Graph-based Management and Mining of Blockchain Data

Arijit Khan
Aalborg University
Aalborg, Denmark
arijitk@cs.aau.dk

Cuneyt Gurcan Akcora
University of Manitoba
Winnipeg, Canada
cuneyt.akcora@umanitoba.ca

ABSTRACT

The mainstream adoption of blockchains led to the preparation of many decentralized applications and web platforms, including Web 3.0, a peer-to-peer internet with no single authority. The data stored in blockchain can be considered as big data – massive-volume, dynamic, and heterogeneous. Due to highly connected structure, graph-based modeling is an optimal tool to analyze the data stored in blockchains. Recently, several research works performed graph analysis on the publicly available blockchain data to reveal insights into its business transactions and for critical downstream tasks, e.g., cryptocurrency price prediction, phishing scams and counterfeit token detection. In this tutorial, we discuss relevant literature on blockchain data structures, storage, categories, data extraction and graphs construction, graph mining, topological data analysis, and machine learning methods used, target applications, and the new insights revealed by them, aiming towards providing a clear view of unified graph-data models for UTXO and account-based blockchains. We also emphasize future research directions.

CCS CONCEPTS

• Mathematics of computing → Graph algorithms; **Exploratory data analysis**; **• Applied computing → Digital cash**.

KEYWORDS

blockchain data, graph analysis, cryptocurrency price prediction

ACM Reference Format:
Arijit Khan and Cuneyt Gurcan Akcora. 2022. Graph-based Management and Mining of Blockchain Data. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3511808.3557502

1 INTRODUCTION

A blockchain is a distributed, digital ledger of records stored in a sequential order. Each record or block is time-stamped and is linked to the previous one. These blocks can be shared openly among its participants to create an immutable sequence of transactions. A blockchain is updated by consensus among its users (an open or a controlled set), who participate in a peer-to-peer network. Therefore, a blockchain contains a secure, tamper-proof, and verifiable record of every transaction validated by the system. Due to the proliferation of cryptocurrencies and decentralized applications (dApps), blockchain technology received tremendous attention recently, such as in decentralized finance (DeFi), smart city, Internet-of-Things (IoT), nonfungible tokens (NFTs), stablecoins, supply chain, identity management, healthcare, voting, and gaming [15]. A blockchain is well-suited for Web 3.0 [28], that is a decentralized, secure internet, where individuals engage with each other in economic transactions without the need for a central authority, e.g., a bank or a credit card company.

Public blockchains (e.g., Bitcoin, Ethereum) are permissionless, i.e., allow anyone to join. With "public permissionless" blockchains, we have access to trusted, transparent, comprehensive, and granular datasets of digital economic behaviors. Blockchain data analytics, also called the distributed ledger analytics (DLA), is an emerging field of research. Data stored in a public blockchain can be considered as *big data*: they are **massive-volume** (e.g., Ethereum archive nodes that store a complete snapshot of the Ethereum blockchain, including all the transaction records, take up to 4TB of space[1]), **dynamic** (e.g., Ethereum blockchain processed more than 1.1 million transactions per day in July 2021[2]), and **heterogeneous** (e.g., Ethereum blockchain contains a vast amount of heterogeneous interactions: user-to-user, user-to-contract, contract-to-user, and contract-to-contract across multiple layers, such as external and internal transactions, tokens, dAapps, etc. [18]); thus, data analytic methods can be applied to extract knowledge hidden in the blockchain. Such data can be modeled as **complex**, **dynamic**, **multi-layer**, and even **higher-order networks** [24, 30, 32]. Public blockchain data are widely investigated in several applications, including cryptocurrency price prediction and abuses, address clustering, criminal usage detection, anti-money-laundering, business transactions analysis, and thereby providing new means for financial data mining [5, 6, 16]. They are critical in emerging fields such as blockchain intelligence, blockchain social networks [25], and blockchain search engines, using data analytic and machine learning tools to guide users avoiding transaction risks and frauds.

This tutorial gives a comprehensive introduction to the topic of graph-based data analysis methods and tools for blockchains data. We shall provide a discussion of relevant literature based on blockchain data structures, storage, and categories, data extraction and usage, graphs construction, graph mining, topological data analysis, and machine learning methods used (on top of blockchain graphs), target applications, and the new insights they revealed. We shall conclude by discussing open problems and the road ahead.

Motivation and benefits: Blockchain data must be analyzed for an ever-growing list of critical tasks such as money laundering and price manipulation detection. In recent years, a growing set

[1] decrypt.co/24779/ethereum-archive-nodes-now-take-up-4-terabytes-of-space
[2] statista.com/statistics/730838/number-of-daily-cryptocurrency-transactions-by-type/

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557502

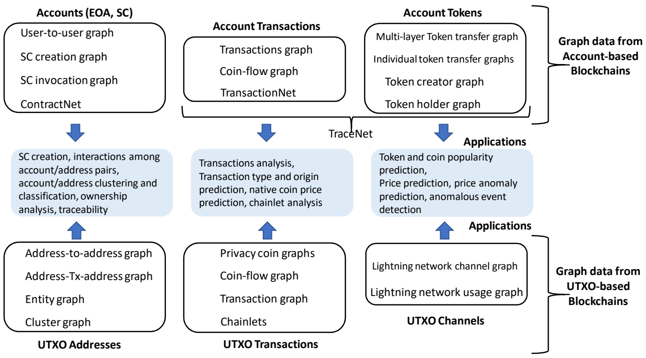

Figure 1: Several graphs derived from interactions in *UTXO* and *account-based* blockchains; and their common applications.

of custom-made tools (e.g., BiVa [26]) and services (e.g., tutela.xyz) have been deployed online with support from the blockchain community to meet the growing needs for blockchain data analytics. Graph-based analysis of crypto-data is an emerging field of study, recently adopted by many institutions, e.g., the USA challenger bank [3], PwC Germany [4], and Elliptic[5] for building new financial services, combating money laundering and terror financing, respectively. However, existing tools work in isolation and do not provide cross-blockchain data-linkage and search capabilities critical in applications such as tracking the chain-hopping behavior of malicious users across cryptocurrencies, asset bridging in Decentralized Finance, and detecting buggy code execution in wrapped assets. We point out that a clear view of unified graph-data models for *UTXO* and *account-based* blockchains (introduced in §3) is lacking. As a result, algorithms and tools that are developed for a specific type of blockchain cannot be easily applied to blockchains with similar data (i.e., transaction and block models). We propose a graph model to unify complex, dynamic, multi-layer, and higher-order network data from diverse blockchains in this premier information and knowledge management conference. Our tutorial also provides the blueprint to develop a unified ecosystem of interconnected services for blockchain data, storage, querying, and analysis capabilities. Our classification and description of data, models, applications, and future directions on blockchains are **timely and critical**.

We present in Figure 1 the **summary diagram** of several important graphs that can be constructed based on interactions between various components of *UTXO* and *account-based* blockchains, such as accounts, transactions, token transfers, and channels; together with their applications. We hope that our summary diagram would be a starting point for future research in this domain.

Target audience and prerequisites. The tutorial is intended for researchers, system designers, data scientists, and practitioners in the broad area of information and knowledge management, complex networks, Web science, machine learning, and financial technology (Fintech). This tutorial does not require any in-depth knowledge on complex graph algorithms and blockchain techniques.

2 OUTLINE OF TUTORIAL

```
1 Introduction
  1.1 Components
    - Ledgers, cryptocurrencies, transactions, users, contracts, tokens, dApps, DeFi,
      stablecoins, channels, use cases
  1.2 Blockchains: Data Structures and Storage
    - Bitcoin, Ethereum, Litecoin, Namecoin, Monero, ZCash, Ripple, IOTA, EOSIO, Steem
  1.3 Categories
     - Public, private, permissionless, permissioned, consortium, hybrid
     - Currency vs. platform; UTXO vs. account; First vs. second layer technologies
2 Data Extraction and Analysis Tools
  - Data Extraction and ETL, Open Datasets, Analytic Tools
3 Graphs Constructed
  3.1 UTXO-based (Bitcoin, Litecoin, Monero, Zcash)
     - Transaction graph, address graph, user graph, k-chainlets, lightning graph
  3.2 Account-based (Ethereum)
      - Trace graph; transaction (money flow) graph
      - (ERC20, ERC721 and ERC1155) token graph(s)
      - token creator, holder, and transfer graphs
      - user-to-user, contract-to-contract, and user-to-contract graphs
      - contract deployment and invocation graphs
      - dApp and DeFi graphs
4 Graph Data Analysis and Machine Learning on Blockchain Graphs
  4.1 Graph Analysis
      - Local and global graph properties analysis
      - Hypergraph and multi-layer graph analysis
      - k-chainlet analysis, clustering heuristics, coin-mixing analysis
  4.2 Topological Data Analysis on Blockchain Graphs
     - persistent homology, simplicial complex, Betti number, functional data depth
     - mapper analysis
  4.3 Machine Learning on Blockchain Graphs
      - Graph representation learning, Chainlets
5 Target Applications
      - Insights into the transaction and token transfers
      - Node classification, link prediction, anomaly detection
      - Temporal price prediction
6 Open Problems
    6.1 dApps and DeFi graphs
    6.2 Individual ERC20 token subnetworks, stablecoin graphs
    6.3 Multi-layer network and hypergraph analysis
    6.4 Users clustering and graph analysis
    6.5 Dynamic graphs and incremental machine learning
```

Length of the tutorial. The intended length of our tutorial is 3 hours. Both tutors will be available to present in-person.

What we shall not cover in this tutorial. We shall not emphasize on applications [15] and distributed databases aspects of blockchains [8, 14, 21], e.g., consensus protocols, confidentiality, fault-tolerance, scalability, and production deployment [22].

3 DESCRIPTION OF TOPICS

• **Introduction.** Bitcoin introduced arguably the first decentralized cryptocurrency [23]. Ethereum started a new way to flourish decentralized applications with its *smart contracts* (SC), which are autonomous agents that can execute complex code across a decentralized network [10]. In addition to ether, its native cryptocurrency, Ethereum blockchain also permits creation and transaction of *tokens*, which are digital assets, through codes defined in the respective smart contracts. Therefore, Ethereum introduces a heterogeneous, financial ecosystem of humans (users) and autonomous agents (smart contracts). Among other important blockchains, we shall introduce Litecoin, Namecoin, Monero, ZCash, Ripple, IOTA, EOSIO, and Steem, together with their ecosystems, data structures, and storage, e.g., hypergraph representation of bitcoin [30], *Merkle Patricia trees* in Ethereum, trust graph in Ripple, and directed acyclic graph in IOTA [4]. We can broadly classify blockchains as (i) private vs. public, (ii) currency-based vs. platform, (iii) UTXO-based vs. account-based, and (iv) first vs. second layer technologies. This tutorial focuses on graph-based analysis of data from public blockchains.

In **account-based transaction model** (e.g., Ethereum), an account can spend a fraction of its coins and keep the remaining balance, similar to bank accounts; thus account-based model updates user balances globally. In the **unspent transaction output**

[3]neo4j.com/news/us-challenger-bank-current-uses-neo4j-graph-technology-to-build-services-centered-around-customer-relationships/
[4]neo4j.com/blog/how-pwc-germany-combats-money-laundering-in-crypto-space-with-neo4j/
[5]elliptic.co/blog/crypto-regulatory-affairs-the-us-shuts-down-3-terror-financing-campaigns

(UTXO) model (e.g., Bitcoin), the entire graph of transaction outputs, spent and unspent, represents the global state. The UTXO model only records transaction receipts, account balances can be calculated by adding up the available unspent transaction outputs.

• **Data extraction and analysis tools.** To get all historic blockchain transactions, one can join the peer-to-peer network through a client, e.g., Bitcoin-Core, Geth, OpenEthereum, and Parity [6] are popular software clients for running a full node. Alternatively, users can also interact with network nodes via the web3 library using managed services, such as Infura, Quicknode, and SoChain.[7] In addition, some well-curated blockchain datasets have also been released, e.g., Google BigQuery [13] and XBlock-ETH [33]. The ETL (extract-transform-load) process converts raw data into convenient formats, such as CSVs, relational databases, and graphs [29].

Participants in blockchain transact under pseudonyms, making it hard to obtain their identity (e.g., exchanges, wallets, frauds, etc.). Such information for many accounts can be collected from Etherscan (etherscan.io) and open forums, e.g., CryptoScamDB (cryptoscamdb.org), they are useful in fraud detection. Ethereum Query Language (EQL) [9] supports SQL-like queries to retrieve information from the Ethereum blockchain data. BiVA is a graph mining tool for the bitcoin network visualization and analysis [25].

Researchers are also working on natural language processing, sentiment analysis using tweets and Google Trends on blockchain [17], providing alternative ways to monitor blockchain ecosystems.

• **Graphs constructed.** UTXO (e.g., Bitcoin, Monero) transaction networks are modeled as address or transaction graphs [4]. Additionally, one can extract *chainlet* substructures [2].

UTXO transaction graphs omit address nodes from the transaction network and create edges among transaction nodes only. The most important aspect of the transaction graph is that a node can appear only once. There will be no future edges that reuse a transaction node, which simplifies graph analysis. By omitting addresses, however, associations between addresses are lost which prevents linking addresses. **UTXO address graph** is the most commonly used graph model for UTXO networks. The address graph omits transactions and creates edges between addresses only. Address nodes may appear multiple times, which implies that addresses may create new transactions or receive coins from new transactions in the future. Address graphs are larger than transaction graphs in node and edge counts. When omitting the intermediate transaction node, we cannot know how to connect input-output address pairs. Thus, one needs to create an edge between every pair, which creates artificial edges in the graph that causes computational overheads.

In Ethereum, Chen et al. [12] studied the **money flow graph** (MFG), **smart contract creation** (CCG), and **invocation graphs** (CIG). MFG is a *weighted, directed graph* denoting transfer of ether between accounts. An weight denotes the total amount of ether transferred along that edge via one or more transactions. CCG, which deals with smart contracts creation, is a *forest* having multiple trees. The root of every tree is an EOA (externally owned account), other nodes of the tree are smart contract accounts that are directly or indirectly created by that EOA. Thus, the edges in CCG are unidirectional. In contrast, CIG is a *weighted, directed graph*; an edge indicates an invocation of a smart contract, either by an EOA or by another smart contract; the edge weight counts the number of invocations, via one or more transactions.

Lee et al. [18] derived four networks: (a) **TraceNet**, consisting of all successful traces with non-null from/to addresses as edges; (b) **ContractNet**, a subgraph of TraceNet, where only those edges with both from_address and to_address belonging to smart contracts, are retained; (c) **TransactionNet**, whose edges are formed by external transactions (this is similar to the money flow graph in [12]); and (d) **TokenNet**, based on explicit transfer of tokens. In [32], Lin et al. studied temporal variations of the four networks.

• **Graph analysis, topological data analysis, and machine learning methods on blockchain graphs.** Majority of the works conducted graph analysis by measuring graph properties, which can be classified as: (a) global properties, also known as "summary features", and (b) local properties of individual nodes and edges [18]. Important local properties analyzed on blockchain graphs are node degree distribution, node centrality measures such as degree, closeness, betweenness, PageRank, and Eigenvector centrality. Among global properties, most prominent ones studied are connected components, reciprocity, assortativity, maximum clique, core decomposition, density, triangle and motif counts, and diameter.

Another emerging approach for ransomware payment detection is topological data analysis (TDA). TDA systematically infers qualitative and quantitative geometric and topological structures of blockchain transaction graphs at multiple resolutions [1, 19]. As a result, TDA allows us to capture subtler patterns in the transaction graphs, including changes in chainlet dynamics, which are often associated with illicit or malicious activity and which are inaccessible with more conventional methods based on various forms of information aggregation [7, 24]. Both [19, 24] conducted topological data analysis on Ethereum networks for anomaly detection, the key concepts include simplicial complex, persistent homology, Betti number, functional data depth, and stacked persistence diagram.

Several works performed graph embedding with blockchain graphs. Lin et al. [20], Poursafaei et al. [27], and Wu et al. [31] designed temporal, node, and edge features-biased random walks for graph representation learning. Chen et al. [11] performed graph convolutional neural network (GCN)-based node embedding.

• **Target applications.** Bulk of the works conducted graph analysis to gain insights into transaction and token transfers. Some of them considered downstream tasks, e.g., node classification, link prediction, anomaly detection, token price prediction. Most tools for blockchain data are related to e-crime or financial (e.g., price, investor) analytics. From ransomware payment detection [7] to sextortion discovery [25], transaction graph analysis has proven useful to study blockchain address importance and to cluster them.

• **Future directions.** We conclude by discussing open problems.

(1) Accounts interact with each other based on different dApps and DeFi protocols, thus forming graph structures. One can investigate their graph properties, embeddings, and anomalous patterns. A specific interest could be analyzing stablecoins' price stabilization mechanisms and understanding the recent crash of Luna Terra [8].

(2) There are relatively less works on graph analysis of individual ERC20 token subnetworks. One may correlate their graph properties with token price and popularity for more accurate forecasting.

[6] geth.ethereum.org openethereum.github.io parity.io/technologies/ethereum

[7] infura.io quicknode.com https://chain.so

[8] bloomberg.com/news/articles/2022-05-19/luna-terra-collapse-reveal-crypto-price-volatility

(3) Due to several modes of interactions among EOAs and contracts, such as external and internal transactions, token transfers, dApps and DeFi usage, one may construct a multi-layer network, where each layer will denote one specific mode of interaction. Multi-layer graphs would be an expressive model of real-world activities.

(4) Accounts can be grouped into various categories and granularity, e.g., miners, mining pools, mixers, exchanges, phishing accounts, ICO contracts, gambling games, etc. One can conduct graph analysis in an OLAP (online analytical processing) manner, by drilling-up/down based on hierarchical categories.

(5) Due to dynamic nature of accounts and transactions, employed ML models must deal with data drifts. Incremental learning, machine unlearning, and continuous learning would be useful.

4 RELATED TUTORIALS

• **Blockchain data analytics.** Akcora, Gel, and Kantarcioglu have given blockchain tutorials at PaKDD, ICDM, ICDE, SDM, and KDD between 2018 and 2021 [3]. These tutorials covered fundamental building blocks of blockchains and data structures of UTXO and account blockchains. However, unlike ours, these tutorials have not proposed unified graph models. Our tutorial also covers data extraction and analysis, as well as the existing methodology in blockchain data analytics. Furthermore, we cover the state-of-the-art in graph analysis, topological data analysis, and graph machine learning, which have seen a considerable body of new work recently.

• **Databases and distributed systems aspects of blockchains.** Several tutorials [8, 14, 21] were presented emphasizing databases and distributed systems aspects of blockchains, such as consensus protocols, confidentiality, verifiability, fault-tolerance, scalability, data management applications, deployment, and benchmarking. Our tutorial is different since we shall discuss the graph-based analysis of transactions and other data stored on public blockchains.

5 PRESENTER BIOGRAPHY

Arijit Khan is an Associate Professor in the Department of Computer Science, Aalborg University, Denmark. He earned his PhD from UC Santa Barbara, USA and did a post-doc at ETH Zurich, Switzerland. He has been an assistant professor at NTU Singapore. Arijit is the recipient of the IBM PhD Fellowship (2012-13). He published several papers in premier data management venues, e.g., SIGMOD, VLDB, TKDE, ICDE, WWW, SDM, EDBT, CIKM, WSDM, and TKDD. Arijit co-presented tutorials on graph queries and systems at ICDE (2012) and VLDB (2014, 2015, 2017), and served in the program committee of KDD, SIGMOD, VLDB, ICDE, ICDM, EDBT, SDM, CIKM, AAAI, WWW, and as an associate editor of TKDE.

Cuneyt G. Akcora is an Assistant Professor of Computer Science and Statistics at the University of Manitoba, Canada. He received his Ph.D. from the University of Insubria, Italy, and completed his postdoctoral studies at the University of Texas at Dallas. He is a Fulbright Scholarship recipient, and published in leading venues including IEEEtran, KDD, VLDB, ICDM, SDM, IJCAI, and ICDE. Cuneyt co-presented tutorials on Blockchain Data Analytics at ICDE, ICDM, PaKDD, KDD, and SDM and served on the program committees of data management, AI, and ML conferences, e.g., ICDE, ICLR, ICML, KDD, and AAAI.

ACKNOWLEDGMENTS

This work was supported by the Novo Nordisk Foundation grant NNF22OC0072415.

REFERENCES

[1] N. C. Abay, C. G. Akcora, Y. R. Gel, M. Kantarcioglu, U. D. Islambekov, Y. Tian, and B. M. Thuraisingham. 2019. ChainNet: Learning on Blockchain Graphs with Topological Features. In *ICDM*.
[2] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu. 2018. Forecasting Bitcoin Price with Graph Chainlets. In *PAKDD*.
[3] C. G. Akcora, M. F. Dixon, Y. R. Gel, and M. Kantarcioglu. 2018. Blockchain Data Analytics. *Intelligent Informatics* (2018), 4.
[4] C. G. Akcora, Y. R. Gel, and M. Kantarcioglu. 2022. Blockchain Networks: Data Structures of Bitcoin, Monero, Zcash, Ethereum, Ripple, and Iota. *WIREs Data Mining Knowl. Discov.* 12, 1 (2022).
[5] C. G. Akcora, M. Kantarcioglu, and Y. R. Gel. 2018. Blockchain Data Analytics. In *ICDM*.
[6] C. G. Akcora, M. Kantarcioglu, and Y. R. Gel. 2021. Data Science on Blockchains. In *KDD*.
[7] C. G. Akcora, Y. Li, Y. R. Gel, and M. Kantarcioglu. 2020. BitcoinHeist: Topological Data Analysis for Ransomware Prediction on the Bitcoin Blockchain. In *IJCAI*.
[8] M. J. Amiri, D. Agrawal, and A. E. Abbadi. 2021. Permissioned Blockchains: Properties, Techniques and Applications. In *SIGMOD*.
[9] S. Bragagnolo, H. Rocha, M. Denker, and S. Ducasse. 2018. Ethereum Query Language. In *WETSEB@ICSE*.
[10] V. Buterin. 2013. Ethereum Whitepaper. https://ethereum.org/en/whitepaper/.
[11] L. Chen, J. Peng, Y. Liu, J. Li, F. Xie, and Z. Zheng. 2021. Phishing Scams Detection in Ethereum Transaction Network. *ACM Trans. Internet Technol.* 21, 1 (2021), 10:1–10:16.
[12] T. Chen, Y. Zhu, Z. Li, J. Chen, X. Li, X. Luo, X. Lin, and X. Zhang. 2018. Understanding Ethereum via Graph Analysis. In *INFOCOM*.
[13] A. Day and E. Medvedev. 2018. Ethereum-ETL. cloud.google.com/blog/products/data-analytics/ethereum-bigquery-public-dataset-smart-contract-analytics.
[14] S. Gupta, J. Hellings, S. Rahnama, and M. Sadoghi. 2020. Building High Throughput Permissioned Blockchain Fabrics: Challenges and Opportunities. *PVLDB* 13, 12 (2020), 3441–3444.
[15] H. Huang, W. Kong, S. Zhou, Z. Zheng, and S. Guo. 2021. A Survey of State-of-the-Art on Blockchains: Theories, Modelings, and Tools. *ACM Comput. Surv.* 54, 2 (2021), 44:1–44:42.
[16] A. Kamišalić, R. Kramberger, and I. Fister. 2021. Synergy of Blockchain Technology and Data Mining Techniques for Anomaly Detection. *Appl. Sciences* 11, 17 (2021).
[17] O. Kraaijeveld and J. D. Smedt. 2020. The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices. *Journal of International Financial Markets, Institutions and Money* 65, C (2020), S104244312030072X.
[18] X. T. Lee, A. Khan, S. S. Gupta, Y. H. Ong, and X. Liu. 2020. Measurements, Analyses, and Insights on the Entire Ethereum Blockchain Network. In *WWW*.
[19] Y. Li, U. Islambekov, C. G. Akcora, E. Smirnova, Y. R. Gel, and M. Kantarcioglu. 2020. Dissecting Ethereum Blockchain Analytics: What We Learn from Topology and Geometry of the Ethereum Graph?. In *SDM*.
[20] D. Lin, J. Wu, Q. Yuan, and Z. Zheng. 2020. Modeling and Understanding Ethereum Transaction Records via a Complex Network Approach. *IEEE Trans. Circuits Syst.* 67-II, 11 (2020), 2737–2741.
[21] S. Maiyya, V. Zakhary, M. J. Amiri, D. Agrawal, and A. E. Abbadi. 2019. Database and Distributed Computing Foundations of Blockchains. In *SIGMOD*.
[22] C. Mohan. 2019. State of Public and Private Blockchains: Myths and Reality. In *SIGMOD*.
[23] S. Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System.
[24] D. Ofori-Boateng, I. Segovia-Dominguez, C. G. Akcora, M. Kantarcioglu, and Y. R. Gel. 2021. Topological Anomaly Detection in Dynamic Multilayer Blockchain Networks. In *ECML PKDD*.
[25] F. E. Oggier, A. Datta, and S. Phetsouvanh. 2020. An Ego Network Analysis of Sextortionists. *Soc. Netw. Anal. Min.* 10, 1 (2020), 44.
[26] F. E. Oggier, S. Phetsouvanh, and A. Datta. 2018. BiVA: Bitcoin Network Visualization & Analysis. In *ICDM Workshops*.
[27] F. Poursafaei, R. Rabbany, and Z. Zilic. 2021. SigTran: Signature Vectors for Detecting Illicit Activities in Blockchain Transaction Networks. In *PAKDD*.
[28] M. Ragnedda and G. Destefanis. 2020. *Blockchain and Web 3.0: Social, Economic, and Technological Challenges.* New York: Routledge.
[29] V. H. Su, S. S. Gupta, and A. Khan. 2022. Automating ETL and Mining of Ethereum Blockchain Network. In *WSDM*.
[30] J. Wu, J. Liu, Y. Zhao, and Z. Zheng. 2021. Analysis of Cryptocurrency Transactions from a Network Perspective: An Overview. *J. Netw. Comput. Appl.* 190 (2021), 103139.
[31] J. Wu, Q. Yuan, D. Lin, W. You, W. Chen, C. Chen, and Z. Zheng. 2022. Who Are the Phishers? Phishing Scam Detection on Ethereum via Network Embedding. *IEEE Trans. Syst. Man Cybern. Syst.* 52, 2 (2022), 1156–1166.
[32] L. Zhao, S. S. Gupta, A. Khan, and R. Luo. 2021. Temporal Analysis of the Entire Ethereum Blockchain Network. In *WWW*.
[33] P. Zheng, Z. Zheng, J. Wu, and H.-N. Dai. 2020. XBlock-ETH: Extracting and Exploring Blockchain Data From Ethereum. *IEEE Open J. Comp. Soc.* 1 (2020).