BitExTract: Interactive Visualization for Extracting Bitcoin Exchange Intelligence

Article *in* IEEE Transactions on Visualization and Computer Graphics · August 2018

DOI: 10.1109/TVCG.2018.2864814

7 authors, including:

Xuanwu Yue
The Hong Kong University of Science and Technology

BitExTract: Interactive Visualization for Extracting Bitcoin Exchange Intelligence

Xuanwu Yue, Xinhuan Shu, Xinyu Zhu, Xinnan Du, Zheqing Yu, Dimitrios Papadopoulos, and Siyuan Liu

Fig. 1. With *BitExTract*, we can observe the evolution of transaction and connection patterns of Bitcoin exchanges from different perspectives A) The comparison view is designed to be highly interactive to compare multiple exchanges' different indices. B) The exchanges list panel reveals Bitcoin exchanges' historical transaction volume. C) The massive sequence view (MSV) demonstrates the overview of Bitcoin exchange market. Users can focus on one exchange to specifically exam its holistic connections. D) The connection view illustrates the connection details intuitively with a node-link design which can facilitate the recognition of unique patterns.

Abstract—The emerging prosperity of cryptocurrencies, such as Bitcoin, has come into the spotlight during the past few years. Cryptocurrency exchanges, which act as the gateway to this world, now play a dominant role in the circulation of Bitcoin. Thus, delving into the analysis of the transaction patterns of exchanges can shed light on the evolution and trends in the Bitcoin market, and participants can gain hints for identifying credible exchanges as well. Not only Bitcoin practitioners but also researchers in the financial domains are interested in the business intelligence behind the curtain. However, the task of multiple exchanges exploration and comparisons has been limited owing to the lack of efficient tools. Previous methods of visualizing Bitcoin data have mainly concentrated on tracking suspicious transaction logs, but it is cumbersome to analyze exchanges and their relationships with existing tools and methods. In this paper, we present *BitExTract*, an interactive visual analytics system, which, to the best of our knowledge, is the first attempt to explore the evolutionary transaction patterns of Bitcoin exchanges from two perspectives, namely, exchange versus exchange and exchange versus client. In particular, *BitExTract* summarizes the evolution of the Bitcoin market by observing the transactions between exchanges over time via a massive sequence view. A node-link diagram with ego-centered views depicts the trading network of exchanges and their temporal transaction distribution. Moreover, *BitExTract* embeds multiple parallel bars on a timeline to examine and compare the evolution patterns of transactions between different exchanges. Three case studies with novel insights demonstrate the effectiveness and usability of our system.

Index Terms—Bitcoin exchange, transaction data, comparative analysis, visual analytics, FinTech

- Xuanwu Yue, Xinhuan Shu, Xinyu Zhu, Xinnan Du and Zheqing Yu are with the Hong Kong University of Science and Technology. E-mail:{xuanwu.yue, xinhuan.shu, xzhuah, xduac, zyuaf} @connect.ust.hk
- Dimitrios Papadopoulos is with the Hong Kong University of Science and Technology. E-mail: dipapado@cse.ust.hk
- Siyuan Liu is with the Penn State University. E-mail: sxl68@psu.edu

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

1 INTRODUCTION

The rocketing Bitcoin value and turbulent market in the past two years have elicited considerable attention from both the finance and technology sectors, making this emerging payment system one of the most compelling topics recently. The booming of Bitcoin has also catalyzed the development of many other cryptocurrencies, including Ethereum, Litecoin [6]. All these alternative coins are built upon variations of blockchain technology, which distinguishes them from the fiat currency system. A blockchain is a public decentralized ledger using a distributed database to verify, propagate, and record digital asset transactions between anonymous addresses in the form of public key hashes. However, for the sake of privacy, public keys generation and usage is usually random, it is difficult to associated the key used in a specific transaction with the key owner in reality. Research and analysis on entity authentication using public keys have revealed the potential to unearth transaction patterns. Various works [21, 39] , thus, have focused on clustering public keys based on external knowledge and heuristic rules into corresponding entities. In our system, we leverage WalletExplorer [10], which has already tagged tens of millions of public keys with their corresponding entities, covering a rich variety of organizations, such as Bitcoin exchanges, Bitcoin mining pools, online gambling sites, cryptocurrency services, etc.

Among the diverse kinds of entities, Bitcoin exchanges distinguish themselves by providing the dominant trading platform where most ordinary customers exchange fiat currency for Bitcoins and vice versa. The thriving interest in Bitcoin has led to the emergence of hundreds of cryptocurrency exchanges of various sizes around the world since it was first introduced by Nakamoto [37]. The overwhelming development of exchanges has promoted the circulation of Bitcoin and brought the growth of its value. A closer look at the Bitcoin blockchain data further reveals the dominance of the exchanges, with a huge number of Bitcoins being transferred between exchanges, thus building up a sophisticated network of transactions. Furthermore, the analysis of Bitcoin transactions between exchanges with the knowledge of exchange trading mechanisms represents the best approach to understand the Bitcoin market. Thus, our intent is to learn the evolution of both individual exchanges and their associated networks so as to understand the development history of Bitcoin exchanges and potential patterns that could be helpful to various interested parties. Moreover, we expect to provide a tool for the public to compare and look for a credible or stable exchange to park their money online.

The critical roles of exchanges have already aroused much research enthusiasm in the area of finance and statistics. Financial researchers focus mostly on the business cooperation and competition between exchanges [22], while the statistics field uses diverse models to study the life cycle and survival probability of exchanges [35]. However, these works are mainly based on empirical hypotheses through a trial-and-error validation process, which often lacks the ability to reveal hidden patterns and relationships. Hence, visualization can be introduced to empower users to utilize their visual senses and intuition to glean insights from the data. Nevertheless, to the best of our knowledge, there is little visual analysis work with respect to the Bitcoin blockchain data. Previous visualization regarding Bitcoin has mainly focused on illegal service discovery [36], anomaly detection [13], and transaction presentation [5]; the core merely concentrates on tracking blockchain transaction logs. However, these visualization techniques are insufficient for exploring the exchange-centered evolution of the Bitcoin market. Despite having the same data source, which is based mainly on the Bitcoin blockchain, the design requirements of our work expect a multi-layer structure for each exchange, containing transaction aggregation instead of raw logs usage directly. Therefore, our work delves into a list of 60 Bitcoin exchanges with millions of transaction records ranging from 2011 to 2018. Our study provides a state-of-the-art visualization system that investigates individual exchanges, analyzes inter-exchange correlation and networks, also allows users to observe the evolving Bitcoin market trend.

It should be noted that developing such a visual analytic system with huge volumes of data presents three major challenges. First, visualizing tens of gigabytes of transaction data involves severe scalability problems. The multi-perspective demonstration of exchange-specific details increases the complexity as well. Second, describing the sophisticated trading networks among exchanges is non-trivial due to a wealth of influencing factors, such as their transaction volumes, trading frequencies, and geographic regions. The dynamic changes of the network should also be further traced. Third, providing a holistic and comprehensive visual design is challenging. The system must take into account people's ability to absorb massive information and the capacity of limited user interfaces to display multi-dimensional data.

To address the aforementioned challenges, we introduce a highly interactive visualization system called *BitExTract*, as it extracts Bitcoin exchange intelligence. The system leverages a massive sequence view (MSV) to depict an overview of the Bitcoin market evolution with all inter-exchange transactions during the 7-year period. Furthermore, the connection view, which is based on a node-link diagram, provides a comprehensive and intuitive demonstration of Bitcoin exchange networks with multiple influential factors token into consideration.Meanwhile, the comparison view embeds multiple parallel bars on a timeline, which allows users to gain insights of individual exchange evolution via interactively tracking the changing momentum compared with others. The well-coordinated system is built on a superimposed hierarchy, which guides users to explore the whole market progressively and interactively. To verify *BitExTract*, we conducted three case studies to evaluate its capability and efficiency. The novel insights into Bitcoin exchanges and further interviews with domain experts have provided promising feedback to our system.

The major contributions of this paper are as follows:

- To the best of our knowledge, the first interactive visualization system which allows users to explore, analyze, and compare the evolution of different Bitcoin exchanges.
- A new dynamic timeline visualization combined with parallel bars that demonstrates time-varying, multi-variable transaction data features for comparison between exchanges.
- Case studies with domain experts and senior practitioners integrate our visualization system into their analysis, leading to the in-depth discovery of and valuable insights into Bitcoin-related events and policies.

2 Related Work

Our work is related to the Bitcoin market and exchange analysis in the financial area, blockchain data visualization, and more generally dynamic graph visualization.

2.1 Bitcoin Market and Exchange Analysis

The thriving interest in Bitcoin transactions has resulted in the emergence of various works in the area. According to Yli-Huumo et al. [50], research related to Bitcoin covers a wide range of topics, such as security, wasted resources, usability, privacy, smart contracts, new cryptocurrencies, botnets, P2P broadcasting protocols, and trustworthiness. From the perspectives of applications, anomaly detection [41], anonymity analysis [43], and Bitcoin price prediction [25] have also drawn great interest in the area of computer science. The majority of previous work focuses on the whole blockchain system without emphasizing specific community. More specifically, this paper concentrates on the analysis of the Bitcoin market and exchanges.

Kiran and Stanett [27] conducted comprehensive research on the risks of the Bitcoin market from social, legal, economic, and security perspectives. Gandal and Halaburda [22] studied the competition among Bitcoin exchanges and claimed that the trading frequency and quantity is related to market news and social community. Other work [16, 45] focused on threats that cryptocurrencies are facing due to the limitation of the intrinsic mechanism of blockchain technology. Moore and Christin [35] employed the survival model to calculate the risk of an exchange going bankrupt based on its transaction volume. While extensive studies on the Bitcoin market have been conducted from the economic and technical perspectives, few studies have been done to empirically analyze the detailed Bitcoin transaction history, since it is difficult to extract inter-exchange transactions. Ranshous et al. [42] made use of transaction data to construct a directed hyper-graph, from which transaction patterns of Bitcoin exchanges can be revealed and leveraged in fraudulent pattern mining, but it mainly aimed at the tasks of characterizing and understanding patterns centered around a specific individual exchange. Thus, to the best of our knowledge, there is no previous work emphasizing the evolution of connections among exchanges and the analysis of event influenced.

2.2 Blockchain Data Visualization

The innate nature of the Bitcoin public ledger has brought an unprecedented opportunity for blockchain visualization. A wealth of websites and tools provides visualization artifacts, including real-time transaction visualization [1, 2, 4, 8] and transaction-network information navigation [5, 7, 9]. However, most of them present only descriptive statistics or transaction details. Recently, visual analytics on blockchain-related data has received increasing attention. BitConeView [13] is a pioneering work inferring illegal activities like money laundering; it utilizes tailored flow charts to illustrate suspicious Bitcoin flows. McGinn et al. [32] deployed high-fidelity visualizations on a large-scale observatory facility, displaying unexpected dynamic transaction patterns. These prior studies are based on block contents to detect abnormal patterns, rarely involving particular users or organizations. Regarding this, heuristic clustering algorithms with graph visualizations [23, 34] have been adapted to analyze the blockchain data for user network characterization. Isenberg et al. extended their previous work on the visual exploration of the activities of entities exchanging Bitcoins [26] and further revealed transaction histories of individual entities [24]. However, these existing works mainly fulfill the classification and presentation tasks, instead of digging into the behavior analysis within the targeted categories, such as exchanges. Moreover, the visualization techniques used are straightforward and unsuitable to tackle the complex problems of revealing connection patterns among exchanges.

In practice, the Bitcoin blockchain consists of linked blocks storing all executed Bitcoin transactions. Visualizations working on transaction data have already arisen in different contexts. SellTrend [31] combines a diverse set of techniques to analyze airline travel purchase requests in real time. Many studies have focused on analyzing pre-collected data. Ko et al. [29] surveyed a series of visual analysis approaches for transaction data. Various types of transaction data, such as wire transfers [18], foreign currency exchanges [30], bids and asks [38] and stock deals [28], have been investigated, resulting in some interesting and insightful visualizations. These studies explore anomaly detection, detail display, pattern mining and violation identification, respectively. Xie et al. [48] also introduced a visualization system called “VAET” to detect salient transactions from large e-transaction time series. In these studies, the buyers and sellers identity were fixed. However, in our case, each exchange can have dual roles, which means each exchange can both buy and sell Bitcoins. A particular design supporting this kind of transaction relationship remains absent.

2.3 Dynamic Graph Visualization

Considering our focus on the evolution of connections between Bitcoin exchanges, dynamic graph visualization techniques should be introduced to address the time-varying process. Two surveys [14, 15] provided a broad perspective of this growing research discipline, a list of visual approaches are proposed which could be categorized into three types, namely, animation, timeline, and a hybrid of the two. Considering the mental map required in the animation techniques, the timeline technique indicated an advantage for the intuitive connection comparison task. In general, this time-to-space mapping consists of node-link-based and matrix-based approaches. Previous dynamic graph visualization techniques mainly focused on the display of entire graph variations. However, the evolution of connections between Bitcoin exchanges gazes at the parts of interest rather than the overview topology, which requires a particular design to highlight the details. Also, the exchange-centric analysis is emphasized. Nevertheless, there is only one application work [47] mentions the financial area but illustrates the simulation of the financial network.

Despite the state-of-the-art methods, most dynamic graph visualizations have a scalability problem: the visualization should change correspondingly according to various data dimensions. For instance, although von Landesberger et al. [47] analyzed a financial network whose dataset was similar to ours, the design proposed for the contagion simulation cannot satisfy our requirements, as it lacked the consideration of other critical variables such as surplus or standing index. Further, there have been several attempts to utilize dynamic multivariate graphs, but the applications were dominated by social networks [17, 49], and software engineering [11]. With respect to the application-specific data characteristics and analytical requirements, it was unsuitable to directly use the existing techniques to analyze the dynamically evolving connections between exchanges, or to display details of multiple data dimensions with millions of records. Thus, we developed an integrated visualization system combining dynamic and multivariate graph techniques to analyze the inter-exchange connections.

Continent	Exchange name
Asia	796.com, Btc38.com, BTCC.com, BtcTrade.com, Bter.com, ChBtc.com, Hashnest.com, Huobi.com, OKCoin.com, Vircurex.com, AnxPro.com, Bitfinex.com, BitVC.com, Exchanging.ir, Korbit.co.kr, UrduBit.com-cold, CoinHako.com, FYBSG.com, MaiCoin.com, BX.in.th
Europe	SimpleCoin.cz, CoinMotion.com, LocalBitcoins.com, Bitcoin.de, C-Cex.com, SpectroCoin.com, TheRockTrading.com, HappyCoins.com, LiteBit.eu, BitBay.net, Bitcurex.com, BTC-e.com, Matbea.com, YoBit.net, BitBargain.co.uk, Bitstamp.net, BitX.co, Bit-x.com, Cex.io, Coinmate.io, Exmo.com, HitBtc.com, Vaultoro.com
North America	Igot.com, Cavirtex.com, Coins-e.com, CoinTrader.net, Bittrex.com, BlockTrades.us, CampBX.com, Cryptsy.com, Kraken.com, Paxful.com, Poloniex.com
South America	Bleutrade.com, FoxBit.com.br, MercadoBitcoin.com.br
Australia	BtcMarkets.net, CoinSpot.com.au, VirWoX.com

Table 1. The Bitcoin exchange companies we studied together with their geographic locations.

3 Background

This section first introduces the background knowledge of blockchain and Bitcoin transaction data. Thereafter, three levels of analytical tasks are proposed and discussed in detail.

3.1 Data Abstraction

3.1.1 Raw data collection and pre-processing

The primal Bitcoin transaction data stored in the blockchain mainly consists of four parts: a list of input public keys, input values, a list of output public keys and output values. Some keys can be recognized as belonging to the same wallet according to generally accepted rules [34]. We began by collecting historical blockchain data maintained by the website WalletExplorer [10] on which every public key has been assigned a wallet ID. Since our focus is on the exchanges, we downloaded the whole transaction history (2011-01-01 to 2017-12-31) of 60 different exchange companies. The companies that we studied are listed in Table 1. The original transactions we obtained from the blockchain might have more than one input wallet and output wallet. This kind of N-to-N transaction structure is not suitable for us to analyze the relationship between two specific entities. Therefore, we broke each transaction into one or more 1-to-1 transactions with the same timestamp, such that each 1-to-1 transaction contained only one input wallet and one output wallet. The transaction value of each 1-to-1 transaction was calculated proportionally. These 1-to-1 transactions are the raw data stored in our database. The following are the fields and descriptions for each record:

- *Transaction ID:* The 64-character hash of the transaction.
- *Exchange ID:* The exchange wallet discriminator provided by WalletExplorer [10]
- *Client ID:* A 16-character wallet discriminator provided by WalletExplorer [10]. It has the same functionality of exchange ID except that we can tell the name of the exchange from its exchange ID, while the client ID is anonymous. When the client is another exchange, the exchange ID will be placed here instead.
- *Time:* The timestamp when the transaction was made.
- *Transaction amount:* The amount of Bitcoin sent by the exchange to the client. It can be a negative number when the exchange is the Bitcoin receiver.
- *Balance:* The amount of Bitcoin that “remained” in the exchange. This field is maintained by WalletExplorer [10].

We also collected some other auxiliary data and information:

- *Historical Bitcoin price:* We retrieved the Bitcoin historical daily prices provided by Coindesk [3]. This is an average value based on daily prices from all Bitcoin exchanges. We converted historical transaction values in Bitcoin to the corresponding values in US dollars.

- *Exchange geographies:* According to the introductory website of each exchange, we manually determined its registered address. This information allows us to explore how the difference in geographic location affects the evolution of an exchange and its relationships with others.
- *News about the Bitcoin exchanges and the Bitcoin market:* We collected numerous online news articles related to exchange activity so that we could link the events to our data, making our tasks more practical, specific and significant. This information allows us to explore how a popular event may have affected the data pattern.

The data mentioned above is considerably large in size: more than 10GB with about 60 million of transactions ranging from January 2011 to December 2017.

3.1.2 Mining for evolution information

We organized the raw data to allow efficient information mining and visualization. We first sorted all the transactions chronologically. Next, we grouped all the transactions according to their exchange IDs. Then we classified the transactions of each group into two sets: exchange-to-exchange transactions and exchange-to-client transactions. The next key step was to define the time granularity to further aggregate the transactions and return summarized statistical information for our tasks. The whole pipeline of our system can be found in Fig. 2.

3.2 Task Analysis

Research into Bitcoin transactions is a growing discipline associated with cryptography, statistics and finance. To better characterize the application domain problems, we have been working closely with four experts during the past six months. In particular, one expert is our internal expert E_A (a co-author of this paper) whose research focuses on Bitcoin, cryptography, and blockchain security. E_A is keen on evolution patterns of the overall exchange network and the impacts caused by significant events. Another internal expert E_B (a co-author of this paper) is a professor in the field of finance, who is curious about the preferences of Bitcoin traders when they try to pick up credible exchanges. Both experts have been actively engaged since the beginning of the work and have shown passionate enthusiasm to integrate visual analytics from the data perspective into their research. The other two external experts, E_C and E_D (not co-authors), are senior participants who have traded Bitcoin since 2013 and are both knowledgeable of Bitcoin exchanges, transaction rules and major milestones of Bitcoin history.

Therefore, our work is built on close cooperation with these four experts and follows a typical user-centered design framework, including discussions, brainstorming, designing, prototyping and presenting. After several iterations, we collected their feedback and condensed it into a set of six primary questions for task analysis that are further classified into three categories as follows.

Two **Overall-Market-level** questions give a full picture of the Bitcoin market development that is dominated by exchanges.

T.1 *How do the patterns of the overall exchange network evolve over time?* The prosperity of exchanges is essential for the circulation of Bitcoin. Rising and falling periods are both crucial for analysis. In different years, what does an exchange's transaction network look like from both global and regional perspectives?

T.2 *What impact does a specific event exert on the overall exchange network evolution?* Owing to the closely-connected network of exchanges, domestic events or policies may have a continuous effect on the whole Bitcoin market. Thus, even if regionally-imposed policies only exert direct pressure on a small number of exchanges, they may also affect other exchanges which possess strong relationships with them. Powerful ripples may be sent across the entire market.

Two **Inter-Exchange-level** questions emphasize the evolution of connections among a set of exchanges.

T.3 *How do exchange connections evolve over time?* Does a particular exchange have a stable connection with others? Or do its connections with others change a lot? If an exchange has strong ties with others, does it switch its ties at some points due to a specific reason? Do exchanges develop relationships with geographically nearby exchanges first and then with geographically distant exchanges?

T.4 *How is the change in "surplus" of an exchange related with its relationship with other exchanges?* The "surplus" of an exchange during a period is the sum of Bitcoin value that exchange received from its clients minus the Bitcoin value sent by that exchange to its clients in the same time period. The surplus is computed by

$$\sum_t Receive_i - \sum_t Send_i \tag{1}$$

for exchange i in a time section t. Will an exchange that gain adequate Bitcoin from its clients play an active role in supporting other exchanges thus influence its connection?

Two **Individual-level** questions focus on a detailed picture of a single exchange over time.

T.5 *How does the performance of an individual exchange evolve over time?* Transaction patterns of an exchange vary from time to time due to its business operations or some specific events. When are these events most likely to affect its performance significantly?

T.6 *What are the periods of importance of each exchange?* Each exchange may possess some unique patterns in both its performance development and connection evolution. These patterns are usually associated with special events or trading strategies of that exchange. Quickly identifying those periods of importance of each exchange presents a question.

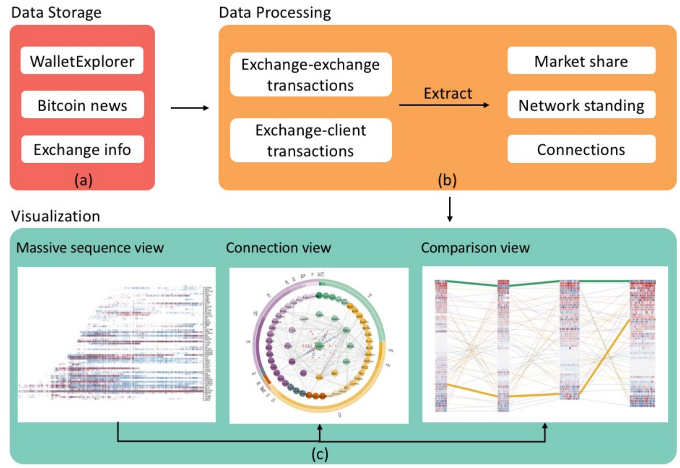

Fig. 2. System Overview. *BitExTract* consists of three major parts: a) Data storage module; b) Data processing module; c) Visualization module.

4 System Overview

BitExTract is a web-based application with three major parts, namely, a data storage module, a data processing module and a visualization module as shown in Fig. 2. The data storage module is based on MongoDB and collects necessary information, such as transactions of Bitcoin exchanges, significant news about Bitcoin exchanges and exchange company information from online websites. The data processing Python3 module then groups transactions into two categories exchange-to-exchange transactions and exchange-to-client transactions, and performs time aggregation on the raw data in the database. The processed data is further utilized to extract high-level statistics and other indices, such as the market share, the network standing index and the connectivity between two exchanges. These two Python-based modules together form the backend and further guarantee the efficiency of our visualization module.

The visualization module empowers users to explore the evolutionary transaction patterns of Bitcoin exchanges with three well-coordinated views. The massive sequence view summarizes the evolution of the Bitcoin market over time. The connection view depicts the trading network of exchanges and their temporal transaction distribution. The comparison view compares the evolution trends of different exchanges, taking network standings into consideration. Besides, an additional panel listing all exchanges is provided to help users get an overview and quickly select a certain exchange.

We describe a common workflow to show how a user utilizes our system. The user can first selects exchanges of interest on the Exchanges List (Fig. 1*B*) on the right side of the interface. Different sorting methods could provide hints for users to select interested exchanges with particular patterns. Three other views then update according to the selected exchanges. When focusing on an exchange, the massive sequence view (Fig. 1*C*) shows inter-exchange transactions involving such exchange. The user can further brush an interested time period on the price panel (Fig. 1C_3). Then the news panel (Fig. 1C_1) displays related events during this period. He can also explore the trading network of selected exchange on the connection view (Fig. 1*D*) and compare evolution trends, surplus situation, different network standing combination on the comparison view (Fig. 1*A*).

5 Data Analysis

In this section, we will introduce our analytic approach to organize tremendous trading data, and we will provide linear ordering and business proximity to the exchanges through an interactive user interface. First, we will describe the adaptive ranking algorithm, which reflects the market share and network standing of exchanges to some extent. Then we will give the definition of business proximity in the context of Bitcoin exchanges.

Notation	Meaning
$Stand_i^t$	Network standing of an exchange company i in time period t
I_i^t	Inverse volatility of exchange i in time period t
E	Representing all the exchange companies examined in this paper
$Share_i^t$	Market share, *i.e.thetradingvolume* of exchange i with other exchanges during time period t
$V_t^{p,i}$	Trading volume between exchange i and p during time section t
$P_t(a,b)$	Business proximity between exchange a and b at the end of time period t
$V_t(a,b)$	Trading volume between exchange a and b at the end of time period t
$F_t(a,b)$	Trading frequency between exchange a and b in time period t

Table 2. Terminology table used in financial analysis for Bitcoin exchanges.

5.1 Adaptive Network Standing Index

A weighted average is widely used in scientific research in order to determine the relative importance of several potential factors [46]. Therefore, we apply this general rule to the context of quantifying Bitcoin exchanges. We combine three attributes suggested by Bitcoin and financial domain experts, and define a *network standing index* of an exchange platform. We define our model as follows. Let $Stand_t^i$ be the network standing of an exchange company i in time period t, which is a function of the balance volatility, its market share during that time section, and also weighted network standing of exchange companies which have incurred transactions with it in the previous time section.

$$Stand_t^i = \alpha \cdot I_t^i + \beta \cdot Share_t^i + \gamma \cdot \frac{\sum_{p \in E}(V_{t-1}^{p,i} \cdot Stand_{t-1}^p)}{\sum_{p \in E} V_{t-1}^{p,i}} \quad (2)$$

In the financial domain, volatility measures are usually described by the standard deviation of change in price or value of a financial security [20]. In this context, we expect an exchange company to send Bitcoins to its clients within its own capability without a volatile record of balance. Therefore, in order to model the potential risk of investing in an exchange company, we should take into account its fluctuation of remaining Bitcoin balance. Moreover, though the distribution of the transaction data is symmetric bell-shaped, it is yet highly skewed. Therefore, we apply the widely used log-normalization on the volatility of balance: $V = log(\frac{balance_t}{balance_{t-1}})$, and pass the inverse of volatility through a Logistic function to make it fit into $[0,1]$ range nicely. The result of this term is then: $I_t^i = \frac{1}{1+e^{1-std(V)}}$

The market share of exchange i is scaled by an activation function so as to nicely fit into the range between zero and one. That is, $Share_t = 2 \cdot (1 - logistic(-\frac{1}{3} \cdot log_{10}(\frac{V_t^i}{V_t})))$

The third term is inspired by the idea of PageRank [40], which propagates the influence of one node to its adjacency nodes so as to measure its relative importance within the set. Each partner p of exchange i will exert influence weighted by $\frac{V_{t-1}^{p,i}}{\sum_{a \in E} V_{t-1}^{a,i}}$, which is p's fraction of transaction volume with i.

α, β, γ are adaptive factors which control the three features' relative importance in terms of network standing. Users can adjust them via comparison view "Network Standing Parameters" sector in order to evaluate different ranking results and dominant factor.

5.2 Business Proximity

Business proximity is an active research area in Information System [44] which measures relatedness of companies in terms of geographical distribution, market, technology, etc. We tailor this idea to measuring the business connection of exchanges as a linear combination of their transaction volume and trading frequency. The formal definition is as follows

$$P_t(a,b) = \alpha \cdot V_t(a,b) + \beta \cdot F_t(a,b) \quad (3)$$

where $V_t(a,b)$ represents trading volume between exchange a and b, and $F_t(a,b)$ represents their inner-transaction frequency. Currently, α and β are set with equal weights since we regard the two factors are both important for proximity analysis.

Business proximity is effective when it comes to understand the market connection among exchanges. Users can easily identify the strongest partner of a specific exchange in the ego-view from Connection View where business proximity serves as the distance metric. The detailed visual encoding will be discussed in Section 6.3.

6 Visual Design

Our visualization system consists of four components, namely comparison view Fig. 1*A*, exchanges list Fig. 1*B*, massive sequence view Fig. 1*C*, and connection view Fig. 1*D*.

We followed the design rationales below to guide the process.

Follow user-centric design. In order to aid domain experts with visual analysis, one of the design goals was providing multi-stage problem-solving views which involve end users from the beginning. Considering the demands and learning-curve of first-time users, our system should be highly interactive and the task flow should be intuitive.

Overview first, zoom and filter, details on demand. Given the massive scale of the temporal transaction data of Bitcoin exchanges, the system should provide sufficient interactions so that users can zoom-in and filter the data on demand. According to the theory of *eyes beat memory* [12], it is easier to compare views side-by-side than the visible item to memory. Therefore, we should fully utilize the limited pixels on the screen so that users would not necessarily rely on a mental map to perceive the comprehensive information and make comparisons.

Quantifying network standings of exchanges. Quantifying network standing of exchange provides hints for Bitcoin investors to seek credible or stable exchanges. Domain experts emphasize the importance of modifiable weights on factors of exchange evaluation. Thus, one of the design goals was providing direct interactions for adaptive ranking parameters with intuitive visual feedback.

We use the following unified color encoding to represent the continents where exchange companies are registered, *i.e.*, yellow for Asia, purple for Europe, green for North America, orange for South America, and blue for Australia.

6.1 Exchanges List Panel

We provide an exchanges list panel (Fig. 1*B*) for users to quickly select a certain exchange and observe its historical transaction volume in USD. Each card in the panel is essentially a bar chart encoding the exchange's transaction volume with respect to time. Users can choose to sort the exchanges according to their entry-time, or continents. When the entry-time sorting criteria is selected, the newest incomers will be stacked on the top, while the most aged players will appear at the bottom.

6.2 Massive Sequence View

Description: The massive sequence view (MSV) (Fig. 1*C*) provides a compact and comprehensive overview of the temporal transaction patterns of all the examined exchanges, which further reflects the whole market evolution of Bitcoin. Based on this view, we can answer the analytical questions about the overall market (T1, T2) as well as inter-exchange relationships (T3).

As Fig. 1C_2 illustrates, the x-axis represents the timeline, and the exchanges are stacked along the vertical axis. Each row represents an individual exchange. We order all exchanges chronologically along the y-axis. That is, the "older" exchange lies in the bottom and the newly emerging ones would be laid on top. In particular, the contour of MSV exhibits the development speed of the market. For example, the S-shaped contour in Fig. 1C_4 presents a concave curve and then turns into a convex curve, showing a great number of exchanges had emerged during that around 2013 to 2015 with concave shape but the acceleration stagnated afterwards.

The diverging red-blue color represents the surplus of an exchange as defined in Eq.1. That is, this exchange i with the larger gap of sending amount minus the receiving one is shown in a reddish color during the time section t. Deviation above and below the zero-surplus is well represented by the diverging color scheme.

This view also supports analysis on two types of transaction data, *i.e.*, the transactions between exchanges and clients ("Surplus"), and the aggregated transaction behavior between exchanges ("Inter-exchange"). Users can switch between these modes from the selection bar above. When a certain exchange row is clicked, inter-exchange transactions involving that particular exchange will be highlighted with other irrelevant records filtered out.

The price panel (Fig. 1C_3) shares the timeline with the MSV (Fig. 1C_2), which shows historical prices of Bitcoin in USD in blue background. It also embeds a time brush function which can trigger interaction in multiple other views. The news panel (Fig. 1C_1) on the top-left will display major Bitcoin-related events corresponding to the selected time period. When the cursor is hovered on a certain news, a blue vertical line will appear on the price panel to help you align news with patterns in MSV as well as Bitcoin price.

Justification: We discussed several candidate designs, such as bar chart or line chart to encode the time-varying transaction amount of all exchange companies. However, we found that these approaches would induce visual clutter, given the massive size of the datasets, thus impede the visual analytics process. Moreover, the MSV provides one more dimension of encoding compared to the bar chart and the line chart, from which we can encode the chronological order of entry-time of exchanges by stacking them along the y-axis. Moreover, we also considered using Wordle [33] for the text representation of Bitcoin news. However, we found it is hard to extract meaningful tokens for reasoning, which is still an active research domain in Natural Language Processing. Currently, we adopt a simple design which relies on Google News. Other techniques can be developed further to summarize the financial text data so as to relieve the scalability pressure.

6.3 Connection View

The connection view, which is essentially a graph, aims at visualizing inter-exchange behavior with geographic information(T3, T6), helping users better understand the exchange relationships, and demonstrating the transaction patterns of an ego-exchange with respect to its partners (T2). Note that the "ego" in social network analysis means a specific individual for detailed investigation.

Description: The connection view (Fig. 1*D*) is built upon the design of node-link diagram and extended with an ego-view layout where users can drag any exchange to the center for detailed exploration. With each node representing an exchange, it is colored with respect to different continents, and its opacity varies to distinguish countries inside the same continent. The link between two nodes shows their connection as is described in Section 5.2. Thicker links denote stronger connections within the selected time period. There are five portions of arcs inside the outermost ring, whose lengths represent the market share of this continent. Once an arc is selected, the diagram will enter the "continent view", which only shows the inner-continent transactions.

Apart from the "world view" and the "continent view" described above, we also support "ego view" for analyzing a specific exchange in this diagram. By dragging a node (*i.e.* ego) into the center of the circle, we arrange the exchanges who trades with the ego during selected time range and those who does not into two concentric rings. Arrows along the path, which is essentially the timeline, that link the ego and its partners encode the trading volume. For example, an exchange having larger volume transactions only at the beginning of the selected time period will have long and dense arrows closer to the center while short and sparse arrows further from the center. The highly-summarized layout provides visual patterns for users to compare and analyze across different time periods. One can always go back to the general inter-exchange view by clicking on the ego node.

Justification: We considered alternative visual designs for illustrating the inter-exchange connection. One option was to use matrix-based design [19]. However, this design emphasize on the hierarchical group structure, which may not be necessary given the shallow hierarchy of our dataset and it may induce steeper learning curve of the users. Our design based on the node-link diagram could better utilize limited pixel space and is extendable to ego-view for further analysis. Specifically, if users find an exchange interesting and drag it to the center, the ego-view appears to illustrate its connections with other exchanges, as well as their detailed transaction patterns from the beginning of the selected period to the end via path-with-arrow encoding, which is hard to achieve through matrix design.

6.4 Comparison View

We draw upon the design of parallel coordinate view for reference to encode the overall evolution pattern of Bitcoin exchanges with their clients. While parallel coordinate targets at encoding high-dimensional data, our comparison view leverages the similar vertically-aligned bars to encode time-varying transaction patterns. Generally speaking, this view first provides users a handy comparison of exchange transaction patterns, and second, it provides a zoom-and-expand feature which supports more detailed visual analysis on a specific exchange (T5, T6). Moreover, through this view users can quickly identify the big exchanges in the Bitcoin market (T5). The visual encoding is as follows.

Description: As shown in Fig. 1*A*, each vertical bar in the comparison view stands for the exchange transaction history in one month. The width of a bar stands for the aggregated client number of all the exchanges in that month. Horizontal rows in one bar represent exchange companies sorted by certain rules. Vertical lines in each row record a summarization of daily transactions of that company.

We support multi-ranking (*i.e.*, network standing, surplus and transaction volume) and two different data filling (*i.e.*, surplus (Eq. 1) and transaction volume) options. This flexible design is intended to help users analyze the influence of the filling index on the ranking index. If sorted by the rule of "surplus", the companies on the top will have a red color since they "send" more than they "receive", while the companies on the bottom will have a blue color. While the "surplus" filling-scheme follows divergent color encoding, the "transaction volume" filling follows sequential color scheme. When network standing ranking-scheme is selected, users can customize three factors to the linear combination of the input, as shown in Section 5.1, by clicking on a point representing α, β, γ in the triangle. The closer distance to a triangle vertex, the more weighted this factor is.

The same companies between consecutive time steps are linked by lines, thus it is easier for users to trace a particular exchange throughout history and observe the change in the rankings. Multiple exchanges can be selected and highlighted simultaneously for easy comparison (Fig. 4). Users can easily grasp the temporal evolution pattern of all exchanges and apply customized rules to re-sort the exchanges through this comparison view, which could compare by different measures.

According to the empirical theory of visualization, it is hard to get an accurate value estimation on the opacity channel. Therefore, we provide an interaction from which the horizontal row inside bars will be

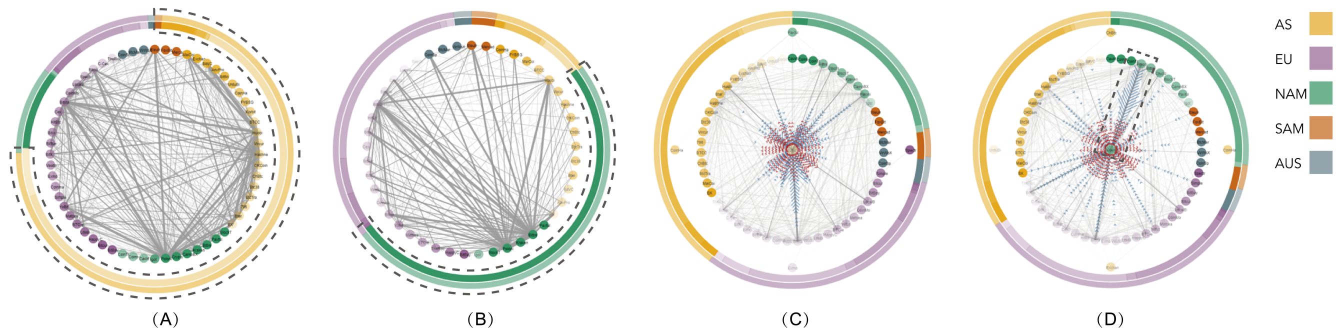

Fig. 3. Exchange network patterns on connection view. A) Summarized inter-exchange transaction from Nov, 2015 to March, 2016. B) Summarized inter-exchange transaction from May 2017 to December 2017. C) Bitfinex gradually reduced its direct connections with its partners. D) Kraken maintained some direct connections with its partners. Both Bitfinex (C) and Kraken(D) stopped sending Bitcoins to their direct partners after integrated with BitGo service as the red arrow totally disappeared.

expanded when users double-click a company in a bar, and transaction amount on each day will be projected onto the length of the expanded line. While the rows encode every vertical line with equivalent length but different opacity, the expanded line encodes the everyday send-versus-receive amount in its left-side month with various length, which enables users to better perceive the quantity through the length channel.

Justification: We considered some standard charts, for instance, line chart, before we adopted the current design. These alternative graphs are usually too dense and difficult to comprehend. Meanwhile, our comparison view can encode information on both vertical bars and horizontal lines, which can provide a compact high-level summary of temporal patterns. Aggregating every month's information on the vertical bars and zooming in on demand becomes necessary to give the massive data volume in this context.

6.5 Cross-view Interaction

BitExTract supports various interactions, empowering users with strong visual analytic abilities. The detailed explanation is as follows.

Temporal-related Interaction. In order to better understand the temporal patterns, our system will by default show the data spanning from 2011 to 2018. However, users can always zoom in to an interesting period on demand. We provide cross-view interaction from the MSV to the comparison view and the connection view. By brushing the time axis from the MSV, the data in the comparison view and the connection view will be filtered correspondingly. This will further facilitate the exploration of the relationship between the tendency on the comparison view and the patterns in the connection view. For example, we can study how the change in surplus of an exchange is related with its connections with other exchanges (T4).

Exchange-centered Interaction. The panel serves as the main controller for selecting a specific exchange for analysis. Clicking on a card on the panel will trigger the MSV to display transactions with regard to that exchange, the comparison view to highlight the corresponding line, and the connection view to display the ego-view. Such interaction will facilitate exchange-centered exploration by displaying visual patterns from multiple views at the same time.

7 Evaluation

This section presents three case studies. The ultimate goal of *BitExTract* is to provide users with an efficient and effective tool to capture and analyze the movements of different scales in the Bitcoin exchange network. To test and evaluate the performance, we conducted three case studies with our domain experts, Bitcoin traders and researchers.

7.1 Detect and compare special periods

Since many people are interested in the development history of the Bitcoin market, the expert E_A planned to review and summarize the key time periods during the history (T1) He was looking forward to dating active market periods, describing the market situation during these periods and getting hints about causes of these activeness.

In this scenario, the expert E_A used *BitExTract* to go through the development of the exchange market, being interested in identifying certain periods where major shifts in the exchange market took place. He began by looking through the comparison view from left to right and noticed that there were two special periods in which the width of the pillar increased significantly, which indicated an increase in the Bitcoin exchanges' activity. As can be seen in Fig. 4, the Bitcoin exchange market showed an increase in activity during the periods of November 2015 to March 2016 and May 2017 to December 2017 (T1). The increase of the bars' width shows an increasing amount of total clients in the market.

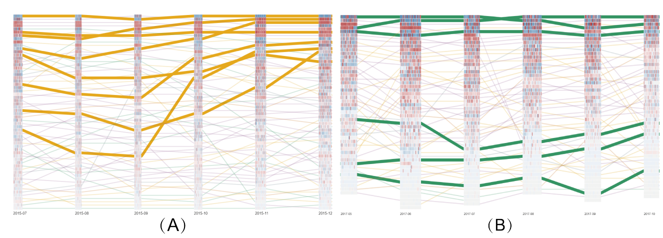

Fig. 4. Comparison views ranked by transaction amount and filled by daily surplus. A) The Chinese exchanges' uprising tendency together with the increasingly active Bitcoin market from July, 2015 to December, 2015. B) The American exchanges' polarized tendency together with the flourishing Bitcoin market from May, 2017 to October, 2017.

In order to identify the reasons behind these changes, E_A then delved into a further comparison between these two periods. He first slid the time window to the first one. The connection view showed the overall market share comparison as well as the networking picture between all the exchanges, as shown in Fig. 3*A*. From this, he was able to notice that exchanges geographically located in Asia (and in particular in China) accounted for more than half of the market share for this period (T1). Aided by this finding from the connection view, E_A moved back to the comparison view and filtered exchanges from China, as in Fig. 4. He found that many Chinese exchanges significantly rose up in terms of transaction amount during these months (T5). Therefore, E_A concluded that the flourishing period between 2015-11 and 2016-03 was mainly led by Chinese exchanges. This is in line with the market behavior during that period, and it was the result of certain market policies in China, in combination with the low economic growth in the greater Asia area, as validated by relevant news articles from the period [1,2].

For the second period, which was between May 2017 and December 2017, E_A followed the same exploration path and the diagrams from Fig. 3*B* and Fig. 4*B*. According to Fig. 3*B*, E_A found that USA-based exchanges account for most of the transactions, despite the fact that the number of such exchanges is relatively small compared to the number of exchanges based in Asia or Europe. For this period, Asian exchanges become less important (T1). Subsequently, in the comparison view E_A filtered out the tendency of American exchanges. He found that this

[1]https://bit.ly/2KceI67
[2]https://bit.ly/2KbzUG1

view included a polarization. In particular, three of the American-based exchanges took the top positions on the transaction amount ranking list while others laid well below the average (T5). Therefore, E_A concluded that this second flourishing period was led by a relatively small number of American exchanges. Again, this finding is supported by news articles from that time period [3,4].

7.2 The impact of a policy event

One of our expert co-authors, E_B, who is a financial researcher interested in Bitcoin risk analysis, mentioned the effect that the introduction of a new policy can have on the Bitcoin market. One of the most influential such events was the Chinese government's decision to prohibit Yuan-to-Bitcoin trading in September 2017 [5]. In the period prior to the release of the official announcement by the Chinese government, various rumors had been circulating, impacting the Bitcoin market. We recorded how E_B, E_C and E_D attempted a comprehensive comparison for this time period with the help of *BitExTract*.

In terms of the capitalization of the 2016-2017 cryptocurrency market, the China-based Huobi exchange was definitely the largest Bitcoin trading platform, both domestically and globally. In February 2017, the People's Bank of China (the Chinese central bank), inspected Huobi and OKCoin [6]. These two exchanges were two of Chinas most widely-used ones, and they both announced their intent to suspend Bitcoin and Litecoin withdrawals effective immediately after the inspection. After this, China's central bank issued a warning to domestic exchanges, going so far as to state it would move to shutter startups that violated its guidance through the necessary government channels. Around four months later, Huobi announced that they would resume Bitcoin withdrawal services[7]. This policy and this event changed the Bitcoin exchanges market globally to a significant extent.

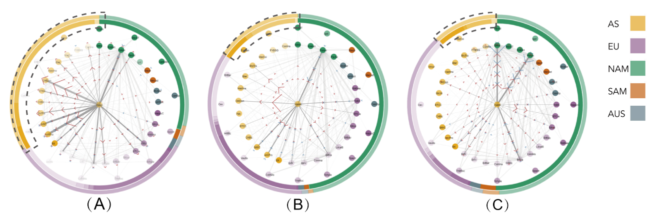

Fig. 5. Huobi's network evolution: A) Before the inspection of Peoples Bank of China, Huobi's connections with Chinese partners were very strong. Asian market share was comparable with the European and the North American. B) Huobi stopped withdrawal services and its connection with partners shrank immediately. Asian market share descended significantly. C) Huobi resumed withdrawal services and restored some connections with partners in Europe and North America, its connection with Chinese partners remained weak. Asian market share didn't recover.

E_B began exploring this period by using the massive sequence view. It is quite obvious that the selected Huobi exchange demonstrates a blank gap from February to June of 2017, which exactly matches the time period during which Huobi suspended transactions (Fig. 1C_3) (T6). Prior to this period, the exchange used to maintain close and dense connections with other exchanges worldwide, as can be inferred from the connection view in Fig. 5. Strong connections were established with most of the Asia-based exchanges, as well as some popular exchanges in North America and Europe (T3). Moreover, from Fig. 5*A*, it is also clear that during this period Asia, Europe, and North America took approximately one-third of the market share each. Meanwhile, Huobi mainly sent out Bitcoins to other exchanges and received from clients, which could be revealed by comparasion view in the surplus ranking mode (T4). During the examined period, connections between Huobi and other exchanges shrank significantly. This change was reflected especially in the Asian area. Note that dominance in the Bitcoin market transferred mostly into North America, followed by Europe (Fig. 5*B*) (T2). E_C mentioned that North American exchanges seized this timing point to attract Asian customers. Even though Huobi eventually recovered some connections with foreign exchanges, the connections with Asia-based ones remained unchanged. Also, the North American exchanges maintained the worldwide top place (Fig. 5*C*). E_D then explained regarding this situation that Huobi was the central point of the Chinese market and, to some extent, even the entire Asian region. Once Huobi suspended its services, the other Chinese exchanges failed to establish relations with other (internal or external) exchanges, possibly due to the regulations imposed by the Chinese government. Beyond these insights gained by the connection view, our experts also noticed an extreme drop in the "Network Standing" and "Transaction Amount" mode in comparison view (Fig. 1*A*).

7.3 The different effects of BitGo adoption on Kraken and Bitfinex

How to choose a reliable exchange for Bitcoin investment is always a huge concern almost for each Bitcoin trader. Many people discuss with each other via online platforms. For instance, there is a discussion raised by Quora users to explore the reliability of three big exchanges: Coinbase, Kraken and Bitfinex[8] (T5). However, online information is sometimes biased. For example, some people mentioned that Kraken and Bitfinex were good and safe because they had lots of currencies available and they had all the information stored on the offline resources. These kinds of information are ubiquitous on the Internet, but they are apparently too general and superficial. How much currencies can be called lots of and why storing information on the offline is safer? We consulted expert E_A on the exchanges comparison problem, like Bitfinex and Kraken. E_A pointed out that these two exchanges were similar to each other because they both adopted the BitGo custodian service to accelerate their customer transactions. Both of them kept part of their Bitcoin in the BitGo wallet which could process the transactions faster. However, E_A had little idea about which exchange was more reliable because he didn't know the detailed strategies used by these two. Therefore, we invited him to make a comparison via *BitExTract*.

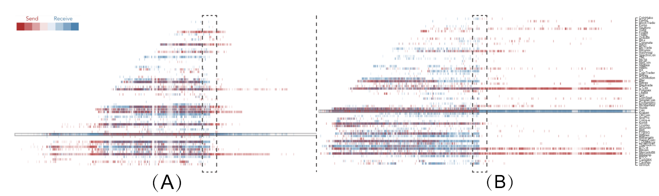

Fig. 6. Bitfinex's and Kraken's historical connections with other exchanges. A) In Jun, 2015, Bitfinex stopped sending out Bitcoins to its partners and in Aug, 2016 Bitfinex terminated all its connections with old partners. B) In November, 2015, Kraken stopped sending out Bitcoins to its partners.

Overview comparison using the massive sequence view. Starting from the massive sequence view, E_A highlighted Bitfinex and Kraken in order to see an overview of the historical connection between them and other exchanges. The resulting patterns can be seen in Fig. 6*A* and Fig. 6*B*. E_A was immediately attracted by the turning points in the diagrams. The first one was at the beginning of June 2015. Originally, Bitfinex had numerous connections with many other exchanges, such as Bitstamp, Poloniex, Huobi, and Kraken. As he observed both red slots and blue slots on these bars, he concluded that Bitfinex both sent Bitcoins to and received Bitcoins from these partners. Immediately after June 2015, the bar corresponding to Bitfinex contains only blue slots, and all other bars are either interrupted or become totally red. In other words, during this period Bitfinex stopped sending Bitcoins to its partners, and it terminated a number of its direct connections to some of its old partners. Yet, it still received Bitcoins directly from some exchanges. However, after August 2016 Bitfinex ended all its

[3] https://bit.ly/2lvNb1G
[4] https://bit.ly/2KhEaqS
[5] http://www.bbc.com/news/business-41320568
[6] https://news.bitcoin.com/happened-to-bitcoin-in-china-2017/
[7] https://news.bitcoin.com/chinese-bitcoin-exchanges-resume-withdrawals/
[8] https://bit.ly/2yAW5Eu

connections with other exchanges. Turning his attention to Kraken, E_A observed an interesting turning point around November 2015. Similar to Bitfinex, Kraken also reduced its connections with other exchanges and no longer sent out Bitcoin. It just received Bitcoins from its partners after that time. E_A swept over these time periods to see what happened to these exchanges. With the help of our real-time news plugin, E_A learned that precisely on June 4th, 2016 Bitfinex started using the BitGo service [9], and on November 10th, 2016 Kraken also integrated with BitGo for multi-sig security [10]. Then on August 2nd, 2016 Bitfinex reported being hacked in an attack that resulted in approximately 70 millions US dollars worth of Bitcoins being stolen [11], and that may have been partially related to its BitGo implementation [12] (T6).

Networking evolution comparison using the connection view. By focusing on the connection views during different time periods, E_A found additional details about these periods as in Fig. 3*C* and Fig. 3*D*. After integrating the BitGo service, the connections of Bitfinex and Kraken with other exchanges grew "weaker", as they no longer sent Bitcoins directly to any of their partners but used BitGo's wallet as an intermediary. The network of Bitfinex kept dropping continuously which made its direct partners fewer and fewer, while Kraken (following a more conservative strategy regarding BitGo) still maintains a number of direct partners (T3).

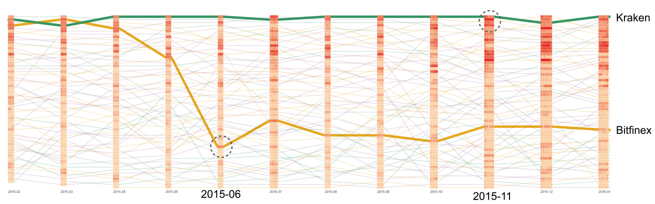

Fig. 7. Bitfinex's and Kraken's network standing. Bitfinex's network standing drops significantly after integrating with BitGo service in Jun, 2015. Kraken's network standing remained high after integrating with BitGo service in Nov, 2015.

Network standing comparison using the comparison view. Switching to the comparison view (Fig. 7), E_A attempted to infer the tendency of the network standing of Bitfinex and Kraken. Around the first timepoint, he observed that the network standing of Bitfinex shows a steep decrease (T5). This indicates that Bitfinex's wallet was at higher risk after the adoption of the BitGo wallet service. By adjusting the weighting scheme, he found that this drop was mainly caused by the increase in balance volatility, since the Market share and connection with its partners did not drop significantly. Bitfinex was still using its old wallet to serve its customers. Apart from that, it first sent Bitcoins to its BitGo wallet and then to its customers so its market share remained roughly the same. The only difference was that Bitfinex maintained only a small amount of Bitcoin in its old wallet, which made its wallet volatility quite high, thus affecting its network standing. On the other hand, Kraken kept most of its Bitcoin in its own wallet and transferred only a necessary amount to the BitGo hot wallet and therefore, its network standing remained virtually unaffected.

8 Discussion

Three case studies demonstrate the effectiveness and usability of our system in extracting intelligence of Bitcoin exchanges. Overall, our collaborating experts are satisfied with its strong analytical abilities, especially with straightforward visualization. The massive sequence view presents a whole picture of the Bitcoin world based on transactions between exchanges and clients. Supplementary information is provided by the news panel. The connection view displays the business proximity with a highly interactive design. Smooth interactions with prompt visual feedback allow users to explore from the worldwide situation to the inner-continent connections. The individual exchange is further displayed in the comparison view considering different attributes. For now, the order of each bar only reflects exchanges' relative status instead of the absolute values. When compared with standard data analysis software, our system has already processed raw transaction data and provided a comprehensive analysis of the whole market and individual exchange with well-coordinated views. To the best of our knowledge, *BitExTract* is the first visualization system analyzing Bitcoin exchanges, which closely follows domain tasks and requirements. Both domain experts and general users with basic knowledge about Bitcoin could step into this world via *BitExTract*.

However, it still faces several limitations. Firstly, scalability is the major concern when designing *BitExTract*. For example, we show approximately a maximum of 10 months once in the comparison view. Although scrolling in this view enables users to explore more than these, interactions still increase users' memory burden. A possible improvement could be aggregating transactions based on different time granularities according to the length of the selected time period. The massive sequence view currently gives a clear demonstration of the evolution of 60 exchanges. However, this view might not provide such good performance when the number of exchanges is much larger than this. Displaying dominant exchanges could be an acceptable trade-off, which still indicates the overall patterns of the whole market. Besides, there might be severe visual clutters in the connection view in some extreme cases such as showing transactions of the main exchange during a long period, while it can be solved by grouping these arrows. Another limitation lies in data uncertainty. Due to the anonymity properties of the blockchain, identifying all the public keys belonging to a specific exchange with 100% accuracy seems impossible, despite the use of the state-of-the-art heuristic algorithms for classification. Although this factor introduces uncertainty to data, we believe the exchanges examined in our system are a representative subset of the current Bitcoin market. It is further supported by our findings in Section 7. Thirdly, three attributes when defining the network standing index are derived from our experts' and senior Bitcoin traders' empirical knowledge. It might be partially incomplete and subjective.

Besides, the current *BitExTract* system can be extended to many promising aspects. First, most exchanges can support multiple cryptocurrencies trading. We can further apply our visualization approach to other cryptocurrency analyses and dig into exchanges' business intelligence. Secondly, the design in the comparison view could be extended. Owing to the characteristics of financial attributes, multiple sortable parallel bars which encode alternative attributes on a timeline are suitable for other time-varying, multi-variate transaction data.

9 Conclusion

In this work, we delved into the visual analysis of the Bitcoin market development encompassing the evolution of Bitcoin exchanges. Close cooperation with domain experts allowed us to characterize two major targeted problems, namely, exchange selection priority for participants and exchange network evolution. Thus, we developed *BitExTract* and integrated this highly interactive visualization system into the exploration process to tackle these problems. We further validated our proposed system through three representative case studies and four well-designed structured interviews with domain experts. The results indicated that *BitExTract* performs efficiently in Bitcoin exchange intelligence extraction and cryptocurrency transaction data exploration.

In the future, we plan to integrate short-term impact analysis, *e.g.*, cryptographic attacks, Bitcoin forks, or dramatic price fluctuation. These kinds of activities take place frequently and influence the whole network thoroughly. Besides, the investigation of clients is an essential part of trading analysis. It would be more encompassing if users' behaviour patterns could be revealed. Moreover, except the exchanges, analysis on other communities like pools, dark markets or gambling websites could also reveal the insights of cryptocurrency world to us.

10 Acknowledgements

We would like to thank all the reviewers and domain experts for their valuable comments. This work is supported in part by grant RGC GRF 16208514 and NFSC Grants-61572488, 61673241.

[9] https://bit.ly/2MOxb7f
[10] https://blog.kraken.com/post/247/kraken-integrates-with-bitgo/
[11] http://fortune.com/2016/08/03/bitcoin-stolen-bitfinex-hack-hong-kong/
[12] https://themerkle.com/is-bitgo-to-blame-for-the-bitfinex-bitcoin-theft/

REFERENCES

[1] Bitbonkers.com. https://bitbonkers.com/.
[2] Bitcoin transaction visualization. http://bitcoin.interaqt.nl/.
[3] Bitcoin (usd) price. https://www.coindesk.com/price/.
[4] Bitnodes.earn.com. https://bitnodes.earn.com/.
[5] Blockchain.info. https://blockchain.info/tree/114688199.
[6] Cryptocurrency market capitalizations. https://coinmarketcap.com/.
[7] Elliptic.co. https://www.elliptic.co/.
[8] Realtime bitcoin globe. https://blocks.wizb.it/.
[9] View the bitcoin cryptocurrency specification in detail. http://mapofcoins.com/bitcoin.
[10] Walletexplorer.com: smart bitcoin block explorer. https://www.walletexplorer.com/.
[11] D. Archambault, J. Abello, J. Kennedy, S. Kobourov, K.-L. Ma, S. Miksch, C. Muelder, and A. C. Telea. Temporal multivariate networks. In *Multivariate Network Visualization*, pp. 151–174. Springer, 2014.
[12] G. J. K. R. Barsky A., Munzner T. Cerebral: Visualizing multiple experimental conditions on a graph with biological context. *IEEE Transactions on Visualization and Computer Graphics*, 14:1253–1260, 2008.
[13] G. D. Battista, V. D. Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia. Bitconeview: visualization of flows in the bitcoin transaction graph. In *Proc. IEEE Symp. Visualization for Cyber Security (VizSec)*, pp. 1–8, Oct. 2015. doi: 10.1109/VIZSEC.2015.7312773
[14] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. *EuroVis STAR*, 2, 2014.
[15] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer Graphics Forum*, vol. 36, pp. 133–159. Wiley Online Library, 2017.
[16] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Proc. IEEE Symp. Security and Privacy*, pp. 104–121, May 2015. doi: 10.1109/SP.2015.14
[17] U. Brandes and B. Nick. Asymmetric relations in longitudinal social networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2283–2290, 2011.
[18] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Wirevis: Visualization of categorical, time-varying data from financial transactions. In *Proc. IEEE Symp. Visual Analytics Science and Technology*, pp. 155–162, Oct. 2007. doi: 10.1109/VAST.2007.4389009
[19] F. B. Corinna Vehlow and D. Weiskopf. Visualizing dynamic hierarchies in graph sequences. *IEEE Transactions on Visualization and Computer Graphics*, 22(10), 2016.
[20] S. A. Davydenko, I. A. Strebulaev, and X. Zhao. A market-based study of the cost of default. *The Review of Financial Studies*, 25(10):2959–2999, 2012. doi: 10.1093/rfs/hhs091
[21] M. Fleder, M. S. Kester, and S. Pillai. Bitcoin transaction graph analysis. *Massachusetts Institute of Technology (MIT), Computer Systems Security*.
[22] N. Gandal and H. Halaburda. Competition in the cryptocurrency market. *NET Institute Working Paper No. 14-17*, 2014.
[23] D. Gauldie, S. Langevin, P. Schretlen, D. Jonker, N. Bozowsky, and W. Wright. Louvain clustering for big data graph visual analytics. 2013.
[24] P. Isenberg, C. Kinkeldey, and J.-D. Fekete. Exploring entity behavior on the bitcoin blockchain. In *Posters of the IEEE Conference on Visualization*, 2017.
[25] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8):e0161197, 2016.
[26] C. Kinkeldey, J.-D. Fekete, and P. Isenberg. Bitconduite: Visualizing and analyzing activity on the bitcoin network. In *EuroVis 2017-Eurographics Conference on Visualization, Posters Track*, p. 3, 2017.
[27] M. Kiran and M. Stanett. Bitcoin risk analysis. *NEMODE Policy Paper*, 2015.
[28] J. D. Kirkland, T. E. Senator, J. J. Hayden, T. Dybala, H. G. Goldberg, and P. Shyr. The nasd regulation advanced-detection system (ads). *AI Magazine*, 20(1):55, 1999.
[29] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert. A survey on visual analysis approaches for financial data. In *Computer Graphics Forum*, vol. 35, pp. 599–617. Wiley Online Library, 2016.
[30] L. Lin, L. Cao, and C. Zhang. The fish-eye visualization of foreign currency exchange data streams. In *proceedings of the 2005 Asia-Pacific symposium on Information visualisation-Volume 45*, pp. 91–96. Australian Computer Society, Inc., 2005.
[31] Z. Liu, J. Stasko, and T. Sullivan. Selltrend: Inter-attribute visual analysis of temporal transaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1025–1032, Nov. 2009. doi: 10.1109/TVCG.2009.180
[32] D. McGinn, D. Birch, D. Akroyd, M. Molina-Solana, Y. Guo, and W. J. Knottenbelt. Visualizing dynamic bitcoin transaction patterns. *Big data*, 4(2):109–119, 2016.
[33] L. P. McNaught, C. Using wordle as a supplementary research tool. *Qualitative Report*, 15:630–643, 2010.
[34] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pp. 127–140. ACM, 2013.
[35] T. Moore and N. Christin. Beware the middleman: Empirical analysis of bitcoin-exchange risk. In *International Conference on Financial Cryptography and Data Security*, pp. 25–33. Springer, 2013.
[36] M. Moser, R. Bohme, and D. Breuker. An inquiry into money laundering tools in the bitcoin ecosystem. In *eCrime Researchers Summit (eCRS), 2013*, pp. 1–14. IEEE, 2013.
[37] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
[38] K. V. Nesbitt and S. Barrass. Finding trading patterns in stock market data. *IEEE Computer Graphics and Applications*, 24(5):45–55, Sept. 2004. doi: 10.1109/MCG.2004.28
[39] T. Neudecker and H. Hartenstein. Could network information facilitate address clustering in bitcoin? In *International Conference on Financial Cryptography and Data Security*, pp. 155–169. Springer, 2017.
[40] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
[41] T. Pham and S. Lee. Anomaly detection in the bitcoin system-a network perspective. *arXiv preprint arXiv:1611.03942*, 2016.
[42] S. Ranshous, C. A. Joslyn, S. Kreyling, K. Nowak, N. F. Samatova, C. L. West, and S. Winters. Exchange pattern mining in the bitcoin transaction directed hypergraph. In *International Conference on Financial Cryptography and Data Security*, pp. 248–263. Springer, 2017.
[43] F. Reid and M. Harrigan. An analysis of anonymity in the bitcoin system. In *Proc. Risk and Trust and 2011 IEEE Third Int 2011 IEEE Third Int. Conf. Privacy, Security Conf. Social Computing*, pp. 1318–1326, Oct. 2011. doi: 10.1109/PASSAT/SocialCom.2011.79
[44] Z. Shi, G. M. Lee, and A. B. Whinston. Toward a better measure of business proximity: Topic modeling for industry intelligence. *MIS Quarterly*, 40(4):1035–1056, 2015.
[45] F. Tschorsch and B. Scheuermann. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys Tutorials*, 18(3):2084–2123, Mar. 2016. doi: 10.1109/COMST.2016.2535718
[46] W. P. Vogt. *Weighted Average*. Dictionary of Statistics & Methodology, 2005.
[47] T. von Landesberger, S. Diel, S. Bremm, and D. W. Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, 2015.
[48] C. Xie, W. Chen, X. Huang, Y. Hu, S. Barlowe, and J. Yang. VAET: A visual analytics approach for e-transactions time-series. *IEEE transactions on visualization and computer graphics*, 20(12):1743–1752, 2014.
[49] J. S. Yi, N. Elmqvist, and S. Lee. Timematrix: Analyzing temporal social networks using interactive matrix-based visualizations. *Intl. Journal of Human–Computer Interaction*, 26(11-12):1031–1051, 2010.
[50] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander. Where is current research on blockchain technology? A systematic review. *PloS one*, 11(10):e0163477, 2016.