

Machine Learning for Blockchain Data Analysis: Progress and Opportunities

Poupak Azad¹, Cuneyt Gurcan Akcora², Arijit Khan³

¹University of Manitoba, Canada

²University of Central Florida, USA

³Aalborg University, Denmark

azadp@myumanitoba.ca, cuneyt.akcora@ucf.edu, arijitk@cs.aau.dk

Abstract

Blockchain technology has rapidly emerged to mainstream attention, while its publicly accessible, heterogeneous, massive-volume, and temporal data are reminiscent of the complex dynamics encountered during the last decade of big data. Unlike any prior data source, blockchain datasets encompass multiple layers of interactions across real-world entities, e.g., human users, autonomous programs, and smart contracts. Furthermore, blockchain’s integration with cryptocurrencies has introduced financial aspects of unprecedented scale and complexity such as decentralized finance, stablecoins, non-fungible tokens, and central bank digital currencies. These unique characteristics present both opportunities and challenges for machine learning on blockchain data.

On one hand, we examine the state-of-the-art solutions, applications, and future directions associated with leveraging machine learning for blockchain data analysis critical for the improvement of blockchain technology such as e-crime detection and trends prediction. On the other hand, we shed light on the pivotal role of blockchain by providing vast datasets and tools that can catalyze the growth of the evolving machine learning ecosystem. This paper serves as a comprehensive resource for researchers, practitioners, and policy-makers, offering a roadmap for navigating this dynamic and transformative field.

1 Introduction

Blockchain, originally designed as the underlying technology for cryptocurrencies, e.g., Bitcoin [Nakamoto, 2008], has evolved into a robust framework for recording and verifying transactions. Its inherent features, including decentralization and cryptographic security, make it an ideal candidate for myriad applications beyond finance, such as internet-of-things, healthcare, and smart city. One of the most intriguing aspects of blockchain is its ability to generate vast and publicly accessible datasets, containing records of transactions involving diverse real-life entities and autonomous agents.

Simultaneously, the field of machine learning (ML) is experiencing an exponential surge in its application to data analysis across domains, thanks to deep neural methods and artificial general intelligence. ML and deep learning algorithms, capable of discerning patterns, trends, and anomalies within vast datasets, have proven indispensable for extracting meaningful insights and enabling predictions from complex data in an automated and end-to-end manner.

The importance of Blockchain is increasingly felt as the United Nations, through its Innovation Fund, has committed substantial resources (\$35M + 2267ETH + 8BTC) to explore and develop blockchain technologies for creating transparent, efficient systems and rethinking problem-solving approaches in enhancing lives and developing communities [Chapiro *et al.*, 2021]. Our exploration reveals that “Machine Learning for Blockchain Data Analysis” has emerged as a vibrant and influential field since 2018 with more than 1750 publications dedicated to this field in the ACM Digital Library.

We apply rigorous criteria to select and evaluate papers that contribute the most to the “ML for Blockchain Data Analysis” field. They encompass factors such as the relevance of the research, the significance of the problem addressed, the quality of the methodology employed, and the impact of the findings on the broader artificial intelligence community. Our search particularly focused on articles that analyzed and built models for data from a public blockchain such as Bitcoin, Ethereum, Litecoin, Eosio, Ripple, Monero, Zcash, and Dash.

Contributions and Roadmap. Our survey offers several key contributions to the field. First, it provides a comprehensive taxonomy (§2) and overview (§4) of the latest advancements in “ML for Blockchain Data Analysis” since 2018, offering insights into the state of the art. Second, in §5 we discuss how the datasets and tools we have highlighted can significantly facilitate future ML research, benchmarking, and the development of innovative applications in the field. Additionally, we discuss the unique challenges (§3) and opportunities (§6) inherent in this domain, shedding light on areas that require further exploration and innovation. Ultimately, our survey aims to guide researchers, practitioners, and policymakers in harnessing the potential of machine learning within the blockchain ecosystem, promoting user-friendly, explainable, and responsible data analysis practices. To the best of our knowledge, this is the first comprehensive survey that covers all five areas of ML on blockchains (see Table 1).

Table 1: Comparison of survey articles across ML for blockchains.

Survey	Graph ML	Seq. ML	Code ML	Temp. ML	Text ML
A Survey on Blockchain Anomaly Detection Using Data Mining Techniques [Li <i>et al.</i> , 2020a]	✓	×	✓	✓	×
Knowledge Discovery in Cryptocurrency Transactions: A Survey [Liu <i>et al.</i> , 2021a]	✓	✓	✓	✓	×
A Survey on Blockchain Data Analysis [Hou <i>et al.</i> , 2021]	✓	✓	✓	×	×
Analysis of Cryptocurrency Transactions from a Network Perspective: An Overview [Wu <i>et al.</i> , 2021]	✓	×	✓	✓	✓
Anomaly Detection in Blockchain Networks: A Comprehensive Survey [Hassan <i>et al.</i> , 2022]	✓	✓	✓	✓	×
Graph Analysis of the Ethereum Blockchain Data: A Survey of Datasets Methods and Future Work [Khan, 2022]	✓	×	✓	✓	×
A survey on machine learning approaches in cryptocurrency: challenges and opportunities [Mujlid, 2023]	×	✓	×	×	×
Blockchain Data Mining with Graph Learning: A survey [Qi <i>et al.</i> , 2023]	✓	✓	✓	✓	×
Machine Learning for Blockchain Data Analysis: Progress and Opportunities [ours]	✓	✓	✓	✓	✓

2 Taxonomy

We discuss our taxonomy of machine learning methods (§2.1), blockchain components (§2.2), data models (§2.3), and applications of blockchain data analysis (§2.4).

2.1 Machine Learning Methods

The integration of machine learning is unlocking new potential in blockchain data analysis and decision-making [Khan and Akcora, 2022]. ML approaches, including graph-based learning, recurrent neural networks (RNN), and transformers, have become pivotal in extracting insights from blockchain’s complex and varied data structures. These methods enable a nuanced understanding of blockchain components, such as transaction networks and smart contracts, by identifying patterns and anomalies that might otherwise remain obscured.

Graph ML approaches such as unsupervised methods, graph embedding, and graph neural networks, e.g., graph convolutional neural networks (GCNs) and graph attention networks (GATs) [Xia *et al.*, 2021] are essential for analyzing complex network structures. *Sequential ML*, e.g., RNNs and transformers are adept at processing sequential data [Wen *et al.*, 2023], thus crucial for transaction analysis. *Code ML* techniques for smart contract analysis focus on interpreting code and bytecode [Pierro *et al.*, 2020]. *Temporal ML* handles time-sensitive data – revealing trends, prices, and patterns over time [Benidis. *et al.*, 2023]. Lastly, *Text ML*, particularly using text and NLP on social media posts, offers insights into public perception and interactions regarding blockchains [Rouhani and Abedin, 2020]. The categories are not mutually exclusive, e.g., temporal graph learning deals with both graph ML and temporal ML; it has been exploited in cryptocurrency e-crimes detection [Akcora *et al.*, 2021].

2.2 Blockchain Components

The key blockchain components include the *transaction network*, which records assets (e.g., cryptocurrency) movements; *token networks*, managing the distribution and interactions of various tokens; and *smart contracts*, which are automated

agreements encoded directly in the blockchain. Additionally, the *peer-to-peer (P2P) network* underpins the decentralized nature of blockchains, allowing direct interactions among users. *User accounts* represent individuals or entities with their transaction histories and balances. A *decentralized application (dApp)* combines one or more smart contracts to support a certain functionality on a distributed, peer-to-peer network; for example, *decentralized finance (DeFi)* are dApps for financial services. One may also consider external sources, including social media data, online blogs, cryptocurrency prices, Google Trends, etc., to mine public sentiments and trends about blockchains. For a detailed survey on blockchain components, we refer to [Khan, 2022].

2.3 Blockchain Data Models

The data model for blockchain analysis in ML includes i) simple graphs that illustrate basic peer-to-peer connections, ii) temporal graphs that capture changes across time, iii) attributed graphs where nodes and edges carry distinct properties and iv) weighted graphs with varying importance assigned to connections. Furthermore, directed graphs indicating transaction directions, dynamic graphs reflecting evolving relationships, stream graphs representing continuous data flows, and higher-order graphs offering a multi-dimensional perspective on interactions, have been considered [Akcora *et al.*, 2022].

Another aspect of the data model is the analysis of smart contract code, which is essential for understanding the functional mechanics of blockchain systems [Bartoletti *et al.*, 2020]. This includes both the source code, which offers insights into the logic and rules governing the contracts; and the bytecode, which is the executable form deployed on the blockchain. Furthermore, analyzing text data from transaction descriptions, user comments, and other textual inputs provides a unique perspective on user behaviors and social dynamics within the blockchain ecosystem. The integration of these varied data types, including sequential data models, e.g., time series, is indispensable for a comprehensive analysis. This integration not only helps in decoding the current state of the blockchain but also in forecasting future trends. We shall highlight graph, time series, and smart contract code data models, as well as their combinations in our survey.

2.4 Applications of Blockchain Data Analysis

Blockchain data analysis has diverse applications pivotal to the advancement of blockchain technology. This domain facilitates predictive analytics in financial cryptocurrency markets and anomaly detection within blockchain networks [Li *et al.*, 2020a]. Furthermore, the field is useful in identifying and mitigating financial crimes, including ransomware, money laundering, darknet markets, and Ponzi schemes [Wu *et al.*, 2023]. Additionally, blockchain data analysis is key in address/transaction clustering and scrutinizing code for duplicates or malicious contents, thus enhancing the security and integrity of blockchain systems.

3 Challenges of Machine Learning for Blockchain Data Analysis

In the realm of blockchain technology, a complex web of challenges emerges from technology, its usage, control mechanisms, the nature of data, and the ML methods employed.

Blockchain Technology. A fundamental aspect of all public blockchains is the anonymous nature of blockchain addresses. The anonymity allows fast and easy access to blockchain for users, but it also presents a significant hurdle when tracking addresses and analyzing transaction patterns. A second technological challenge in blockchain arises from the fact that only the compiled binary of smart contract code is visible on the blockchain. This limited visibility restricts our understanding of the underlying source code, obscuring the logic and potential vulnerabilities of these contracts. This opacity is a significant concern for ensuring the integrity and security of the blockchain network, as it hinders comprehensive auditing and analysis of smart contracts.

Blockchain Usage. A blockchain is characterized by the dynamic nature of its data. With new transactions arriving in blocks every 15 seconds (as seen on Ethereum [Wood, 2018]) to 10 minutes (as on Bitcoin [Nakamoto, 2008]), the data is in a constant state of evolution. This poses a significant challenge in maintaining updated and relevant analyses in real-time. The sheer volume of this data, compounded by its sparse and graph-like structure, exacerbates computational and analytical difficulties. Additionally, the complexity is further intensified by coin-mixing schemes [Wu *et al.*, 2022a], which deliberately muddle the process of tracking transaction flows, often to obscure the origins of funds for purposes such as coin-laundering [Akcora *et al.*, 2020].

Blockchain Control Mechanisms. The open and decentralized nature of blockchains, while one of its strengths, also invites a range of adversarial behaviors. This includes long-range attacks and manipulations, challenging the system’s integrity and reliability. The lack of a centralized review mechanism for both code and users in the blockchain further heightens these risks, leaving the network vulnerable to malicious smart contracts and abusive users.

Blockchain Data. Data-related challenges in blockchains are multifaceted. When utilizing labeled data in blockchain analysis, the rarity of the positive class (such as instances of ransomware or money laundering) compared to the vast size of the networks results in a significant bias in the methods employed. Such a skewed distribution can lead to misleadingly high accuracy metrics. The scarcity of verified, reliable ground truth data hampers the development and validation of robust analytical models. Furthermore, the challenge of train-test mismatch in blockchain analytics is accentuated by the ever-evolving nature of blockchains, which are frequently impacted by real-world events such as government regulations or bans [Xie, 2019]. These external influences can significantly alter the nature of the data within a given period, leading to a scenario where the blockchain’s state during the training phase may be different from that in the testing phase. This divergence between training and testing data distributions severely compromises the accuracy and generalizability

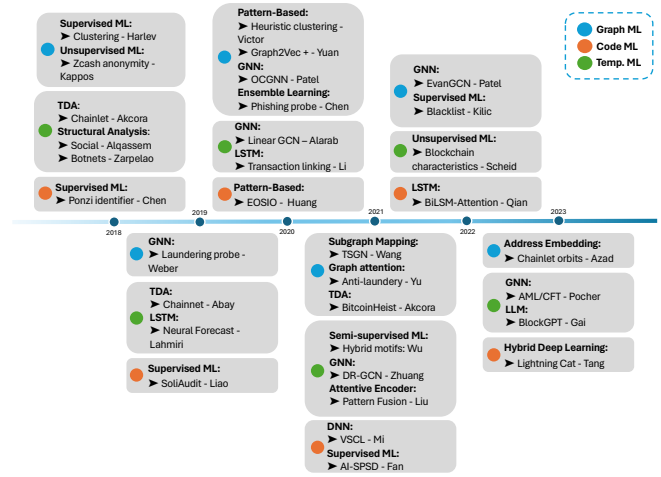


Figure 1: The timeline of machine learning for Blockchain research.

of models, presenting a substantial obstacle to the effectiveness of machine learning applications in blockchain analysis.

ML Models. The challenges extend into the domain of machine learning methods used for blockchain data analysis. The “black-box” neural models, particularly deep learning, raise concerns about explainability and interpretability. These are critical issues in a field that demands transparency and accountability to comply with financial regulations. Inherent biases in ML algorithms pose risks of unfairness, contradicting the ethos of blockchain technology. Furthermore, the high computational demands, including extensive training and inference times and the need for large volumes of labeled training data, present substantial challenges, especially when data is often scarce, dynamic, and unlabeled.

4 Survey: Blockchain Data Models, Machine Learning Methods, and Applications

We primarily investigate three non-exclusive ML approaches: graph machine learning (§4.1), temporal machine learning (§4.2), and machine learning for smart contracts (§4.3). We survey their methods for blockchain data analysis, respective data models, and applications. A schematic diagram connecting various articles in our survey is illustrated in Figure 1.

4.1 Graph Machine Learning on Blockchains

4.1.1 Graph Data Models

UTXO Data Models. Blockchain technology, which started with Bitcoin, utilizes a distinctive data structure known as an “output” that contains an address and an amount. Such blockchains are referred to as the UTXO (Unspent Transaction Output) blockchains. An address is a unique string representation of the holder within the transaction network. A Bitcoin transaction, where a later transaction consumes one or more outputs to generate new outputs, can effectively be modeled as heterogeneous graphs comprising two primary node types: addresses and transactions. However, a significant challenge arises with most graph libraries, e.g., NetworkX [Hagberg *et al.*, 2008], which are designed to handle

graphs with a single node type. This limitation has led researchers to frequently model the Bitcoin transaction network as either an address graph [Spagnuolo *et al.*, 2014] by omitting transactions, or a transaction graph [Ron and Shamir, 2013] by omitting addresses. Specifically, both the address graph and the transaction graph are edge-weighted, directed graphs with nodes representing their respective namesakes, and directed edges record the flow of coins. An edge weight represents the amount of coins transferred.

Account Data Models. The emergence of Ethereum introduced a shift in blockchain data models. Unlike Bitcoin, Ethereum employs an account-based model that eschews the output data structure. Instead, the representation shifts to a graph of address nodes. A key feature of these networks is the variety of edge types, which can represent different forms of value transfer, such as the native cryptocurrency (Ether), tokens, or other user-defined assets. This complexity transforms the network into a multiplex network [Dickison *et al.*, 2016], where address nodes are shared, but the edges differ in their types and meanings. Therefore, these graphs are categorized as directed, edge-weighted multigraphs.

Moreover, the application of hypergraphs [Antelmi *et al.*, 2023] presents a new dimension in modeling blockchain transactions, particularly beneficial in e-crime scenarios where coins flow between seemingly different addresses which are, in reality, owned by the same user. For instance, in coin mixing networks such as Tornado Cash [Wu *et al.*, 2022b], the flow of coins creates a hyper-edge that connects more than two nodes, providing a more nuanced view of asset transfers in such systems.

4.1.2 Graph Machine Learning Methods

We categorize the discussion based on unsupervised and supervised graph ML, as well as techniques to scale graph ML.

Unsupervised Learning. The evolution of blockchain analytics has been significantly influenced by the application of unsupervised learning techniques. Initial research in this domain mainly focused on examining transaction patterns within blockchain networks to understand the flow of digital currencies, identify trends, and detect anomalies [Ron and Shamir, 2013]. This analysis typically included studying aspects such as transaction volumes, frequency, and the interrelationships between different addresses [Lee *et al.*, 2020].

As the research progressed, a shift towards more address and transaction-centric views emerged. Address clustering, aiming to deduce which addresses are controlled by the same user, gained considerable attention [Victor, 2020; Harrigan and Fretter, 2016]. Address clustering employs various heuristics that exploit the characteristics of UTXO transactions. This process is largely unsupervised and focuses on linking entities behind blockchain addresses. Clustering plays a crucial role in identifying and understanding address behaviors and transaction patterns [Spagnuolo *et al.*, 2014]. Similar unsupervised analyses have been performed on reportedly “anonymous” cryptocurrencies, e.g., Monero [Möser *et al.*, 2017], Zcash [Kappos *et al.*, 2018], and a diverse set of cryptocurrency ledgers [Yousaf *et al.*, 2019].

Supervised Learning. The advent of public datasets, e.g., Elliptic [Weber *et al.*, 2019] signified a pivotal moment in

the realm of blockchain graph machine learning, providing a rich source of labeled data. This marked a transition towards more supervised learning approaches, broadening the scope and precision of blockchain data analysis. We categorize these supervised methods into three classes: graph features extraction, graph embeddings, and graph neural networks.

Graph Features Extraction. Harlev *et al.* [Harlev *et al.*, 2018] first use unsupervised clustering on the transaction graph to link bitcoin addresses owned by the same user. Next, supervised machine learning based on cluster features has been employed to de-anonymize entities on the Bitcoin blockchain. This approach relies on known data about entities whose identities were previously exposed to form a training dataset, thereby reducing the level of anonymity inherent in Bitcoin transactions. Supervised learning has also been effectively used in detecting blacklisted addresses in the Ethereum blockchain [Kılıç *et al.*, 2022]. The approach involved using both local and global features extracted from the Ethereum transaction graph to train various machine learning models. This method’s feature extraction process, employing techniques such as random undersampling and SMOTE [Chawla *et al.*, 2002], is designed to address label scarcity.

Graph Embeddings. Graph embeddings map each node in a graph to a low-dimensional vector, e.g., for supervised node classification, which has been pivotal in detecting phishing activities within blockchain networks. Yuan *et al.* [Yuan *et al.*, 2020] introduce a graph-based classification framework leveraging an improved Graph2Vec algorithm to analyze Ethereum transaction networks for this purpose. The paper’s focus on Ether flow in phishing scams integrates this aspect into the machine learning model, enhancing phishing detection capabilities. Similarly, Wang *et al.* [Wang *et al.*, 2021] develop the transaction subgraph network model to identify phishing accounts in the Ethereum blockchain, utilizing a directed version of the model that retains transaction flow information crucial for identifying such illicit activities.

Graph Neural Networks. GNNs are deep learning models developed for graph-related tasks in an end-to-end manner. A notable contribution in this domain is the work on detecting Ponzi schemes within the Ethereum blockchain [Yu *et al.*, 2021b]. Here, a model based on a graph convolutional network is developed to classify nodes in the Ethereum transaction network as Ponzi or non-Ponzi. This approach demonstrates the efficacy of supervised learning in identifying fraudulent schemes by examining the topological structure and transactional characteristics of smart contracts. The development of graph attention network models to identify abnormal transactions in dynamically generated data is also a key area where supervised learning has shown great promise. Yu *et al.* [Yu *et al.*, 2021a] introduce a GAT approach, focusing on exploiting the graph structure of transactions. The method’s dynamic graph handling capability and weight assignment to nodes based on their relevance to abnormal transactions offer advanced capabilities.

Moreover, the concept of anomaly detection in Ethereum’s blockchain network has been explored. Patel *et al.* [Patel *et al.*, 2020] employ the “one-class” graph neural network capturing complex relationships and interactions between ac-

counts for more effective identification of anomalous patterns. Analogously, the paper by Patel et al. [Patel et al., 2022] develops EvAnGCN, a dynamic GCN for detecting anomalous behaviors in blockchain networks by structuring the data as temporal graphs. This model efficiently learns from the dynamic and evolving structures of blockchain networks, utilizing both temporal and structural features.

Furthermore, the identification of illicit Bitcoin addresses has been enhanced through the integration of structure and temporal information of Bitcoin transactions. Tian et al. [Tian et al., 2021] develop an attention-based graph neural network that refines address embeddings through neighbor embedding and attention mechanisms. An LSTM-based auto-encoder is used to capture hidden temporal features from transaction records, augmenting identification accuracy.

Scaling Graph Machine Learning. Scaling graph machine learning on blockchains is crucial for handling the vast and continuously growing volume of data within transaction networks. For example, Bitcoin has $\approx 700,000$ unique addresses daily in 500,000 transactions.¹ Examining the Bitcoin transaction network for even a single day poses a computationally demanding challenge for graph neural networks which are considered state-of-the-art in a multitude of predictive tasks, such as node classification [Yang et al., 2023].

In their initial efforts to analyze large graphs, researchers typically focus on extracting information from the local neighborhoods of nodes. Kılıç et al. employ easily calculable features, including neighbor counts and the time difference between the first and last transactions of a given address [Kılıç et al., 2022]. If computing power permits, e.g., using parallel computing, researchers may extend their analysis to higher-hop neighborhoods [Yu et al., 2021a].

One common scaling approach is node sampling. This technique has been widely employed to manage large transaction networks. For instance, Harlev et al. classify entities based on transactional behaviors without necessitating analysis of the entire network [Harlev et al., 2018]. Similarly, Yu et al. identify Ponzi schemes within the Ethereum blockchain by node sampling to create subgraphs for analysis [Yu et al., 2021b]. The authors randomly sample centered contracts to obtain their first-order neighbors, significantly reducing the computational load. Another scaling strategy involves the use of subgraph sampling, where transaction subgraphs are extracted and analyzed. This is evident in the work of Yu et al., where the dynamic graph structures employ a GAT model that relies on the structure of the sampled edges, rather than requiring a complete graph for analysis [Yu et al., 2021a]. This method is particularly effective in processing dynamic graph structures, and adapting to real-time transaction data.

4.1.3 Open Questions and Challenges

Graph machine learning for blockchains faces several critical challenges. Label scarcity is a prominent but well-known issue. An under-reported issue is the undisclosed e-crime transactions (e.g., ransomware payments), which may create false positives in node classification tasks. The scale of blockchain graphs presents a computational hurdle, demanding efficient

algorithms and scalable systems. Real-time analysis is crucial as blockchain data evolves rapidly where latency in detecting anomalies can cause billions of dollars in lost value (e.g., in the LunaTerra collapse). Integrating machine learning across multiple blockchains is complex, involving data heterogeneity and interoperability challenges (e.g., in UTXO-account data integration). Detecting data shifts within blockchain graphs is essential for maintaining model accuracy as usage patterns by ordinary users, as well as e-crime operators, change. Tackling these challenges is essential for harnessing machine learning’s potential in blockchain data analysis.

4.2 Temporal Machine Learning on Blockchains

The integration of ML with blockchain’s temporal data offers unique opportunities for enhanced security, predictive analytics, and understanding dynamic market behaviors.

4.2.1 Temporal Data Models

Temporal data on blockchains offer a rich variety, including time series of crypto asset prices; temporal, multilayer graphs of transaction and asset networks; discrete and continuous dynamic graphs; and graphs with temporal node and edge features. The market volumes of native coins have reached billions of dollars. Hence, the most critical temporal data relates to the price of the native coins, such as Ether on the Ethereum network, denominated in fiat currency. The price data also exists for a subset of crypto assets on blockchains, such as tokens on Ethereum due to global trading activities, thereby establishing an external pricing dataset. Transaction and asset trading networks provide temporal transaction data in the form of networks where both node and edge attributes, as well as edge types, may change. When a blockchain has a short block creation interval (e.g., Ethereum’s ≈ 12 sec gap between two blocks), the network can be effectively modeled as an (almost) continuous-time dynamic graph.

4.2.2 Temporal Machine Learning Methods

Time Series Analysis. Early work in time series analysis for cryptocurrencies used abundant transaction network data to extract predictive signals. Abay et al. [Abay et al., 2019] use Bitcoin graph substructures, called chainlets [Akcora et al., 2018], to predict Bitcoin prices. Kwon et al. [Kwon et al., 2019] use the long short-term memory (LSTM) model [Schmidhuber and Hochreiter, 1997] on the historic cryptocurrency price time series data to classify the time series. Livieris et al. use ensemble-averaging, bagging, and stacking with deep learning models for forecasting hourly cryptocurrency prices [Livieris et al., 2020].

Unsupervised Learning. The transaction network provides a dynamic dataset abundant in user behavior, enabling the mining of complex patterns. For instance, Alqassem et al. analyze the Bitcoin transaction graph from its inception [Alqassem et al., 2018]. They observe changes in network diameter, node connectivity, and community structure over time. Their findings include patterns like the densification power law and shrinking diameter. Importantly, they underscore the influence of anonymity-seeking behavior on Bitcoin’s network dynamics. Zhao et al. investigate the evolutionary nature of the Ethereum blockchain network such as

¹<https://www.blockchain.com/charts/n-unique-addresses>

the growth rate, active lifespan of high-degree nodes, detecting anomalies based on temporal changes in global network properties, and forecasting the survival of network communities [Z. *et al.*, 2021]. In the context of blockchain selection, Scheid *et al.* [Scheid *et al.*, 2022] introduce an ML-based approach to simplify the selection process for non-technical individuals. The authors present a novel metric to quantify the subjective popularity of blockchain platforms, contributing to the feature set used in their ML model. This work emphasizes the temporal flexibility of their ML model, which adapts over time to new parameters and data.

Supervised Learning. Many temporal ML articles study graph ML topics with a temporal view. Alarab *et al.* divide the popular Elliptic dataset into 49 time-steps, each representing a distinct set of transactions within a three-hour window [Alarab *et al.*, 2020]. This temporal division of data ensures that the model can handle real-time transaction data and be trained on temporally coherent subsets. Temporal information is also useful in profiling blockchain addresses. Harlev *et al.* focus on de-anonymizing entities on the Bitcoin blockchain by analyzing transactions over time and extracting useful features, such as transaction patterns and time-series data [Harlev *et al.*, 2018]. This temporal dimension enables predicting behaviors based on transaction history.

In e-crime research, temporal transaction patterns exhibited by operators such as ransomware hackers [Akcora *et al.*, 2021] is invaluable. Pocher *et al.* effectively utilize patterns by first grouping Bitcoin transactions into distinct time steps and then using a chronological analysis of transaction patterns to find characteristic of e-crime activities [Pocher *et al.*, 2023]. In anonymity-seeking behavior, users employ different addresses for each transaction to maintain their anonymity. The anonymous behavior is further strengthened by coin-mixing services where one can launder the coins through a mixing service. Wu *et al.* propose a feature-based network analysis framework to identify such mixing services on Bitcoin [Wu *et al.*, 2022a]. In their work, temporal motifs are crucial to distinguish normal transactions from those associated with mixing services.

Sequence-based Models. Li *et al.* focus on identifying illicit Bitcoin addresses by extracting temporal features from the change in the balance of addresses over time [Li *et al.*, 2020b]. They use an auto-encoder with LSTM to generate discriminating temporal features, enhancing the model’s ability to identify illicit addresses based on temporal patterns. This approach highlights the importance of temporal analysis in distinguishing normal transaction behavior from illicit activities. Lahmiri *et al.* used LSTM neural networks for predicting cryptocurrency prices [Lahmiri and Bekiros, 2019]. Their model memorizes both long-term and short-term temporal information, which is crucial for predicting the volatile and dynamic nature of cryptocurrency markets. One recent contribution in this field is BlockGPT, a dynamic, real-time approach for detecting anomalous blockchain transactions [Gai *et al.*, 2023]. This tool is notable for its ability to generate tracing representations of blockchain activity and train an LLM as a real-time intrusion detection system. Unlike traditional methods, BlockGPT does not rely on prede-

defined rules or patterns, making it significantly more effective in detecting anomalies in Ethereum transactions.

Graph Neural Networks. Zhuang *et al.* propose a novel method for detecting vulnerabilities in smart contracts using graph neural networks [Zhuang *et al.*, 2021]. They introduce a degree-free graph convolutional neural network and a temporal message propagation network for automatic detection. The temporal aspect is central to their approach, considering the sequence of operations and interactions within smart contracts to detect vulnerabilities over time. Liu *et al.* introduce a method for detecting vulnerabilities in smart contracts by combining graph neural networks with expert knowledge [Liu *et al.*, 2021b]. They transform smart contract source code into a contract graph, focusing on critical nodes through a node elimination phase. A temporal message propagation network is employed to extract graph features, considering the sequential nature of smart contract execution. This approach is pivotal in detecting vulnerabilities by capturing the temporal dynamics of data and control flows within smart contracts. Other notable works include [Patel *et al.*, 2022; Yu *et al.*, 2021a] for detecting anomalous transactions; due to the non-exclusive nature of our categorization, they have been discussed earlier in graph ML (§4.1.2).

4.2.3 Open Questions and Challenges

Linking temporal data across multiple blockchains (e.g., between Bitcoin and Monero in money laundering) to identify behavior patterns presents a complex challenge. Blockchains operate independently, and cross-chain data analysis requires addressing issues related to data heterogeneity, interoperability, and privacy while uncovering valuable insights into cross-blockchain behaviors. Identifying significant changes or anomalies in temporal blockchain data is critical for understanding and responding to emerging trends or irregularities such as hacked blockchain bridges, seized addresses, and external events [Xie, 2019]. Developing effective change point detection algorithms tailored to blockchain data remains an open question on (sparse) transaction graphs. Another challenge is dealing with data staleness issues. As blockchain data continuously evolves, ensuring that ML models operate on informative and up-to-date information is essential.

4.3 Machine Learning for Smart Contracts

4.3.1 Smart Contract Data Models

We consider four types of smart contract data: transaction, contract state, event log, and source code. Transaction data includes information on each transaction executed on the blockchain, e.g., sender and receiver addresses, and block numbers. Smart contracts have a state, which is essentially the current data stored in the contract. This state includes variables, balances, and other information specific to the contract’s functionality. Events, emitted by contracts, record specific occurrences, such as the completion of a task, or the occurrence of an event-triggering condition. The source code of a smart contract (in bytecode or higher level languages, e.g., Solidity) is another critical element for ML analysis.

4.3.2 Machine Learning Methods for Smart Contracts

Contract Graph Analysis. Ferreira *et al.* automate detection and investigation of attacks on Ethereum smart contracts,

utilizing logic-driven and graph-driven analysis of transactions [Ferreira T. *et al.*, 2021]. Zhuang *et al.* construct a contract graph to represent both syntactic and semantic structures of contract functions [Zhuang *et al.*, 2021]. Liu *et al.* propose a method that transforms smart contract source code into a contract graph, highlights critical nodes via a node elimination phase, and employs a temporal message propagation network to extract graph features [Liu *et al.*, 2021b]. These features, combined with expert-designed security patterns, contribute to an effective and scalable vulnerability detection system on platforms, e.g., Ethereum and VNT Chain.

Source Code Analysis. Mi *et al.* propose a metric learning-based deep neural network for vulnerability detection in smart contracts, focusing on analyzing bytecode [Mi *et al.*, 2021]. Fan *et al.* detect smart Ponzi schemes in blockchain systems by extracting smart contract features from Op-Codes [Fan *et al.*, 2021]. Qian *et al.* present a deep learning model, BiLSTM-Attention, for detecting defects in smart contracts, treating contract operation codes as sequential sentences, and utilizing attention mechanisms for accurate detection [Qian *et al.*, 2022]. Tang *et al.* identify vulnerabilities by analyzing code snippets of functions [Tang *et al.*, 2023].

Community and Transaction Analysis. Huang *et al.* provide a large-scale analysis of the EOSIO blockchain ecosystem, identifying bot activities at both community-level and account-level [Huang *et al.*, 2020]. SoliAudit combines ML and fuzz testing for vulnerability assessment using Solidity machine code as learning features and incorporating gray-box fuzz testing [Liao *et al.*, 2019]. Chen *et al.* detect Ponzi schemes in Ethereum by extracting features from user accounts and operation codes of contracts [Chen *et al.*, 2018].

4.3.3 Open Questions and Challenges

One significant challenge in code machine learning for blockchains is the difficulty in finding the high-level code of smart contracts. Smart contracts often have their bytecode uploaded to the blockchain, making it challenging to access their human-readable source code. Lack of access to high-level code hinders comprehensive analysis and interpretation.

The decentralized and distributed nature of blockchain networks can introduce vulnerabilities, such as reentry attacks, not found in typical software projects. Analyzing the script languages of blockchains for these vulnerabilities requires blockchain domain knowledge as well as a good understanding of how distributed systems work. As a result, coding for blockchains is a challenging software domain.

Additionally, functions and opcodes on blockchains often lack direct equivalents in conventional programming languages, which makes it challenging to apply standard code analysis techniques, as the mapping between blockchain code and traditional code constructs may not be straightforward.

5 Datasets and Tools

Graphs. Blockchain network data have become increasingly valuable in research for financial transactions, network dynamics, and user behavior. The Elliptic dataset [Weber *et al.*, 2019] stands out with its labeled Bitcoin transaction graph,

which has been utilized in GNNs. However, the dataset employs anonymized addresses, and descriptions of node features are not shared due to intellectual property rights issues. The BitcoinHeist dataset shares address and labels for about 30K addresses linked to ransomware, facilitating more direct transaction pattern analysis [Akcora *et al.*, 2021].

The evolution of blockchain datasets has been notable. Initially, datasets were released in conjunction with academic articles in isolated repositories [Anoiaica and Levard, 2018; Liang *et al.*, 2018; Lee *et al.*, 2020]. However, recent trends, particularly highlighted in benchmark tracks of conferences, e.g., NeurIPS, have led to the development of standardized and accessible benchmarks, such as Chartalist [Shamsi *et al.*, 2022] and NFTGraph [Zhang *et al.*, 2023]. These benchmarks provide large-scale, labeled graph data crucial for diverse research areas, from financial fraud detection to network dynamics analysis. The datasets are also used in the analysis of real-life phenomena where datasets are quite difficult to access. For example, Zhang *et al.* have proposed to use blockchain networks for studying the resilience of power networks [Zhang and Y., 2021].

Code. Smart contract code datasets, such as [Ortner and Eskandari, 2024; di Angelo *et al.*, 2023], include vulnerable smart contract codes, offering valuable insights into security vulnerabilities within blockchain applications. Ibba *et al.* [Ibba, 2022] provide token and non-fungible token contract code datasets, shedding light on the intricacies of these specialized smart contract types.

Tools. Kushwaha *et al.* provide a comprehensive overview of tools and methodologies for analyzing Ethereum-based smart contracts [Kushwaha *et al.*, 2022]. Additionally, [Durieux *et al.*, 2020] provides a comprehensive resource for an empirical review of automated analysis tools on a dataset of 47,587 Ethereum smart contracts.

6 Conclusion and Future Direction

The field of machine learning for blockchains has made significant progress in addressing numerous challenges, as highlighted in this survey. However, several promising future directions await further advancement. Firstly, ensuring that ML model decisions are transparent and interpretable is crucial for responsible and trustworthy blockchain data analysis. As blockchain data continues to grow in size and complexity, the development of scalable learning and inference techniques becomes imperative. Efficient algorithms and distributed computing approaches will play a pivotal role in handling the ever-expanding datasets. Furthermore, exploring the application of machine learning to complex blockchain networks, including cross-chain analysis, offers new insights and opportunities for research. Moreover, the dynamic nature of blockchain data requires the development of machine unlearning and continuous learning techniques, enabling models to adapt to evolving data distributions and maintain accuracy over time. Lastly, harnessing the capabilities of large language models for understanding natural language, interacting with data, and generating source code can revolutionize blockchain data and smart contract analysis.

References

- [Abay *et al.*, 2019] N. C. Abay, C. G. Akcora, Y. R. Gel, et al. Chainnet: Learning on blockchain graphs with topological features. In *ICDM*, 2019.
- [Akcora *et al.*, 2018] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu. Forecasting bitcoin price with graph chainlets. In *PAKDD*, 2018.
- [Akcora *et al.*, 2020] C. G. Akcora, S. Purusotham, et al. How to not get caught when you launder money on blockchain? *arXiv:2010.15082*, 2020.
- [Akcora *et al.*, 2021] C. G. Akcora, Y. Li, Y. R. Gel, and M. Kantarcioglu. Bitcoinheist: Topological data analysis for ransomware prediction on the bitcoin blockchain. In *IJCAI*, 2021.
- [Akcora *et al.*, 2022] C. G. Akcora, Y. R. Gel, and M. Kantarcioglu. Blockchain Networks: Data Structures of Bitcoin, Monero, Zcash, Ethereum, Ripple, and Iota. *WIREs Data Mining Knowl. Discov.*, 12(1), 2022.
- [Alarab *et al.*, 2020] I. Alarab, S. Prakoonwit, and M. I. Nacer. Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In *ICMLT*, 2020.
- [Alqassem *et al.*, 2018] I. Alqassem, I. Rahwan, and D. Svetinovic. The anti-social system properties: Bitcoin network data analysis. *IEEE Trans Syst Man Cybern.*, 50(1):21–31, 2018.
- [Anoica and Levard, 2018] A. Anoica and H. Levard. Quantitative description of internal activity on the ethereum public blockchain. In *NTMS*, 2018.
- [Antelmi *et al.*, 2023] A. Antelmi, G. Cordasco, et al. A survey on hypergraph representation learning. *ACM Comp. Sur.*, 56(1):1–38, 2023.
- [Bartoletti *et al.*, 2020] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia. Dissecting ponzi schemes on ethereum: identification, analysis, and impact. *Future Generation Computer Systems*, 102:259–277, 2020.
- [Benidis. *et al.*, 2023] K. Benidis., Syama S. R., et al. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput. Surv.*, 55(6):121:1–121:36, 2023.
- [Chapiro *et al.*, 2021] C. Chapiro, M. Hydary, and C. Lomazzo. Linking blockchain to impact, 2021.
- [Chawla *et al.*, 2002] N. V. Chawla, K. W. Bowyer, et al. Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
- [Chen *et al.*, 2018] W. Chen, Z. Zheng, et al. Detecting ponzi schemes on ethereum: Towards healthier blockchain technology. In *WWW*, 2018.
- [di Angelo *et al.*, 2023] M. di Angelo, T. Durieux, J. F. Ferreira, and G. Salzer. SmartBugs 2.0: An execution framework for weakness detection in Ethereum smart contracts. In *ASE*, 2023. to appear.
- [Dickison *et al.*, 2016] M. E. Dickison, M. Magnani, and L. Rossi. *Multilayer social networks*. Cambridge University Press, 2016.
- [Durieux *et al.*, 2020] T. Durieux, J. F. Ferreira, et al. Empirical review of automated analysis tools on 47, 587 ethereum smart contracts. In *ICSE*, pages 530–541. ACM, 2020.
- [Fan *et al.*, 2021] S. Fan, S. Fu, H. Xu, and X. Cheng. Al-spsd: Anti-leakage smart ponzi schemes detection in blockchain. *IPM*, 58(4):102587, 2021.
- [Ferreira T. *et al.*, 2021] Christof Ferreira T., A. K. I., A. Gervais, and R. State. The eye of horus: Spotting and analyzing attacks on ethereum smart contracts. In *FC*, 2021.
- [Gai *et al.*, 2023] Y. Gai, L. Zhou, K. Qin, D. Song, and A. Gervais. Blockchain large language models. *arXiv preprint arXiv:2304.12749*, 2023.
- [Hagberg *et al.*, 2008] A. Hagberg, P. Swart, and D. S. Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab, United States, 2008.
- [Harlev *et al.*, 2018] M. A. Harlev, H. Sun Yin, et al. Breaking bad: De-anonymising entity types on the bitcoin blockchain using supervised machine learning. *HICSS*, 2018.
- [Harrigan and Fretter, 2016] M. Harrigan and C. Fretter. The unreasonable effectiveness of address clustering. In *UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld*, 2016.
- [Hassan *et al.*, 2022] Muneeb Ul Hassan, Mubashir Husain Rehmani, and Jinjun Chen. Anomaly detection in blockchain networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 2022.
- [Hou *et al.*, 2021] Wenhan Hou, Bo Cui, and Ru Li. A survey on blockchain data analysis. In *COMPSAC*, 2021.
- [Huang *et al.*, 2020] Y. Huang, H. Wang, L. Wu, et al. Understanding (mis) behavior on the eosio blockchain. *AMCS*, 4(2):1–28, 2020.
- [Ibba, 2022] G. Ibba. A smart contracts repository for top trending contracts. In *IWETSEB*, pages 17–20, 2022.
- [Kappos *et al.*, 2018] G. Kappos, H. Yousaf, M. Maller, and S. Meiklejohn. An empirical analysis of anonymity in zcash. In *USENIX Security*, 2018.
- [Khan and Akcora, 2022] Arijit Khan and Cuneyt Gurcan Akcora. Graph-based management and mining of blockchain data. In *CIKM*, 2022.
- [Khan, 2022] Arijit Khan. Graph analysis of the ethereum blockchain data: A survey of datasets, methods, and future work. In *Blockchain*, 2022.
- [Kılıç *et al.*, 2022] B. Kılıç, A. Sen, and C. Özturan. Fraud detection in blockchains using machine learning. In *BCCA*, 2022.
- [Kushwaha *et al.*, 2022] S. S. Kushwaha, S. Joshi, et al. Ethereum smart contract analysis tools: A systematic review. *IEEE Access*, 10:57037–57062, 2022.
- [Kwon *et al.*, 2019] D. Kwon, J. Kim, J. Heo, C. Kim, and Y. Han. Time series classification of cryptocurrency price trend based on a recurrent lstm neural network. *Journal of Information Processing Systems*, 15(3):694–706, 2019.
- [Lahmiri and Bekiros, 2019] S. Lahmiri and S. Bekiros. Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118:35–40, 2019.
- [Lee *et al.*, 2020] X. T. Lee, A. Khan, et al. Measurements, analyses, and insights on the entire ethereum blockchain network. In *WebConf*, 2020.
- [Li *et al.*, 2020a] Ji Li, C. Gu, F. Wei, and Xi Chen. A survey on blockchain anomaly detection using data mining techniques. In *BlockSys*, pages 491–504. Springer, 2020.
- [Li *et al.*, 2020b] Y. Li, Y. Cai, H. Tian, G. Xue, and Z. Zheng. Identifying illicit addresses in bitcoin network. In *BlockSys*, pages 99–111. Springer, 2020.
- [Liang *et al.*, 2018] J. Liang, L. Li, and D. Zeng. Evolutionary dynamics of cryptocurrency transaction networks: An empirical study. *PLOS ONE*, 13(8):1–18, 08 2018.
- [Liao *et al.*, 2019] J. Liao, T. Tsai, C. He, and C. Tien. Soliaudit: Smart contract vulnerability assessment based on machine learning and fuzz testing. In *IOTSMS*, pages 458–465. IEEE, 2019.
- [Liu *et al.*, 2021a] X. Liu, X. Jiang, et al. Knowledge discovery in cryptocurrency transactions: A survey. *Ieee access*, 9:37229–37254, 2021.
- [Liu *et al.*, 2021b] Z. Liu, P. Qian, X. Wang, et al. Combining graph neural networks with expert knowledge for smart contract vulnerability detection. *IEEE TKDE*, 2021.

- [Livieris *et al.*, 2020] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas. Ensemble deep learning models for forecasting cryptocurrency time-series. *Algorithms*, 13(5):121, 2020.
- [Mi *et al.*, 2021] F. Mi, Z. Wang, et al. Vscf: automating vulnerability detection in smart contracts with deep learning. In *ICBC*, pages 1–9. IEEE, 2021.
- [Möser *et al.*, 2017] M. Möser, Kyle Soska, et al. An empirical analysis of traceability in the monero blockchain. *arXiv preprint arXiv:1704.04299*, 2017.
- [Mujlid, 2023] Hana Mujlid. A survey on machine learning approaches in cryptocurrency: Challenges and opportunities. In *iCoMET*, pages 1–6. IEEE, 2023.
- [Nakamoto, 2008] S. Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System, 2008.
- [Ortner and Eskandari, 2024] M. Ortner and S. Eskandari. Smart contract sanctuary, 2024.
- [Patel *et al.*, 2020] V. Patel, L. Pan, and S. Rajasegarar. Graph deep learning based anomaly detection in ethereum blockchain network. In *ICNSS*, pages 132–148. Springer, 2020.
- [Patel *et al.*, 2022] V. Patel, S. Rajasegarar, et al. Evangcn: Evolving graph deep neural network based anomaly detection in blockchain. In *ICADMA*, pages 444–456. Springer, 2022.
- [Pierro *et al.*, 2020] G. A. Pierro, R. Tonelli, and M. Marchesi. An Organized Repository of Ethereum Smart Contracts’ Source Codes and Metrics. *Future Internet*, 12(11):197, 2020.
- [Pocher *et al.*, 2023] N. Pocher, M. Zichichi, et al. Detecting anomalous cryptocurrency transactions: An aml/cft application of machine learning-based forensics. *Electronic Markets*, 33(1):37, 2023.
- [Qi *et al.*, 2023] Y. Qi, J. Wu, H. Xu, and M. Guizani. Blockchain data mining with graph learning: A survey. *IEEE Trans. on Patt. An. and Ma. Int.*, 2023.
- [Qian *et al.*, 2022] C. Qian, T. Hu, and B. Li. A bilstm-attention model for detecting smart contract defects more accurately. In *QRS*, pages 53–62. IEEE, 2022.
- [Ron and Shamir, 2013] D. Ron and A. Shamir. Quantitative analysis of the full bitcoin transaction graph. In *FC 2013*, pages 6–24. Springer, 2013.
- [Rouhani and Abedin, 2020] S. Rouhani and E. Abedin. Cryptocurrencies narrated on tweets: a sentiment analysis approach. *IJES*, 36(1):58–72, 2020.
- [Scheid *et al.*, 2022] E. J. Scheid, R. Hy, et al. On the employment of machine learning in the blockchain selection process. *IEEE Transactions on Network and Service Management*, 19(4):3835–3846, 2022.
- [Schmidhuber and Hochreiter, 1997] J. Schmidhuber and S. Hochreiter. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.
- [Shamsi *et al.*, 2022] K. Shamsi, F. Victor, et al. Chartalist: Labeled graph datasets for utxo and account-based blockchains. *NeurIPS*, 35:34926–34939, 2022.
- [Spagnuolo *et al.*, 2014] M. Spagnuolo, F. Maggi, and S. Zanero. Bitiodine: Extracting intelligence from the bitcoin network. In *FC*, 2014.
- [Tang *et al.*, 2023] X. Tang, Y. Du, A. Lai, et al. Deep learning-based solution for smart contract vulnerabilities detection. *Scientific Reports*, 13(1):20106, 2023.
- [Tian *et al.*, 2021] H. Tian, Y. Li, Y. Cai, X. Shi, and Z. Zheng. Attention-based graph neural network for identifying illicit bitcoin addresses. In *BlockSys*, 2021.
- [Victor, 2020] F. Victor. Address clustering heuristics for ethereum. In *FC*, 2020.
- [Wang *et al.*, 2021] J. Wang, P. Chen, S. Yu, and Q. Xuan. Tsgn: Transaction subgraph networks for identifying ethereum phishing accounts. In *BlockSys*, 2021.
- [Weber *et al.*, 2019] M. Weber, G. Domeniconi, et al. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv:1908.02591*, 2019.
- [Wen *et al.*, 2023] M. Wen, R. Lin, et al. Large sequence models for sequential decision-making: a survey. *Frontiers Comput. Sci.*, 17(6):176349, 2023.
- [Wood, 2018] G. Wood. Ethereum: A secure decentralised generalised transaction ledger. <https://github.com/ethereum/yellowpaper>, 2018.
- [Wu *et al.*, 2021] J. Wu, J. Liu, Y. Zhao, and Z. Zheng. Analysis of cryptocurrency transactions from a network perspective: An overview. *Journal of Network and Computer Applications*, 190:103139, 2021.
- [Wu *et al.*, 2022a] J. Wu, J. Liu, W. Chen, et al. Detecting mixing services via mining bitcoin transaction network with hybrid motifs. *IEEE Trans. Syst. Man Cybern. Syst.*, 52(4):2237–2249, 2022.
- [Wu *et al.*, 2022b] M. Wu, W. McTighe, et al. Tutela: An open-source tool for assessing user-privacy on ethereum and tornado cash. *arXiv:2201.06811*, 2022.
- [Wu *et al.*, 2023] J. Wu, K. Lin, Dan Lin, et al. Financial crimes in web3-empowered metaverse: Taxonomy, countermeasures, and opportunities. *IEEE Open Journal of the Computer Society*, 4:37–49, 2023.
- [Xia *et al.*, 2021] F. Xia, K. Sun, et al. Graph learning: A survey. *IEEE Trans. Artif. Intell.*, 2(2):109–127, 2021.
- [Xie, 2019] Rain Xie. Why china had to ban cryptocurrency but the us did not: a comparative analysis of regulations on crypto-markets between the us and china. *Wash. U. Global Stud. L. Rev.*, 18:457, 2019.
- [Yang *et al.*, 2023] Z. Yang, G. Zhang, J. Wu, et al. A comprehensive survey of graph-level learning. *arXiv preprint arXiv:2301.05860*, 2023.
- [Yousaf *et al.*, 2019] H. Yousaf, G. Kappos, and S. Meiklejohn. Tracing transactions across cryptocurrency ledgers. In *USENIX Security*, 2019.
- [Yu *et al.*, 2021a] L. Yu, N. Zhang, and W. Wen. Abnormal transaction detection based on graph networks. In *COMPSAC*, 2021.
- [Yu *et al.*, 2021b] S. Yu, J. Jin, Y. Xie, J. Shen, and Q. Xuan. Ponzi scheme detection in ethereum transaction network. In *BlockSys*, 2021.
- [Yuan *et al.*, 2020] Z. Yuan, Q. Yuan, and J. Wu. Phishing detection on ethereum via learning representation of transaction subgraphs. In *BlockSys*, 2020.
- [Z. *et al.*, 2021] Lin Z., S. S. Gupta, A. Khan, and R. Luo. Temporal analysis of the entire ethereum blockchain network. In *WebConf*, 2021.
- [Zhang and Y., 2021] X. Zhang and Gel Y. Eager: Collaborative research: Blockchain graphs as testbeds of power grid resilience and functionality metrics, 2021.
- [Zhang *et al.*, 2023] Z. Zhang, B. Luo, S. Lu, and B. He. Live graph lab: Towards open, dynamic and real transaction graphs with NFT. *CoRR*, abs/2310.11709, 2023.
- [Zhuang *et al.*, 2021] Y. Zhuang, Z. Liu, P. Qian, Q. Liu, X. Wang, and Q. He. Smart contract vulnerability detection using graph neural networks. In *IJCAI*, 2021.