa.2022.102023.

[42] R. Bandara, M. Fernando, and S. Akter, "Privacy concerns in E-commerce: A taxonomy and a future research agenda," *Electron Markets*, vol. 30, pp. 629-647, 2020, https://doi.org/10.1007/s12525-019-00375-6.

[43] S. Phetsouvanh, A. Datta, and F. Oggier, "Analysis of multi-input multi-output transactions in the Bitcoin network," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 1, 2021, https://doi.org/10.1002/cpe.5629.

[44] X. He, K. He, S. Lin, J. Yang, and H. Mao, "Bitcoin address clustering method based on multiple heuristic conditions," *IET Blockchain*, vol. 2, no. 2, pp. 44-56, 2022, https://doi.org/10.1049/blc2.12014.

[45] Y. He, H. Dong, H. Wu, and Q. Duan, "Formal Analysis of Reentrancy Vulnerabilities in Smart Contract Based on CPN," *Electronics*, vol. 12, no. 10, p. 2152, 2023, https://doi.org/10.3390/electronics12102152.

[46] 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," *2018 51st Hawaii International Conference on System Sciences (HICSS)*, pp. 5647-5656, 2018, https://doi.org/10.24251/HICSS.2018.443.

[47] M. Coscia, F. Giannotti, and D. Pedreschi, "A classification for community discovery methods in complex networks," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 5, pp. 512-546, 2011, https://doi.org/10.1002/sam.10133.

[48] R. R. Tubino, C. Robardet, and R. Cazabet, "Towards a better identification of Bitcoin actors by supervised learning," *Data and Knowledge Engineering*, vol. 142, p. 102094, 2022, https://doi.org/10.1016/j.datak.2022.102094.

[49] 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," in *Proceedings of the 2013 conference on Internet measurement conference*, pp. 127-140, 2013, https://doi.org/10.1145/2504730.2504747.

[50] D. Goldsmith, K. Grauer, and Y. Shmalo, "Analyzing hack subnetworks in the bitcoin transaction graph," *Appl. Netw. Sci*., vol. 5, no. 22, 2020, https://doi.org/10.1007/s41109-020-00261-7.

[51] L. Peng, W. Feng, Z. Yan, Y. Li, X. Zhou, and S. Shimizu, "Privacy preservation in permissionless blockchain: A survey," *Digital Communications and Networks*, vol. 7, no. 3, pp. 295-307, 2021, https://doi.org/10.1016/j.dcan.2020.05.008.

[52] J. Newell, Q. Mamun, and M. Z. Islam, "A Generalised Logical Layered Architecture for Blockchain Technology," *arXiv preprint arXiv:2110.09615*, 2021, https://doi.org/10.48550/arXiv.2110.09615.

[53] P. Kayal and P. Rohilla, "Bitcoin in the economics and finance literature: a survey," *SN Business &*

Economics, vol. 1, no. 88, 2021, https://doi.org/10.1007/s43546-021-00090-5.

[54] E. Saiedi, A. Broström, and F. Ruiz, "Global drivers of cryptocurrency infrastructure adoption," *Small Business Economics*, vol. 57, pp. 353-406, 2021, https://doi.org/10.1007/s11187-019-00309-8.

[55] P. Weichbroth, K. Wereszko, H. Anacka, and J. Kowal, "Security of Cryptocurrencies: A View on the State-of-the-Art Research and Current Developments," *Sensors*, vol. 23, no. 6, p. 3155, 2023, https://doi.org/10.3390/s23063155.

[56] D. Kondor, N. Bulatovic, J. Stéger, I. Csabai, and G. Vattay, "The rich still get richer: Empirical comparison of preferential attachment via linking statistics in Bitcoin and Ethereum," *Frontiers in Blockchain*, vol. 4, p. 668510, 2021, https://doi.org/10.3389/fbloc.2021.668510.

[57] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information," in *IEEE Access*, vol. 6, pp. 5427-5437, 2018, https://doi.org/10.1109/ACCESS.2017.2779181.

[58] Y. Sun, H. Xiong, S. M. Yiu, and K. Y. Lam, "BitAnalysis: A Visualization System for Bitcoin Wallet Investigation," in *IEEE Transactions on Big Data*, vol. 9, no. 2, pp. 621-636, 2023, https://doi.org/10.1109/TBDATA.2022.3188660.

[59] E. A. Obukhova, "ICO as a modern method for financing high-tech projects," *Problems of Economic Transition*, vol. 62, no. 4-6, pp. 249-260, 2020, https://doi.org/10.1080/10611991.2020.1968745.

[60] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on LightGBM," *Finance Research Letters*, vol. 32, p. 101084, 2020, https://doi.org/10.1016/j.frl.2018.12.032.

[61] H. Halaburda, G. Haeringer, J. Gans, and N. Gandal, "The microeconomics of cryptocurrencies," *Journal of Economic Literature*, vol. 60, no. 3, pp. 971-1013, 2022, https://doi.org/10.1257/jel.20201593.

[62] A. Trozze *et al*., "Cryptocurrencies and future financial crime," *Crime Science*, vol. 11, pp. 1-35, 2022, https://doi.org/10.1186/s40163-021-00163-8.

[63] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants," *International Journal of Forecasting*, vol. 37, no. 1, pp. 28-43, 2021, https://doi.org/10.1016/j.ijforecast.2020.02.008.

[64] B. Zhong, X. Pan, L. Ding, Q. Chen, and X. Hu, "Blockchain-driven integration technology for the AEC industry," *Automation in Construction*, vol. 150, p. 104791, 2023, https://doi.org/10.1016/j.autcon.2023.104791.

[65] H. Mujlid, "A Survey on Machine Learning Approaches in Cryptocurrency: Challenges and Opportunities," *2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1-6, 2023, https://doi.org/10.1109/iCoMET57998.2023.10099130.

[66] X. Cao, J. Zhang, X. Wu, and B. Liu, "A survey on security in consensus and smart contracts," *Peer-to-Peer Networking and Applications*, vol. 15, pp. 1008-1028, 2022, https://doi.org/10.1007/s12083-021-01268-2.

[67] D. Sanz-Bas, C. del Rosal, S. L. N. Alonso, and M. Á. E. Fernández, "Cryptocurrencies and Fraudulent Transactions: Risks, Practices, and Legislation for Their Prevention in Europe and Spain," *Laws*, vol. 10, p. 57, 2021, https://doi.org/10.3390/laws10030057.

[68] Y. Zhang, J. Wang, and J. Luo, "Heuristic-Based Address Clustering in Bitcoin," in *IEEE Access*, vol. 8, pp. 210582-210591, 2020, https://doi.org/10.1109/ACCESS.2020.3039570.

[69] M. Campbell-Verduyn, "Bitcoin, crypto-coins, and global anti-money laundering governance," *Crime, Law and Social Change*, vol. 69, pp. 283-305, 2018, https://doi.org/10.1007/s10611-017-9756-5.

[70] L. J. Trautman, "Bitcoin, virtual currencies, and the struggle of law and regulation to keep peace," *Marquette Law Review*, vol. 102, p. 447, 2018, https://doi.org/10.2139/ssrn.3182867.

[71] W. Chen *et al*., "Sadponzi: Detecting and characterizing ponzi schemes in ethereum smart contracts," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 5, no. 2, pp. 1-30, 2021, https://doi.org/10.1145/3460093.