Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

SuPoolVisor: a visual analytics system for mining pool surveillance*

Jia-zhi XIA†1, Yu-hong ZHANG†1, Hui YE1, Ying WANG1, Guang JIANG1, Ying ZHAO†‡1, Cong XIE2, Xiao-yan KUI1, Sheng-hui LIAO1, Wei-ping WANG1

[1]*School of Computer Science and Engineering, Central South University, Changsha 410083, China*

[2]*Facebook, New York 10003, USA*

†E-mail: xiajiazhi@csu.edu.cn; zhangyuhong@csu.edu.cn; zhaoying@csu.edu.cn

Received Sept. 28, 2019; Revision accepted Feb. 2, 2020; Crosschecked Mar. 6, 2020

Abstract: Cryptocurrencies represented by Bitcoin have fully demonstrated their advantages and great potential in payment and monetary systems during the last decade. The mining pool, which is considered the source of Bitcoin, is the cornerstone of market stability. The surveillance of the mining pool can help regulators effectively assess the overall health of Bitcoin and issues. However, the anonymity of mining-pool miners and the difficulty of analyzing large numbers of transactions limit in-depth analysis. It is also a challenge to achieve intuitive and comprehensive monitoring of multi-source heterogeneous data. In this study, we present SuPoolVisor, an interactive visual analytics system that supports surveillance of the mining pool and de-anonymization by visual reasoning. SuPoolVisor is divided into pool level and address level. At the pool level, we use a sorted stream graph to illustrate the evolution of computing power of pools over time, and glyphs are designed in two other views to demonstrate the influence scope of the mining pool and the migration of pool members. At the address level, we use a force-directed graph and a massive sequence view to present the dynamic address network in the mining pool. Particularly, these two views, together with the Radviz view, support an iterative visual reasoning process for de-anonymization of pool members and provide interactions for cross-view analysis and identity marking. Effectiveness and usability of SuPoolVisor are demonstrated using three cases, in which we cooperate closely with experts in this field.

Key words: Bitcoin mining pool; Visual analytics; Transaction data; Visual reasoning; FinTech
https://doi.org/10.1631/FITEE.1900532 **CLC number:** TP39

1 Introduction

Cryptocurrencies like Bitcoin (Nakamoto, 2008), with their innovative decentralized design, have profound impacts on payment and monetary systems today (Böhme et al., 2015). However, frequent issues, such as illegal transaction and mining monopoly, have made the security and stability of Bitcoin questionable. Volatile Bitcoin values in the past two years have involved it in the crisis of the economic bubble. Due to surveillance difficulties, the implementation and development of cryptocurrencies have been seriously hindered.

Mining pools (Lewenberg et al., 2015) play a critical role in cryptocurrency systems because they generate most of the cryptocurrencies. Therefore, they are the focus of cryptocurrency surveillance. A mining pool is a community composed of operators and miners. Miners integrate their computing

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61872389, 61502540, 61672538, 61872388, and 61772556), the Natural Science Foundation of Hunan Province, China (Nos. 2015JJ4077, 2019JJ40406, and 2017JJ2330), the Changsha Science and Technology Plan Key Project, China (No. kq1801066), and the Fundamental Research Funds for the Central Universities of Central South University, China (No. 2018zzts065)

ORCID: Jia-zhi XIA, https://orcid.org/0000-0003-4629-6268; Ying ZHAO, https://orcid.org/0000-0002-4200-5200

power to calculate the puzzle and win the competition, and pool operators distribute the block reward to each miner's wallet address according to its contributions. Thus, pooled mining is the result of the mining market evolution toward concentration and normativeness, which may lead to the crisis of a mining monopoly if feasible regulation is not deployed. The mining monopoly will then seriously undermine decentralization, and finally trigger a 51% attack (Kroll et al., 2013), meaning that blocks are no longer believable. Mining pools also play a key role in the construction, development, and maintenance of the market infrastructure (Bohr and Bashir, 2014). Many mining pools have emerged since 2010, and they found more than 62% of all blocks so far. The loss of computing power of mining pools will lead to a decline of the blocks of the whole network, and insufficient blocks recording transactions will cause system congestion. In addition, the covert relationship between the pool and other roles will result in a crisis of confidence in the market. Further regulation of the mining pool is imminent.

Many studies of cryptocurrency have emerged. Research on macro market regulation can help us understand the external phenomena of cryptocurrencies, such as price changes (Kim et al., 2016) and community distribution (Bohr and Bashir, 2014). In detail, some studies, such as Meiklejohn et al. (2013) and Belotti et al. (2018), correlate the addresses with public identities by heuristic clustering rules, which can bring partial de-anonymization, but these rules cannot keep up with the changes in deceptive scenarios. Analysis relying on public identities is limited, and de-anonymization requires expert experience and reasoning capabilities. Ron and Shamir (2013) and Ranshous et al. (2017) have tracked and traced the Bitcoin flow between blocks case by case, but they cannot analyze the impacts of the Bitcoin flow in mining pools or determine the correlation between addresses and mining pools. Visual analytics can help users improve data perception, reveal hidden patterns, and interact with data (Liu SX et al., 2014, 2018). Yue et al. (2019) proposed intuitive visualization of the evolution of exchange, but their system cannot explore the underlying fine-grained transactions, making its domain-aware analysis difficult for users. McGinn et al. (2016) used large-screen visualization technology and force-directed layout to give a top-down presentation of transaction data in blocks, but there was a lack of comprehensive multi-aspect analysis capabilities and guidance (Chen HD et al., 2014; Wang XM et al., 2018, 2019).

In-depth research on pool surveillance is still in its infancy. Previous works fail to analyze mining pools from transactions and lack verification of real data due to the following challenges: First, the anonymity feature makes it difficult to find mining pools in the transaction data. Apart from a few public addresses, other addresses can be regarded only as accounts, which are difficult to connect with real roles. Second, the transactions are too massive to analyze. Third, it is challenging to achieve intuitive and comprehensive monitoring of large-scale, multi-dimensional, and time-series data.

To address these challenges, we propose the pool-centered visualization system called SuPoolVisor, which supports comprehensive visual surveillance of pools in the community from the pool level and address level. At the pool level, the interface presents the key statistics of the mining pools and the relationship between mining pools, which are used to supervise the external performance and macro impact of the mining pool. At the address level, the interface presents the patterns of temporal address behaviors, distribution structures, and income sources. It supports interactive identification of address identities to help users find miners in reward distribution, i.e., de-anonymization. Five views are designed and implemented to achieve these purposes. In addition, we have examined three cases to verify the effectiveness of the system in solving important domain problems.

The major contributions of this study are as follows: (1) a visual analytics system for supervising mining pools at the pool level and the address level; (2) a set of features for describing the behaviors and effects of mining pools; (3) an interactive address identification approach for disclosing miners.

2 Related works

2.1 Bitcoin market and mining pool analysis

Various studies have sprung up from many related fields due to the thriving interest in Bitcoin (Yli-Huumo et al., 2016), and cover a wide range of topics including security (Vasek et al., 2014; Vasek and Moore, 2015), privacy (Koshy et al.,

2014; Meiklejohn and Orlandi, 2015), usability (Spagnuolo et al., 2014), and wasted resources (Barkatullah and Hanke, 2015). In specific applications, these works can be summarized into two categories: macro-analysis of statistical data and micro-analysis of transaction data.

1. Macro-analysis

The macro-analysis of cryptocurrency includes its price, market security and risk, and decentralization. Athey et al. (2016) found that the main factor affecting Bitcoin price is user confidence through the analysis of virtual currency data and social media data. Kim et al. (2016) trained a prediction model via machine learning based on commentary data from the cryptocurrency community to predict fluctuations in the price. Moore and Christin (2013) built an exchange risk model to predict the likelihood of its closure and security risks. Kiran and Stannett (2015) introduced agents in the traditional modeling process to analyze the risk of using Bitcoin. Gencer et al. (2018) measured the decentralization of Bitcoin and Ethereum by measuring the characteristics of the Bitcoin network to respond to the public questions about the decentralization of cryptocurrencies. Bohr and Bashir (2014) examined how cryptocurrencies work by the statistical distribution of cryptocurrency accounts. Most of the macro-analyses focus on the statistical properties of cryptocurrencies such as prices, and analyze their relationship with external data. However, these studies do not provide a deep analysis of the transaction data, and it is difficult to find the underlying causes of external performance.

2. Micro-analysis

The micro-analysis of transaction data has been focused on the tracking and analysis of trading cases. Ron and Shamir (2013) defined typical trading patterns of the use, storage, and transfer of Bitcoins by tracking large transactions. Ranshous et al. (2017) explored trading patterns by training classifiers for directed graph nodes to reveal criminal activities such as money laundering. For account anonymity issues, Fleder et al. (2015) used a part of public identity, Meiklejohn et al. (2013) proposed several rules of heuristic account consolidation, and Neudecker and Hartenstein (2017) explored how the network information pair helps for address clustering. In addition, Ober et al. (2013) analyzed the factors influencing anonymity in terms of transaction networks. However, these micro-analyses are mainly case by case analyses, so it is difficult to effectively analyze and mine large-scale data. This study focuses mainly on the study of mining pools.

Research on mining pools is still in its infancy. Current works focus on the simulation analysis of pooled mining and the case by case exploration of miners. Luu et al. (2015) verified the flaws in the protocol and found possible attacks by simulating the computational power splitting game. Lewenberg et al. (2015) used game theory models to verify the transfer of miners for maximum income. Wang LQ and Liu (2015) explored miners in the pool through expert experience and revealed the typical iterative distribution model. Belotti et al. (2018) applied a method based on Union Find to find miners, and analyzed the pool-hopping of the miners between pools. Pool-hopping is a way to maximize miners' income by switching mining pools. However, these works rely on limited public data and do not help find a general way to disclose mining-pool structures and identify miners.

2.2 Blockchain data visualization

The nature of blockchain data makes it closely linked to visualization, which combines computational intelligence of the machine with human perception intelligence to give insight into the underlying patterns. On some websites, visualization has been applied to the analysis of blockchain data. The visualization of real-time transaction data (https://bitbonkers.com, https://bitcoin.interaqt.nl, https://bitnodes.earn.com, and https://blocks.wizb.it) and the traceability of transaction data (https://www.elliptic.co and https://mapofcoins.com/bitcoin) become possible. However, these instruments present only static data details and do not provide efficient interactive analysis. Recently, visual analytics has been gradually applied to block data analysis. Di Battista et al. (2015) visualized the mixed patterns of flows in the Bitcoin transaction graph, and used tailored flow charts to illustrate suspicious Bitcoin flow for illegal transactions such as money laundering. Bistarelli and Santini (2017) explored miners, sources, and branches of the Bitcoin flow using visual analytics. McGinn et al. (2016) deployed the top-down visualization of blockchain transaction data on large display devices, which can present overall dynamic transaction patterns in the

block. Most works use transactions as the objects of analysis, and rarely involve research on entities such as users and organizations. Regarding users, Meiklejohn et al. (2013) used heuristic clustering algorithms with graph visualization to characterize user networks. Isenberg et al. (2017) proposed the design of an interactive visual interface for presenting an individual transaction behavior. Kinkeldey et al. (2017) revealed the transaction histories of entities with visual analytics. Regarding exchanges, Yue et al. (2019) proposed BitExTract, which is the first attempt to explore the evolutionary transaction patterns of Bitcoin exchanges from two perspectives, namely, exchange versus exchange and exchange versus clients. However, existing works present mainly the external performance, but do not conduct an internal analysis such as exploring transaction. More importantly, the interaction of visual analysis has not been used for facilitating de-anonymization. At present, research on visual analysis of mining pools is still scarce.

2.3 Time-series visualization

There are many visualization technologies for time-series data (Aigner et al., 2011; Luo et al., 2019), which are widely used in smart manufacturing (Zhou FF et al., 2019; Zhao et al., 2020), sports analysis (Wu et al., 2019), situational awareness (Ying et al., 2019), and traffic analysis (Zeng et al., 2017). In visual space, time is generally represented as an axis. The most prevalent approach is using a horizontal axis and encoding data along it, such as a histogram (Chen W et al., 2016; Chen SM et al., 2018), a stacked stream graph (Xia et al., 2019; Zhou ZG et al., 2020), or a line chart (Li et al., 2020). Custom glyphs or views are placed in chronological order in some designs (Jie et al., 2019). Advantages of them are intuitive and interactive. For example, Mei et al. (2019) used river views with different time granularities to visualize large-span time data and used the brush to filter and zoom. Many works use circular axes to represent time, making full use of space (Zhou ZG et al., 2019). Circular axes are widely used in the form of time wheels as tools for exploring spatial-temporal patterns (Zhou ZG et al., 2017). Multiple concentric circular axes make it convenient to observe the periodicity (Zhu et al., 2019). In three-dimensional (3D) space, Chen W et al. (2018a) used the Z axis to represent the time information of urban data.

Visualizing e-transaction time series is critical for understanding transaction behavior and market conditions. Many studies focus on discovering temporal trends of transaction behavior (Liu ZC et al., 2009; Wei et al., 2012). However, interesting transactions that are situation relevant might be ignored. Xie et al. (2014) designed KnotLines to show detailed transaction information. It was inspired by the musical notation in which different transactions and their connections were placed along the time axis. Although relationships among the transactions made by the same seller are emphasized, visual analysis of the complete transaction network is still insufficient. The visualization of funding flows requires more comprehensive methods.

3 Problem characterization

As shown in Fig. 1, our system comprises three parts: (1) the data manager, which pre-processes the raw data and constructs transaction networks; (2) the features, which characterize the behaviors of addresses and mining pools for analysis and de-anonymization, respectively; (3) the visual analysis, which includes five well-designed views and interaction methods.

3.1 Data abstraction

3.1.1 Raw data collection and pre-processing

Bitcoin raw transaction data is the most accurate and basic analysis material, and contains all the anonymous transactions recorded on the block. There are multiple public download channels. We chose the transaction data (January, 2009) from blockchain.com, which is considered reliable (Möser et al., 2013). It is convenient for pre-processing because it has already been parsed into JSON format. As shown in Fig. 2, the attributes

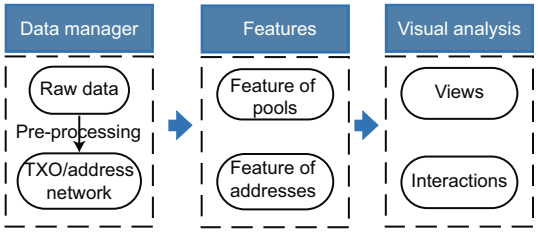

Fig. 1 Overview of SuPoolVisor

of transactions are as follows:

1. Block ID: the identifier of the block, which indicates the block that records the transaction.

2. Block time: the timestamp when the transaction is recorded. All the transactions on the same block have the same time.

3. Transaction ID: the 64-character hash of the transaction, which identifies solely a transaction.

4. isCoinbase: the symbol indicating whether the transaction is a coinbase transaction. The coinbase transaction is the first transaction in the block, where the system pays the block reward to the mining pool or the miner who found the block.

5. Output list: the array recording the transaction outputs (TXOs) and the corresponding addresses. The unspent transaction outputs (UTXOs) will not be separable until it is spent in the next transaction.

6. Input list: the array recording the previous TXOs which are spent in this transaction and the corresponding addresses. If the transaction is a coinbase transaction, this item is empty.

To know the pools that found the blocks, we collected broadcast data from BTC.com, which records the messages when pools find a specific block. Each broadcast data includes the block ID, block time, and the pool name who broadcasts it. Correlating the block and broadcasting, the data can reveal the transactions from which pools receive rewards. Also,

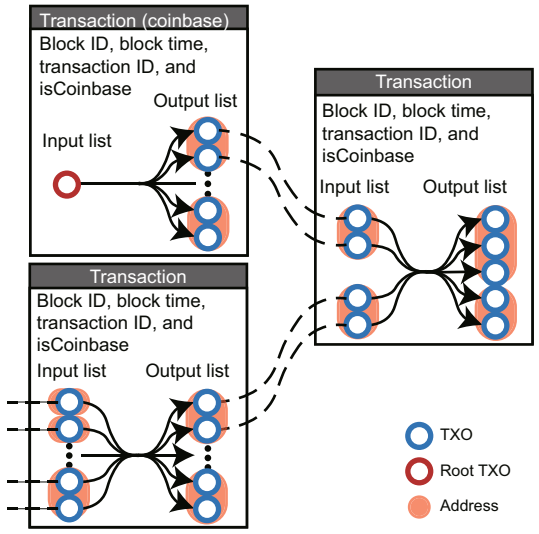

Fig. 2 Transaction attributes and their connections
New transactions spend previous transactions' outputs (TXOs) and generate new outputs (TXOs), making Bitcoin transfer among addresses

we collected publicly identified addresses from WalletExplorer.com, including some exchanges, service providers, mining pools, and other addresses. Identity data is incomplete, but it can be used to assist in our analysis process.

3.1.2 Transaction network construction

The transaction network starting from the coinbase can help find miners. Due to the block verification delay and payment methods, instead of paying miners in coinbase transactions, the pool operator often uses the transferring transaction to transfer funds and then pays the miner by the rewarding transaction. The transaction sub-network consists of the coinbase, and transferring and rewarding represents the Bitcoin flow inside the mining pool.

To observe the data from both the transaction and the address, we constructed two transaction networks: one is the TXO network, which takes TXO as the node and is used to express the original Bitcoin flow, and the other is the address network, which takes the address as the node and is used to express the transaction relationship between addresses. In the TXO network, as shown in Fig. 2, since the TXO in the coinbase is not derived from any other transactions but the system, we virtualized the block rewards into a special type of TXO (i.e., the root TXO) and treated them as nodes. Then, based on the principle that the source of each output in each transaction is proportionally derived from each input, we split the M-to-N transaction into $M \times N$ 1-to-1 transactions. We constructed the address network through the TXO network. According to the ownership relationship between TXOs and addresses, we can obtain the relationship between addresses through the accumulation of the relationship between TXOs. The edge between addresses represents the set of 1-to-1 transactions. Based on the block broadcast, we added identity tags to the address network.

3.2 Task analysis

Research and regulation of Bitcoin transactions is a subject related to cryptography, statistics, and finance. To better describe the domain issues and meet the needs, we have been working closely with three experts for the past six months. Expert E_1 focuses on Bitcoin, cryptography, and blockchain

security. He is curious about pool surveillance. E_2 has been a deep participant in the Bitcoin mining pool since 2013. E_3 is a Bitcoin regulatory practitioner who is looking for an efficient method of pool management. All experts have shown great enthusiasm and have been actively involved in our work. Following a typical user-centered design framework, including discussion, brainstorming, design, prototyping, and presentation, we collected their feedback and summarized two aspects of the analysis tasks to implement mining pool surveillance:

T_1: analysis at the pool level

T_1-1: Present the computing power of mining pools in the market. The actual computing power level of each mining pool is the core competitiveness of the mining pool. The computing power of the whole network is closely related to the health of the system.

T_1-2: Present member migration between mining pools. Pool-hopping of miners, changes of the pool itself, and sharing of nodes between pools will all result in migration of members between pools.

T_1-3: Present the impacts of mining pools in the market. Mining pools act as upstream agencies and have impacts on the market by generating Bitcoin.

T_2: de-anonymization and analysis at the address level

T_2-1: View time-series transaction patterns of addresses. Mining pools often pay rewards to miners through the rewarding transaction, which is considered an internal transaction in the pool.

T_2-2: View the transaction structure and path of address. Experts want to know the structure of different internal members in the distribution process.

T_2-3: View the source distribution of address income. In the mining pool, the source analysis of the addresses can help reveal the stability of the market.

4 Feature abstraction

We used features of addresses to initially determine their identity in mining pools. We defined three identities: pool operators, miners, and external addresses. Pool operators are the internal administrators of pools. Miners are the final recipients of the rewards because of providing computing power. External addresses are the general participants who have nothing to do with the mining process. We proposed the following characteristics as an important basis to infer the identities:

1. The amount of income

The amount of address income can describe the transactions in which addresses participate. The amount paid by the operators to the miners is stable or fluctuates within a certain range, and there is always a minimum for the payment in most pools. For example, an address that receives only 0.000 000 1 Bitcoin at a time is unlikely to be a miner, because the amount does not reach the threshold of payment for most pools. We can filter out non-pool transactions by this feature. For a given address, we divide the total income by the number of times to obtain its average amount of income.

2. The frequency of income

The payment relationship between operators and miners is more frequent and periodic than the ordinary transaction relationship. We can filter out non-pool transactions by this feature. For a given address a, we abstract the receiving fund in n days as one-hot vector: $\boldsymbol{V}_a = [r_1, r_2, \ldots, r_n]$. If the address received the fund on the k^{th} day, r_k is 1; otherwise, it is 0. Then we calculate the frequency by $F_a = r_1 + r_2 + \ldots + r_n$.

3. The purity of income

The income of ordinary miners is relatively simple in single or multiple transactions. An address that receives funds from only a specific pool is more likely to be an operator or loyal miner of it. We propose TXO purity (TP) and address purity (AP) to measure the purity of these two cases. TP represents the proportion of TXO from each pool. We pre-calculated the TP of each TXO as follows:

(1) We defined $\text{amount}_{i,j}$ as how many Bitcoins in TXO_i are generated by pool_j. All Bitcoins in root TXO are generated by a specific mining pool.

(2) We obtained all amount from each pool to each TXO by tracing the source. Because TXO is ordered in generation time and can be consumed only once, according to the TXO network, the amount of output TXOs can be calculated by the amount of input TXOs.

(3) We calculated the TP of the i^{th} TXO by the following formula:

$$\text{TP}_i = \max \frac{\text{amount}_{i,j}}{\sum_{j=1}^{m} \text{amount}_{i,j}}. \tag{1}$$

(4) Let address a have n TXOs which are from m pools. We calculated the AP of a to j by the

following formula:

$$\mathrm{AP}_{a,j} = \frac{\sum_{i=1}^{n} \mathrm{amount}_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{amount}_{i,j}}. \quad (2)$$

4. Recommended internal address network

Before users' identity reasoning process, we recommended every pool's internal address network to users by the automation process, which is divided into three steps. First, we selected the TXO network in which TP=1, namely, the transaction sub-network. As shown in Fig. 3a, the transaction sub-network consists of the coinbase, transferring, and rewarding, which are not mixed with the TXOs of other pools. Second, we constructed the corresponding address network according to the TXO network, as shown in Fig. 3b. Finally, we traversed the address network from the coinbase to terminal addresses that do not meet thresholds or are labeled as exchanges or service providers. The terminal addresses are regarded as the destinations of the internal address network.

5. Migration between pools

After the recommendation, we could find the migration addresses between the pools by calculating the intersection of their internal address network. Based on the order in which addresses appeared in different pools, we divided the migration addresses into moving in, moving out, and hopping in units of days.

5 System design

As illustrated in Fig. 4, our visual analysis system consists of six components: sorted stream view (SSV), migration timeline view (MTV), threshold panel, massive sequence view (MSV), pool structure view (PSV), and Radviz. The SSV, MTV, and Radviz support pool-level comparison and analysis. The threshold panel, MSV, PSV, and Radviz support address-level de-anonymization and visual analysis.

5.1 Visualization design

5.1.1 Sorted stream view

Description: SSV provides an overview of the temporal distribution of the number of blocks generated by each pool, which further reflects the computing power distribution and evolution of Bitcoin (T_1-1). Streams are dynamically sorted to compare the evolution of computing power in different pools. The ability of pools to generate blocks reflects their influence (T_1-3).

As illustrated in Fig. 4a, the X axis represents a timeline, and different pools are stacked as streams along the Y axis based on the number of blocks they found. The different colored streams represent the corresponding pools, and their width on axes indicates the fluctuating number of blocks found with time. Unlike the usual stream graph which has a

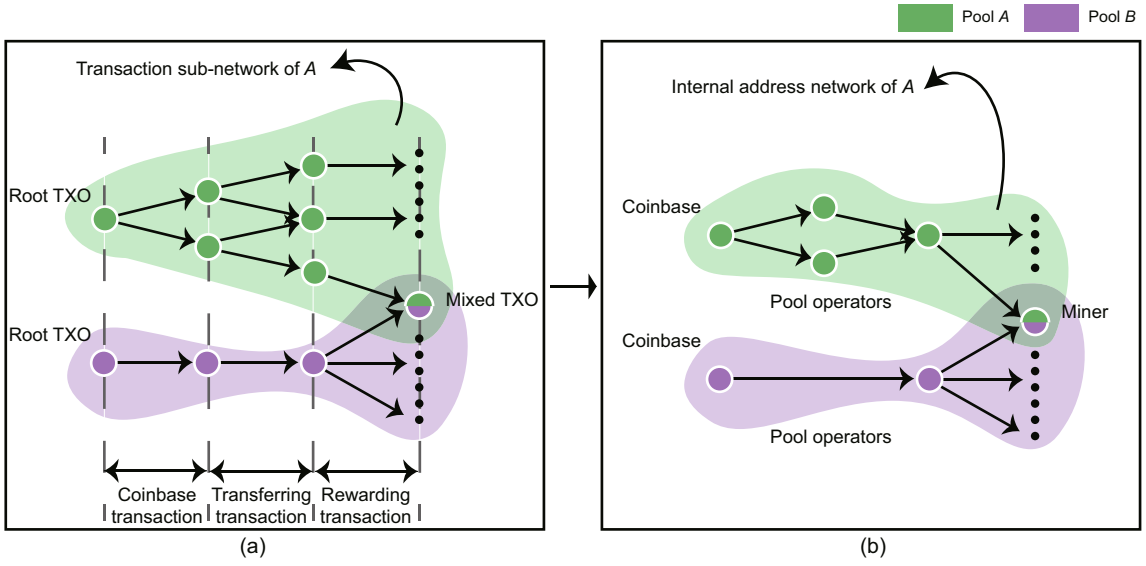

Fig. 3 Construction of the transaction sub-network (a) and the internal address network (b)

Unmixed TXOs are used to find the sub-network of mining pools and the recommended internal address network can be constructed based on the mapping from TXOs to addresses

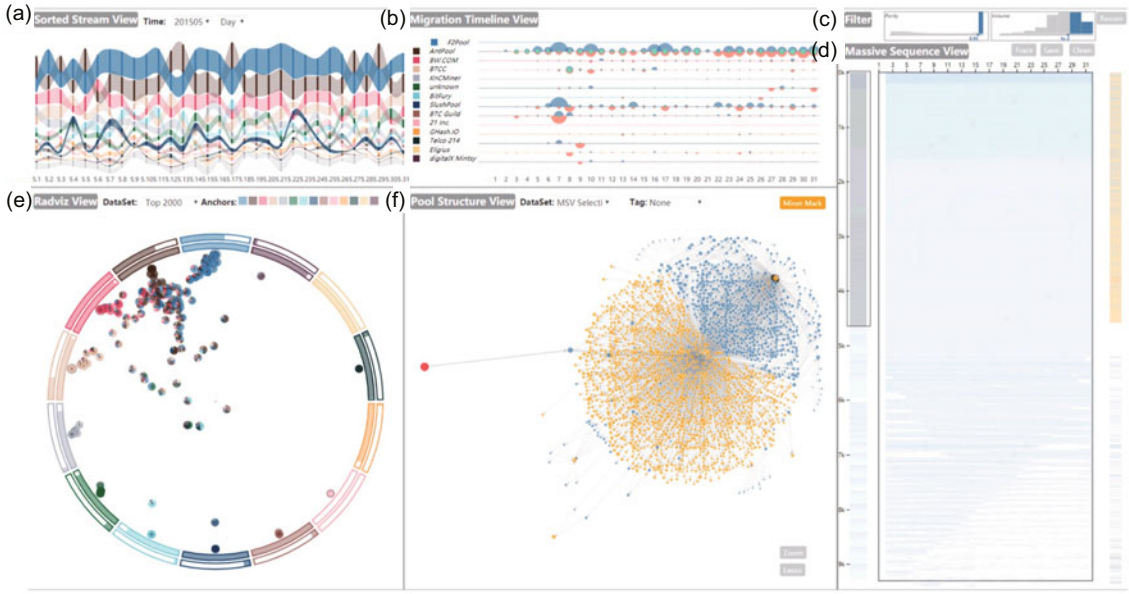

Fig. 4 Interfaces of SuPoolVisor: (a) sorted stream view; (b) migration timeline view; (c) threshold panel; (d) massive sequence view; (e) Radviz view; (f) pool structure view

References to color refer to the online version of this figure

fixed order, streams are re-sorted from top to bottom on the axis of each day in SSV. To reduce the visual clutter, the colored top-ranked pools are manageable, while the insignificant small pools are grayed to the bottom. The legend is given on the right side, which can interact synchronously with the view and MTV. From this view, users can intuitively identify the dominant mining pools.

Justification: We considered several alternatives for time-series data visualization, such as a stacked stream graph and multi-line charts. However, the stacked stream graph lacks ranking, which is essential for finding dominant mining pools and comparing evolutionary differences between mining pools (T_1-1). Although multi-line charts can present ranking, the uneven distribution of the data leads to serious visual occlusion; that is, lines at the bottom are too dense to be distinguished. In the sorted stream graph, visual clutter can be resolved by setting transparency. The computing power of each pool is highlighted on every axis, while the semitransparent crossing parts are just for smooth transitions and have no practical numerical significance. The sorted stream graph is effective for the visualization of multiple continuous time-series data, and its nature of stacking supports sorting and displaying the sum.

5.1.2 Migration timeline view

Description: MTV provides an overview of the temporal pattern of the member migration between pools, which reflects pool-hopping and miner migration (T_1-2). For a given pool, users can observe when and how many the addresses move in, move out, and pool-hopping to other pools at different timestamps.

As illustrated in Fig. 4b, the X axis represents the timeline. Consistent with SSV, different pools are stacked on the Y axis and the order is the same. Each row expresses the member migration between this pool and another specific pool every day in the selected month. We used the same color encoding scheme to the baseline to facilitate the identification of pools. At the nodes of each day in each line, we designed a glyph to visualize the migration details.

As illustrated in Fig. 5a, the glyph is designed to visualize the patterns of moving in, moving out, and pool-hopping of migratory addresses. The area of the upper half ring represents the number of addresses moving in. The area of the lower half ring represents the total number of addresses moving out. The inner-circle area represents the number of addresses that are pool-hopping. The metaphor of the design is that pool-hopping can be interpreted as a common part of moving in and moving out. Each glyph is placed in a limited square to avoid overlap, and the mapping

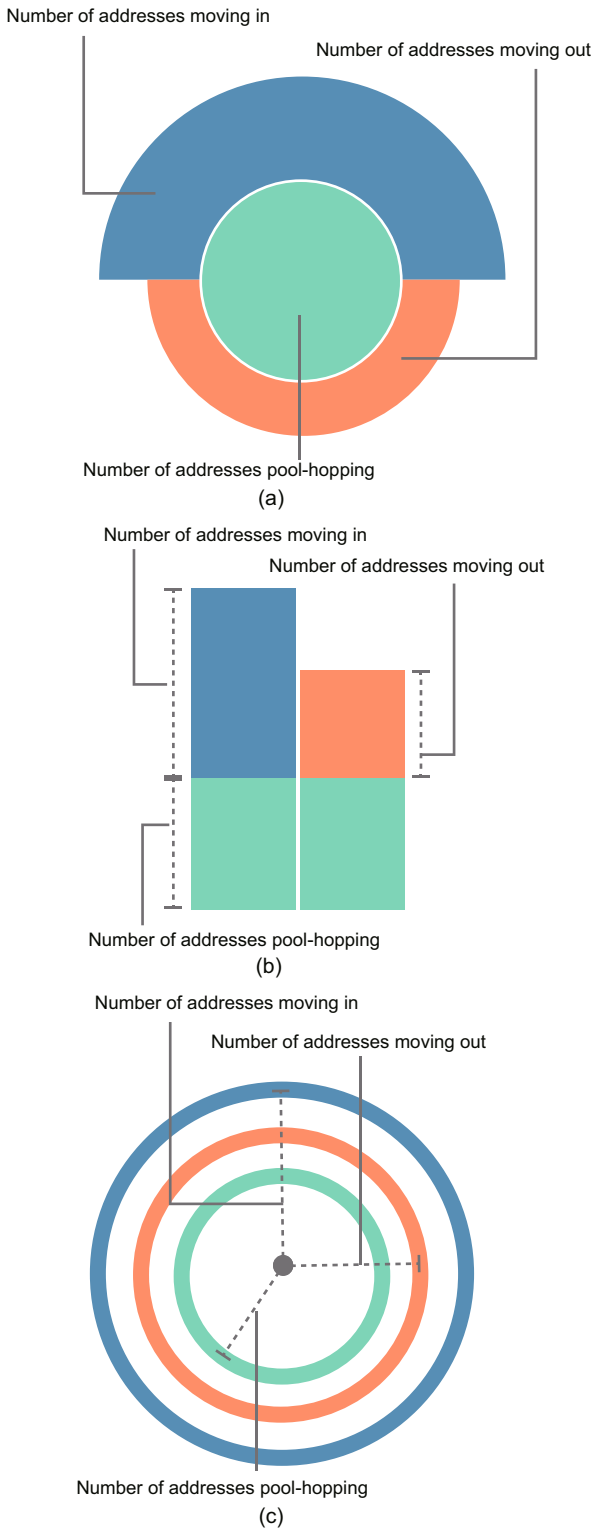

Fig. 5 Alternatives to glyph design in MTV: (a) design based on the radial histogram and donut; (b) design based on the stacked histogram design; (c) design based on ring

References to color refer to the online version of this figure

scale of values to space can be adjusted.

Justification: Experts want to observe the hopping pattern and the comparison between addresses that are moving in and moving out (T_1-2). Besides the chosen glyph design, we considered some alternatives. As Figs. 5b and 5c illustrate, two other glyphs are designed. We were inspired by the stacked histogram (Fig. 5b). The heights of the blue and orange bars represent the numbers of addresses moving in and moving out, respectively. The height of the public part of the lower part represents the number of addresses that are pool-hopping, which uses the same metaphor as the chosen design. However, this scheme does not make full use of the encoding space in the horizontal direction. Relying on the limited vertical space will result in visual occlusion and clutter when the value is too large and visual obscurity when the value is too small. In Fig. 5c, we used the radial layout to make full use of the space. The radii of the three rings represent the numbers of addresses that are moving in, moving out, and pool-hopping. However, three concentric circles sharing the same coding area may also occlude each other.

5.1.3 Threshold panel

As illustrated in Fig. 4c, we provided a visual control component to select the filtering threshold. Addresses in the selected ranges of purity and income will be retained as potential miners after filtering. The threshold can be set by the user or be defaulted to the recommended value. To help users understand the data intuitively, we show the histogram of the data distribution above the slider, allowing users to set the threshold based on the reality and domain knowledge, and the selected part will be highlighted.

5.1.4 Massive sequence view

Description: MSV provides temporal behavior visualization of addresses after filtering (T_2-1). As illustrated in Fig. 4d, the X axis represents the time, and massive addresses are stacked along the Y axis. For each address, the timeline represents the funds it received during the month, and saturation encodes the amount of the funds.

To see an overview of many addresses and explore details on demand, we selected the context+focus design. A guidance axis is applied as context on the left of MSV. The scroll bar on the

axis can be adjusted and dragged to change the visible addresses as the focus. Also, MSV provides two tag lists to save the identity tags and explore history for the addresses when analysis begins.

To reveal address groupings and identify pivotal addresses in MSV, we designed an address-ranking algorithm based on principal component analysis (PCA). The goal is to rank the addresses based on how often they receive funds and to make addresses with similar behavior patterns closer. The following is the process: (1) based on the frequency, divide all addresses into n bins; (2) reduce the dimensionality of $\boldsymbol{V}_i$ to one by PCA in each bin, and sort addresses by their 1-d values; (3) remove the sorted addresses from the n^{th} bin to the 1^{st} bin in order.

Justification: Revealing an overall pattern and observing details while visualizing large-scale time-series data is a challenge. Context+focus with guidance can resolve the contradiction between an overview and observing details. The address-ranking algorithm based on PCA can group addresses in the order of frequency and amount of funds received. Another important function of MSV is to help explore and save the identity of the address by interaction. The addresses selected in MSV are used as de-anonymization entries. The temporal patterns, network structures, and income sources of the addresses are further analyzed in multiple views, and the final identity results can be saved on the right side of MSV.

5.1.5 Pool structure view

Description: PSV visualizes pool's internal address network (T_2-2), which can be used to disclose the structures of mining pools and reveal the destination of the fund stream. For specific anonymous addresses, the sub-network from the coinbase to the addresses will be presented as important evidence of de-anonymization.

As illustrated in Fig. 4f, PSV presents an address network using a force-directed layout, which encodes the identity information with color and encodes the amount of the transactions with the transparency of the edge. A variety of color encoding schemes are provided for marking different identity types. The default color scheme is that the coinbase is red, the middle node is blue, and the end node is yellow, representing the coinbase, pool operator, and miner, respectively. When users interact with MTV, the node uses the color scheme of MTV to display the migrated addresses. The unselected nodes are coded in gray.

5.1.6 Radviz view

Description: We designed an enhanced Radviz (Hoffman et al., 1997; Zhao et al., 2019) to present the distribution of income purity of specific addresses, which reflects the influence of pools. As shown in Fig. 4e, the Radviz view consists of a circle with anchor points on the circumference, which represents pools, and the addresses are projected as glyphs in the circle to characterize the closeness to the different pools (T_2-3).

The purity of n pools of address a_i is considered a high-dimensional vector $\boldsymbol{V}_i = [p_{i,1}, p_{i,2}, \dots, p_{i,n}]$. The location of each address within the circle is computed as a function of its relative attraction to anchors. The attraction to each anchor is proportional to the magnitude of the coordinate for that dimension. Formulas for calculating the resulting transformed coordinates for a_i are expressed as

$$\begin{cases} x_i = \dfrac{\sum_{j=1}^{n} p_{i,j} \cos\theta_j}{\sum_{j=1}^{n} p_{i,j}}, \\ y_i = \dfrac{\sum_{j=1}^{n} p_{i,j} \sin\theta_j}{\sum_{j=1}^{n} p_{i,j}}, \end{cases} \tag{3}$$

where θ_j is the angle on the circle corresponding to dimension j, namely, pool j.

Each Radviz anchor point consists of two arcs. Angles of the inner and outer arcs encode the number of covered user addresses and currency circulation in the market, respectively. Mining pools are encoded by arc color, which is consistent with SSV. The node glyph adopts a pie-chart-based design. The angle of each part in the pie chart represents the purity of the corresponding pool. The radius of the pie itself is uniform, while the radius of the outer black ring encodes the total number of Bitcoins that the account has received.

5.2 Interactions

We provide the following interactions to help experts apply their experience and perform comprehensive analysis:

1. Cross-view analysis

Other views can be used for further analysis if users select objects in a view. When analyzing data

at the pool level, an overview of members' migration will be presented in MTV when users click a pool in SSV. When inferring identities, the same addresses will be highlighted in Radviz and MSV when the user selects a group of addresses in PSV by lasso. Constructing cross-view selections can achieve multi-view and multi-aspect comprehensive analysis of the target of interest.

2. Custom settings

To apply experts' experience, SuPoolVisor will present different visualizations based on users' input. Before recommendation, users can set the threshold panel to customize the recommendation range. In addition, different color schemes can be chosen in PSV for a variety of analysis purposes, such as observing the network location or highlighting the public identity of addresses.

3. Detail amplification

Important details often need to be found and amplified from massive information. In MSV, we deployed the focus+context design to help amplify the focus area on-demand. Part of the complex network can be observed more clearly by zooming in on PSV.

4. Visual reasoning

As shown in Fig. 6, our system supports identity reasoning of addresses by the "project-create-project cycle" (Kirsh, 2009; Chen W et al., 2018b). First, users initialize the threshold panel to obtain the recommended internal address network. Second, users infer the identities of addresses based on the distribution structure, temporal behavior, source distribution, and label information in PSV, MSV, and Radviz. Finally, users save results in the tag list, and these identities can be used in the next loop.

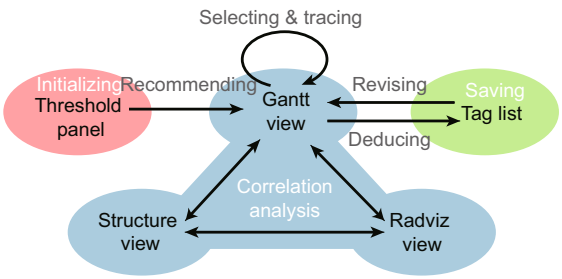

Fig. 6 Interactive reasoning process

6 Case study

The ultimate purpose of SuPoolVisor is to provide users with visual analysis tools for surveillance. To verify the effectiveness of the system, we conducted three case studies with our three experts, who provided the original analysis tasks and were deeply involved in our design process. Experts had different cases based on their tasks and interests. After a brief introduction to the use of the system, experts began their exploration and provided their feedback, which we recorded in separate interviews.

6.1 Identifying miners and operators in mining pools

E_2 is an expert with background knowledge and is interested in exploring the community in pools. His analysis has long relied on WalletExplorer's public data set. It includes many deprecated addresses, but many important and emerging pools are not included. E_2 is looking forward to using SuPoolVisor to infer and find operators and miners.

E_2 selected the data in May 2015 for analysis at the pool level (T_1). It can be seen from Fig. 4a that F2Pool is the dominating pool with the most blocks at that time (T_1-1). It is demonstrated by Radviz (Fig. 4e) that rich addresses gather around F2Pool, and that the Bitcoin generated by F2Pool is widely used in the whole market according to the anchors (T_1-3). E_2 said that F2Pool is the biggest mining pool in China and has maintained its first place in the generation of blocks (8%). However, its addresses were not included in WalletExplorer, so E_2 began to explore its internal addresses (T_2).

1. Initialization

As shown in Fig. 4c, E_2 set the filter condition (AP $\geq$ 95%, amount $\geq$ 0.0001 BTC), and obtained an initial internal address network of F2Pool, which includes about 8000 addresses.

2. Correlation analysis

After recommendation, the overview of addresses was figured out in MSV (Fig. 4d). Obviously, about 2800 wallet addresses received rewarded every day (T_2-1). E_2 selected them and observed their complete process of distribution in Fig. 7b (T_2-2). The start (i.e., coinbase), intermediate, and terminal addresses are in different colors. In PSV, E_2 found that the distribution is divided into three steps: (1) the coinbase gives the rewards to address a each time,

and a transfers them to address c; (2) c transfers the rewards to groups S_1 and S_2; (3) S_2 gathers all the rewards and transfers them to address b. E_2 speculated that addresses a and c are operators who undertake the tasks of receiving, transferring, and paying rewards. E_2 further analyzed them in MSV and Radviz (T_2-1 and T_2-3, respectively). As shown in Fig. 7a, the sources of a, c, and S_2 remain pure, but those of S_1 and address b are complicated. In Fig. 7b, the addresses a, b, and c are the top three in MSV; they maintain a very high income and stable periodicity. Based on this evidence, E_2 believed that a and c are operators, and that S_1 represents stable ordinary miners. E_2 considered S_2 miners to be controlled by the same entity based on their uniform behavior pattern and the same destination, and that b is a controller.

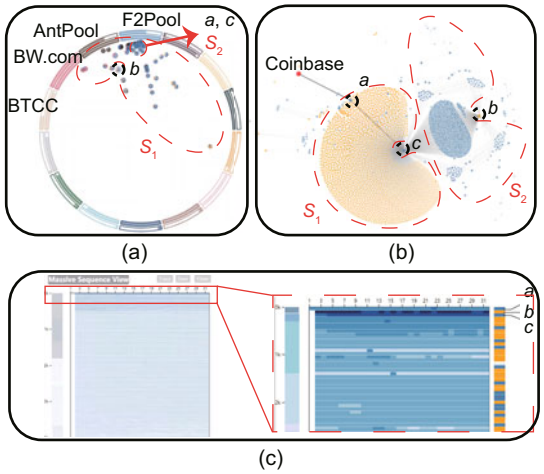

Fig. 7 multi-aspect analysis of special groups and addresses in F2Pool: (a) special groups and addresses in Radviz; (b) special groups and addresses in PSV; (c) three special addresses in MSV

3. Saving

After analysis, the identities of the selected addresses were inferred. E_2 recorded color tags representing different identities in the exploration history column of the tag list. Then, he cleaned up the addresses behind miners because they are no longer considered miners.

E_2 iterated the above processes until all addresses were identified. He highly praised SuPoolVisor and believed that it can help users find key addresses in mining pools and provide multi-aspect analysis. E_2 believed that the view design and interactions can effectively help and guide users in de-anonymization, and that SuPoolVisor can significantly show reward destinations.

6.2 Summarizing the internal structures of different mining pools

Different mining pools are considered to have different distribution processes because of different payment methods and different numbers of miners. E_1 wanted to use SuPoolVisor to summarize the typical internal structure of mining pools. He selected AntPool, BW.com, and Bitfury for inference and analysis, and identified their internal structures for further summary and comparison (T_2-2).

1. One-to-many payment

E_1 first analyzed BW.com and obtained the internal network structure (Fig. 8a). Two types of operators were found, similar to the first case: one is for receiving rewards and the other for paying rewards to all miners at once. E_1 learned from the background that BW.com adopts the PPS payment model in which the miners play the role of employees. Regardless of the income of the mining pool, it is necessary to pay rewards to miners based on their workloads every day. The pool operator uses "receiving addresses" and "payment addresses" to resolve the contradiction between unstable income and stable expenditure.

2. Iterative payment

E_1 analyzed AntPool. Transaction network of the operators of AntPool is shown in Fig. 8b. The operators were reused to recycle funds. Then, E_1 observed local networks between two operators and found that when an operator transfers Bitcoin to another operator, a small portion is distributed to the same miners; the operation is then repeated in the next transaction. E_1 believed that AntPool transfers rewards by constantly reusing operators. It pays miners iteratively in the process, while the same miners may receive their rewards from different operators in different paying rounds. The iterative payment is mainly due to the excessive number of miners. The payment cannot be completed at one time due to the limitation of block capacity, so it needs multiple iterations. In addition, E_1 found a similar structure for iterative distribution in ViaBTC. It is a widely used payment method, which is verified in Wang LQ and Liu (2015).

3. Linear transferring

E_1 was interested in private pools. He believed

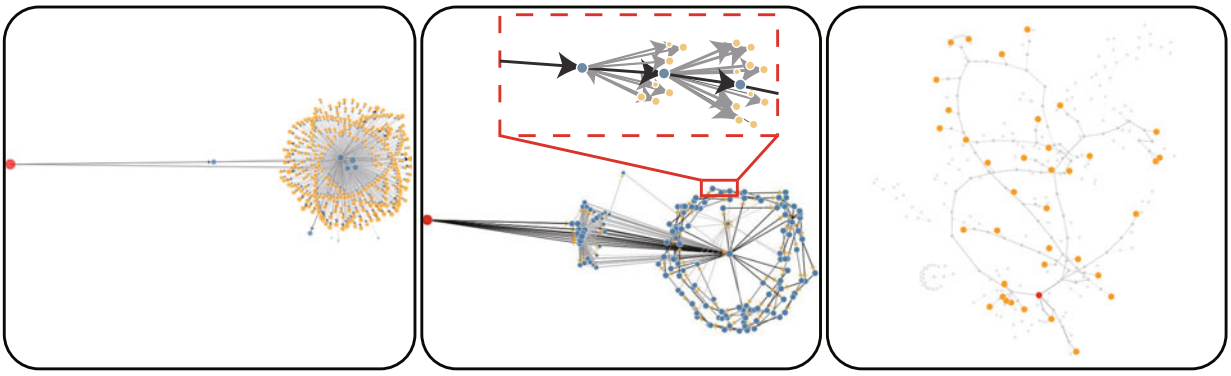

Fig. 8 Three typical internal structures of mining pools: (a) one-to-many payment; (b) iterative payment; (c) linear transferring

that analysis of private pools is meaningful because they are not open to public miners and where the rewards go is a mystery. E_1 chose the largest private pool, BitFury, for analysis. It has a long history, and its private computing power is still at the forefront in the fierce competition in computing power. In the MSV overview, E_1 found that there are very few addresses in the pool (T_2-1), so he selected all the addresses to be drawn in PSV (Fig. 8c). He found that the payment process is similar to the iterative payment, but its operators appeared once and there was only one actual receiver in a transaction. He believed that Bitfury constantly transfers rewards to temporal operator addresses iteratively by making changes. To know the identities of the actual receivers, E_1 switched color encoding to highlight the known identities. PSV showed that half of them are exchanges, and the other half have the same distribution with exchanges in other views. E_1 believed that these receivers are all exchanges and that BitFury sells Bitcoins to exchanges rather than to miners.

6.3 Analyzing the effect of hard forks

E_3 is a Bitcoin regulatory practitioner who wants to investigate the impacts of the hard fork event. The hard fork refers to permanent blockchain disputes. After the new agreed rule is released, some addresses that have not been upgraded cannot verify the blocks produced by the upgraded addresses. Communities could not reach an agreement, so the blockchain is divided into two different chains. The most famous blockchain hard fork event happened on August 1, 2017, when Bitcoin cash (BCC) was generated and became the competitor of BTC. The pool called ViaBTC was an active promoter of the event.

1. Pool level

To investigate the overall status of Bitcoin before and after the fork (T_1), E_3 checked SSV (Fig. 9a) in July and August 2017 (T_1-1). In July 2017, the overall block production was stable, and it did not fluctuate when the fork happened (Fig. 9, P1). However, the pool ranking fluctuated sharply, and the number of blocks decreased significantly on August 20–25, 2017 (Fig. 9, P2). It recovered at the end of the month and the share of blocks of ViaBTC increased significantly. As illustrated in Fig. 9c, Radviz compared the change of income purity of the top 2000 rich addresses in July and August, 2017. E_3 found that the distribution of these addresses is more concentrated after the fork (T_2-3). This means that the Bitcoin produced by the mining pool mixes evenly. E_3 speculated that the profit-driven miners become more sensitive and unstable after the fork, leading to more pool-hopping. To verify this, E_3 observed the MTV (T_1-2) of AntPool, which was the largest pool at that time. As shown in Fig. 9b, the migration in August is much more intense than that in July, and many addresses are indeed pool-hoppers.

E_3 believed that the miners in Bitcoin did not lose when the fork happened. The generation of BTC was not affected by BCC until the revenue of BCC exceeded that of BTC and many big pools started to support BCC mining on August 20, 2017. Lots of miners migrated between pools or went to BCC mining. With the decreased revenue caused by the enhanced difficulty in BCC mining, the overall computing power of the BTC system has been restored.

2. Address level

E_3 wanted to investigate the changes inside pools such as ViaBTC and F2Pool, which are more reactive in this incident (T_2). The former is the promoter and the latter is the opponent. As illustrated in Fig. 10a, E_3 found that ViaBTC miners are periodic and that a lot of miners joined in at the end of August (S1), which agrees with the findings in SSV. In F2Pool, the miners are more obvious and remain active almost every day, but the abnormality is that miners keep silent during P2 and P3 periods. Generally, miners tend to decrease. To determine the cause, E_3 looked up relevant news and found that F2Pool issued several statements to stop paying the miners for financial security in July and August, 2017. As a result, the planner of the fork, i.e., ViaBTC, ultimately benefited, and F2Pool inevitably received a negative impact.

After the top-down investigation on the fork event, E_3 learned the impact of the incident on the Bitcoin mining market and pools. He highly praised SuPoolVisor for its ease of use and intuitiveness. He believed that the system can directly present the state of pools and that multiple views can be used for multi-level analysis and interpretation. In addition, he suggested adding news event tips to the timeline for correlating important events.

7 Conclusions and future work

In this study, we described our development of a visual analytics system, SuPoolVisor, to facilitate regulators and researchers in surveillance and de-anonymization in Bitcoin. SuPoolVisor was divided into the pool level and address level. At the pool level, the evolution of computing power, influence, and members' migration among pools were illustrated. Glyphs were designed to help discover patterns and enhance the performances of MTV and Radviz. At the address level, SuPoolVisor supported visualizing temporal behavior patterns, distribution structures, and income sources of addresses, and multi-aspect analysis can be used to infer their

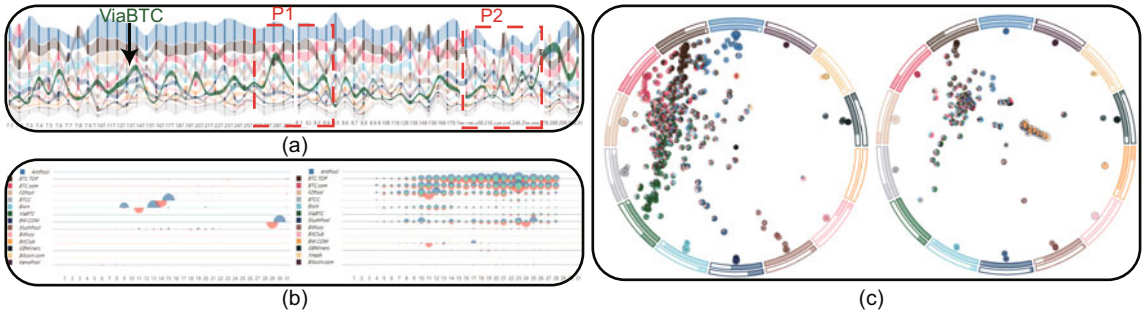

Fig. 9 Overview of Bitcoin market before and after the fork: (a) comparison of SSV in July and August 2017; (b) comparison of MTV in July and August 2017; (c) comparison of Radviz in July and August 2017

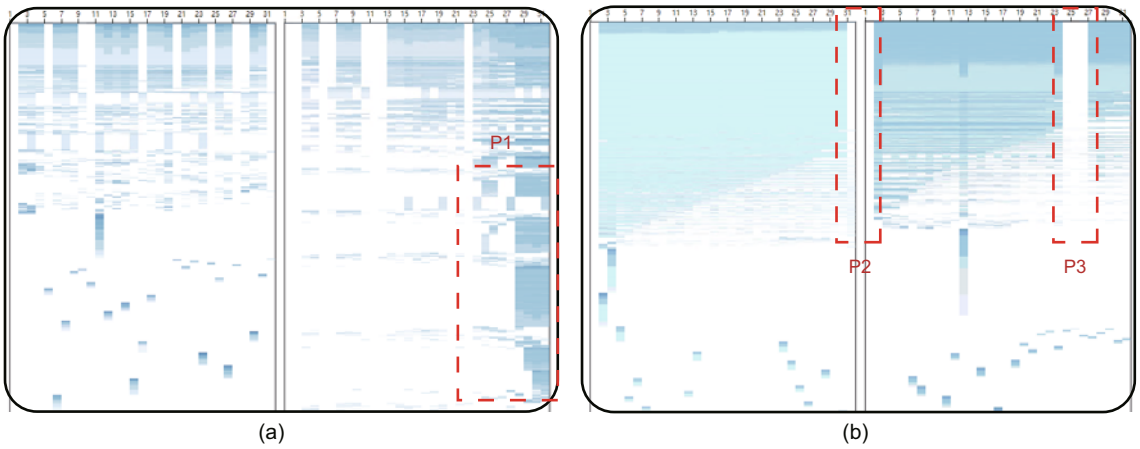

Fig. 10 Comparison of miners' behavior before and after the fork: (a) MTV of ViaBTC; (b) MTV of F2Pool

identities. In particular, we proposed a series of characteristics to identify miners, and the initial internal address network was recommended based on them:

1. Scalability

At the pool level, our designs made full use of space to present aggregated data. Considering the possible over-drawing and visual occlusion problems at the address level, we provided parameter adjustment interfaces and multiple filtering methods. Flexible rendering strategies enabled scalability to explore 10 000-node-level address networks.

2. Target users

This system was mainly for government regulators and researchers. It provided many interesting indicators and rich details. Regulators can use it to monitor, evaluate, and investigate the market and mining pool conditions. Researchers or individuals can explore the history and status of mining pools based on their interests and curiosity.

3. Limitations

When presenting large-scale dynamic networks, the force-directed graph in current systems may look huge and messy. The interaction efficiency of identity reasoning also needs to be improved. In addition, different color schemes for different views cause the system to use too many colors. This is mainly due to the diversity of analysis tasks and data types. Before more consistent color schemes are proposed, users need longer learning time to use the current color scheme.

Further research will focus on two aspects. The first task is to improve graph representation and visualization. We plan to express large-scale dynamic graphs in a vectorized form (Chen W et al., 2019; Wang X et al., 2020). Another useful improvement in the future is to use deep learning (Liu MC et al., 2017, 2018) in de-anonymization and recommendation to reduce repetitive interactions and provide guidance.

Contributors

Jia-zhi XIA designed the research. Jia-zhi XIA and Yu-hong ZHANG drafted the manuscript. Yu-hong ZHANG implemented the system. Hui YE and Guang JIANG processed the data. Ying ZHAO helped design the system. Cong XIE and Ying WANG helped organize the manuscript. Xiao-yan KUI, Sheng-hui LIAO, Wei-ping WANG, and Ying ZHAO revised the manuscript. Jia-zhi XIA and Yu-hong ZHANG finalized the paper.

Compliance with ethics guidelines

Jia-zhi XIA, Yu-Hong ZHANG, Hui YE, Ying WANG, Guang JIANG, Ying ZHAO, Cong XIE, Xiao-yan KUI, Sheng-hui LIAO, and Wei-ping WANG declare that they have no conflict of interest.

References

Aigner W, Miksch S, Schumann H, et al., 2011. Visualization of Time-Oriented Data. Springer, London, UK. https://doi.org/10.1007/978-0-85729-079-3

Athey S, Parashkevov I, Sarukkai V, et al., 2016. Bitcoin Pricing, Adoption, and Usage: Theory and Evidence. Research Papers 3469, Stanford University, San Francisco, USA.

Barkatullah J, Hanke T, 2015. Goldstrike 1: CoinTerra's first-generation cryptocurrency mining processor for Bitcoin. *IEEE Micro*, 35(2):68-76. https://doi.org/10.1109/MM.2015.13

Belotti M, Kirati S, Secci S, 2018. Bitcoin pool-hopping detection. Proc IEEE 4th Int Forum on Research and Technology for Society and Industry, p.1-6. https://doi.org/10.1109/RTSI.2018.8548376

Bistarelli S, Santini F, 2017. Go with the Bitcoin flow, with visual analytics. Proc 12th Int Conf on Availability, Reliability and Security, Article 38.

Böhme R, Christin N, Edelman B, et al., 2015. Bitcoin: economics, technology, and governance. *J Econom Persp*, 29(2):213-238. https://doi.org/10.1257/jep.29.2.213

Bohr J, Bashir M, 2014. Who uses Bitcoin? An exploration of the Bitcoin community. Proc 12th Annual Int Conf on Privacy, Security and Trust, p.94-101. https://doi.org/10.1109/PST.2014.6890928

Chen HD, Chen W, Mei HH, et al., 2014. Visual abstraction and exploration of multi-class scatterplots. *IEEE Trans Vis Comput Graph*, 20(12):1683-1692. https://doi.org/10.1109/TVCG.2014.2346594

Chen SM, Li J, Andrienko G, et al., 2018. Supporting story synthesis: bridging the gap between visual analytics and storytelling. *IEEE Trans Vis Comput Graph*, 14(8):1. https://doi.org/10.1109/TVCG.2018.2889054

Chen W, Lao TY, Xia J, et al., 2016. Gameflow: narrative visualization of NBA basketball games. *IEEE Trans Multim*, 18(11):2247-2256. https://doi.org/10.1109/TMM.2016.2614221

Chen W, Huang ZS, Wu FR, et al., 2018a. Vaud: a visual analysis approach for exploring spatio-temporal urban data. *IEEE Trans Vis Comput Graph*, 24(9):2636-2648. https://doi.org/10.1109/TVCG.2017.2758362

Chen W, Xia J, Wang XM, et al., 2018b. RelationLines: visual reasoning of egocentric relations from heterogeneous urban data. *ACM Trans Intell Syst Technol*, 10(1):2. https://doi.org/10.1145/3200766

Chen W, Guo FZ, Han DM, et al., 2019. Structure-based suggestive exploration: a new approach for effective exploration of large networks. *IEEE Trans Vis Comput Graph*, 25(1):555-565. https://doi.org/10.1109/TVCG.2018.2865139

Di Battista G, Di Donato V, Patrignani M, et al., 2015. Bitconeview: visualization of flows in the Bitcoin transaction graph. Proc IEEE Symp on Visualization for

Cyber Security, p.1-8. https://doi.org/10.1109/VIZSEC.2015.7312773

Fleder M, Kester MS, Pillai S, 2015. Bitcoin transaction graph analysis. https://arxiv.org/abs/1502.01657v1

Gencer AE, Basu S, Eyal I, et al., 2018. Decentralization in Bitcoin and Ethereum networks. Proc 22nd Int Conf on Financial Cryptography and Data Security, p.439-457. https://doi.org/10.1007/978-3-662-58387-6_24

Hoffman P, Grinstein G, Marx K, et al., 1997. DNA visual and analytic data mining. Proc 8th IEEE Visualization Conf, p.437-441. https://doi.org/10.1109/VISUAL.1997.663916

Isenberg P, Kinkeldey C, Fekete JD, 2017. Exploring entity behavior on the Bitcoin blockchain. Université Paris-Saclay, Paris, France.

Jie L, Chen SM, Zhang K, et al., 2019. COPE: interactive exploration of co-occurrence patterns in spatial time series. *IEEE Trans Vis Comput Graph*, 25(8):2554-2567. https://doi.org/10.1109/TVCG.2018.2851227

Kim YB, Kim JG, Kim W, et al., 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS ONE*, 11(8):e0161197. https://doi.org/10.1371/journal.pone.0161197

Kinkeldey C, Fekete JD, Isenberg P, 2017. BitConduite: visualizing and analyzing activity on the Bitcoin network. Eurographics Conf on Visualization, p.3. https://diglib.eg.org:443/handle/10.2312/eurp20171160

Kiran M, Stannett M, 2015. Bitcoin Risk Analysis. NEMODE Policy Paper, p.1-28.

Kirsh D, 2009. Projection, problem space, and anchoring. Proc 31st Cognitive Science Society, p.2310-2315.

Koshy P, Koshy D, McDaniel P, 2014. An analysis of anonymity in Bitcoin using P2P network traffic. Proc 18th Int Conf on Financial Cryptography and Data Security, p.469-485.

Kroll JA, Davey ID, Felten EW, 2013. The economics of Bitcoin mining, or Bitcoin in the presence of adversaries. Proc 12th Workshop on the Economics of Information Security, p.1-21.

Lewenberg Y, Bachrach Y, Sompolinsky Y, et al., 2015. Bitcoin mining pools: a cooperative game theoretic analysis. Proc Int Conf on Autonomous Agents and Multiagent Systems, p.919-927.

Li J, Chen SM, Chen W, et al., 2020. Semantics-space-time cube. a conceptual framework for systematic analysis of texts in space and time. *IEEE Trans Vis Comput Graph*, 26(4):1789-1806. https://doi.org/10.1109/TVCG.2018.2882449

Liu MC, Shi JX, Li Z, et al., 2017. Towards better analysis of deep convolutional neural networks. *IEEE Trans Vis Comput Graph*, 23(1):91-100. https://doi.org/10.1109/TVCG.2016.2598831

Liu MC, Shi JX, Cao KL, et al., 2018. Analyzing the training processes of deep generative models. *IEEE Trans Vis Comput Graph*, 24(1):77-87. https://doi.org/10.1109/TVCG.2017.2744938

Liu SX, Cui WW, Wu YC, et al., 2014. A survey on information visualization: recent advances and challenges. *Visual Comput*, 30(12):1373-1393. https://doi.org/10.1007/s00371-013-0892-3

Liu SX, Andrienko G, Wu YC, et al., 2018. Steering data quality with visual analytics: the complexity challenge. *Vis Inform*, 2(4):191-197. https://doi.org/10.1016/j.visinf.2018.12.001

Liu ZC, Stasko J, Sullivan T, 2009. SellTrend: inter-attribute visual analysis of temporal transaction data. *IEEE Trans Vis Comput Graph*, 15(6):1025-1032. https://doi.org/10.1109/TVCG.2009.180

Luo XN, Yuan Y, Zhang KY, et al., 2019. Enhancing statistical charts: toward better data visualization and analysis. *J Vis*, 22(4):819-832. https://doi.org/10.1007/s12650-019-00569-2

Luu L, Saha R, Parameshwaran I, et al., 2015. On power splitting games in distributed computation: the case of Bitcoin pooled mining. Proc 28th Computer Security Foundations Symp, p.397-411. https://doi.org/10.1109/CSF.2015.34

McGinn D, Birch D, Akroyd D, et al., 2016. Visualizing dynamic Bitcoin transaction patterns. *Big Data*, 4(2):109-119. https://doi.org/10.1089/big.2015.0056

Mei HH, Chen W, Wei YT, et al., 2019. Rsatree: distribution-aware data representation of large-scale tabular datasets for flexible visual query. https://arxiv.org/abs/1908.02005

Meiklejohn S, Orlandi C, 2015. Privacy-enhancing overlays in Bitcoin. Int Conf on Financial Cryptography and Data Security, p.127-141. https://doi.org/10.1007/978-3-662-48051-9_10

Meiklejohn S, Pomarole M, Jordan G, et al., 2013. A fistful of Bitcoins: characterizing payments among men with no names. Proc Conf on Internet Measurement, p.127-140. https://doi.org/10.1145/2504730.2504747

Moore T, Christin N, 2013. Beware the middleman: empirical analysis of Bitcoin-exchange risk. Proc 17th Int Conf on Financial Cryptography and Data Security, p.25-33. https://doi.org/10.1007/978-3-642-39884-1_3

Möser M, Böhme R, Breuker D, 2013. An inquiry into money laundering tools in the Bitcoin ecosystem. Proc APWG eCrime Researchers Summit, p.1-14. https://doi.org/10.1109/eCRS.2013.6805780

Nakamoto S, 2008. Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf

Neudecker T, Hartenstein H, 2017. Could network information facilitate address clustering in Bitcoin? Proc Int Conf on Financial Cryptography and Data Security, p.155-169. https://doi.org/10.1007/978-3-319-70278-0_9

Ober M, Katzenbeisser S, Hamacher K, 2013. Structure and anonymity of the Bitcoin transaction graph. *Fut Int*, 5(2):237-250. https://doi.org/10.3390/fi5020237

Ranshous S, Joslyn CA, Kreyling S, et al., 2017. Exchange pattern mining in the Bitcoin transaction directed hypergraph. Proc Int Conf on Financial Cryptography and Data Security, p.248-263. https://doi.org/10.1007/978-3-319-70278-0_16

Ron D, Shamir A, 2013. Quantitative analysis of the full Bitcoin transaction graph. Proc Int Conf on Financial Cryptography and Data Security, p.248-263. https://doi.org/10.1007/978-3-319-70278-0_16

Spagnuolo M, Maggi F, Zanero S, 2014. Bitiodine: extracting intelligence from the Bitcoin network. Proc 18th Int Conf on Financial Cryptography and Data Security, p.457-468. https://doi.org/10.1007/978-3-662-45472-5_29

Vasek M, Moore T, 2015. There's no free lunch, even using Bitcoin: tracking the popularity and profits of virtual currency scams. Proc 19th Int Conf on Financial Cryptography and Data Security, p.44-61. https://doi.org/10.1007/978-3-662-47854-7_4

Vasek M, Thornton M, Moore T, 2014. Empirical analysis of denial-of-service attacks in the Bitcoin ecosystem. Proc Int Conf on Financial Cryptography and Data Security, p.57-71. https://doi.org/10.1007/978-3-662-44774-1_5

Wang LQ, Liu Y, 2015. Exploring miner evolution in Bitcoin network. Proc 16th Int Conf on Passive and Active Network Measurement, p.290-302. https://doi.org/10.1007/978-3-319-15509-8_22

Wang X, Cui ZW, Jiang L, et al., 2020. WordleNet: a visualization approach for relationship exploration in document collection. *Tsinghua Sci Technol*, 25(3):384-400. https://doi.org/10.26599/TST.2019.9010005

Wang XM, Chou JK, Chen W, et al., 2018. A utility-aware visual approach for anonymizing multi-attribute tabular data. *IEEE Trans Vis Comput Graph*, 24(1):351-360. https://doi.org/10.1109/TVCG.2017.2745139

Wang XM, Chen W, Chou JK, et al., 2019. GraphProtector: a visual interface for employing and assessing multiple privacy preserving graph algorithms. *IEEE Trans Vis Comput Graph*, 25(1):193-203. https://doi.org/10.1109/TVCG.2018.2865021

Wei JS, Shen ZQ, Sundaresan N, et al., 2012. Visual cluster exploration of web clickstream data. Proc IEEE Conf on Visual Analytics Science and Technology, p.3-12. https://doi.org/10.1109/VAST.2012.6400494

Wu YC, Xie X, Wang JC, et al., 2019. ForVizor: visualizing spatio-temporal team formations in soccer. *IEEE Trans Vis Comput Graph*, 25(1):65-75. https://doi.org/10.1109/TVCG.2018.2865041

Xia JZ, Ye FJ, Zhou FF, et al., 2019. Visual identification and extraction of intrinsic axes in high-dimensional data. *IEEE Access*, 7:79565-79578. https://doi.org/10.1109/ACCESS.2019.2922997

Xie C, Chen W, Huang XX, et al., 2014. VAET: a visual analytics approach for e-transactions time-series. *IEEE Trans Vis Comput Graph*, 20(12):1743-1752. https://doi.org/10.1109/TVCG.2014.2346913

Ying Z, Luo XB, Lin XR, et al., 2019. Visual analytics for electromagnetic situation awareness in radio monitoring and management. *IEEE Trans Vis Comput Graph*, 26(1):590-600. https://doi.org/10.1109/TVCG.2019.2934655

Yli-Huumo J, Ko D, Choi S, et al., 2016. Where is current research on blockchain technology?—a systematic review. *PLoS ONE*, 11(10):e0163477. https://doi.org/10.1371/journal.pone.0163477

Yue XW, Shu XH, Zhu XY, et al., 2019. Bitextract: interactive visualization for extracting Bitcoin exchange intelligence. *IEEE Trans Vis Comput Graph*, 25(1):162-171. https://doi.org/10.1109/TVCG.2018.2864814

Zeng W, Fu CW, Arisona SM, et al., 2017. A visual analytics design for studying rhythm patterns from human daily movement data. *Vis Inform*, 1(2):81-91. https://doi.org/10.1016/j.visinf.2017.07.001

Zhao Y, Luo F, Chen MH, et al., 2019. Evaluating multidimensional visualizations for understanding fuzzy clusters. *IEEE Trans Vis Comput Graph*, 25(1):12-21. https://doi.org/10.1109/TVCG.2018.2865020

Zhao Y, Wang L, Li SJ, et al., 2020. A visual analysis approach for understanding durability test data of automotive products. *ACM Trans Intell Syst Technol*, 10(6):1-23. https://doi.org/10.1145/3345640

Zhou FF, Lin XR, Liu C, et al., 2019. A survey of visualization for smart manufacturing. *J Vis*, 22(2):419-435. https://doi.org/10.1007/s12650-018-0530-2

Zhou ZG, Ye ZF, Liu YN, et al., 2017. Visual analytics for spatial clusters of air-quality data. *IEEE Comput Graph Appl*, 37(5):98-105. https://doi.org/10.1109/MCG.2017.3621228

Zhou ZG, Meng LH, Tang C, et al., 2019. Visual abstraction of large scale geospatial origin-destination movement data. *IEEE Trans Vis Comput Graph*, 25(1):43-53. https://doi.org/10.1109/TVCG.2018.2864503

Zhou ZG, Zhang XL, Guo ZY, et al., 2020. Visual abstraction and exploration of large-scale geographical social media data. *Neurocomputing*, 376:244-255. https://doi.org/10.1016/j.neucom.2019.10.072

Zhu MF, Chen W, Xia JZ, et al., 2019. Location2vec: a situation-aware representation for visual exploration of urban locations. *IEEE Trans Intell Transp Syst*, 20(10):3891-3990. https://doi.org/10.1109/TITS.2019.2901117