BITCOIN PRICE ANALZYE USING AND PREDICTION USING DATA SCIENCE PROCESS

Syed Farras Hussain, Syed Ibrahim, Syed Meeran Durabudeen

Final year student, Dept of CSE, Dhaanish Ahmed College Of Engineering, Chennai.

Abstract—We present BitConduite, a visual analytics approach for explorative analysis offinancial activity within the Bitcoin network, offering a view on transactions aggregated by entities, i. e. by individuals, companies or other groups actively using Bitcoin. BitConduitemakes Bitcoin data accessible to non-technical experts through a guided workflow aroundentities analyzed according to several activity metrics. Analyses can be conducted at differentscales, from large groups of entities down to single entities. BitConduite also enables analysts to cluster entities to identify groups of similar activities as well as to explore characteristics and temporal patterns of transactions. To assess the value of our approach, we collected feedback from domain experts.

INTRODUCTION:

. Bitcoin is a digital pseudo-currency and payment system based on strong public cryptography: a crypto currency [1], [2]. It challenges several notions of traditional banking as well as government-regulated currencies and transactions: using Bitcoin people can by pass traditional centrally governed payment systems. Bitcoin is legal to use virtually everywhere and a number of countries have officially accepted I as 'private money' [3]. Millions of people have directly transferred Bitcoin virtual money through its peer-to-peer network while building a large open data source called the Bitcoin blockchain: transactions bundled in blocks that form a chain. Bitcoin, and in particular users' transaction activities, are an important data source to study be-cause little is known about how Bitcoin compares to fiat currencies. Understanding behavior around the currency can help to explain certain Bitcoinphenomena such as its large volatility. In addition ,a high level of technical expertise is required to extract, store and analyze Bitcoin transactions that domain experts who are interested in Bitcoin usually do not have. Only few approaches exis that lower the threshold of Bitcoin analysis and help with a deeper analysis.We present BitConduite (Fig. 1), a visual analytics approach for the analysis of different types of activities and actor profiles in the Bitcoin network. It focuses on identifying and characterizing (but not deanonymizing) entities: individuals,

1. Related Work

With the growing interest in Bitcoin as a financial and social phenomenon, methods and approaches to analyze Bitcoin data have emerged.

In this section we present the most closely related past approaches for visual analysis of Bitcoin data.Many websites offer simple visual analy-ses of Bitcoin blockchain data. For instance,blockchain.info [5] provides information suchas the Bitcoin market value or the number of transactions per block. Most of these websitesprovide information in the form of simple charts that resemble stock charts and presumably provide information for investors as target users. Only a small number of systems support more complex visual analyses of different Bitcoin characteristics. One example is SuPoolVisor [6] that supports surveillance of mining pools and de-anonymization of pool members. It visualizes information about mining pools (e. g. computing power) and their transactions. Similarly, Bit Ex-Tract [7] supports the analysis of Bitcoin ex-changes, i. e. platforms to buy and sell Bitcoin. Transactions between exchanges can be analysed over time as well as between exchanges and their clients. Both approaches focus on specific types of entities in Bitcoin (mining pools and exchange platforms) and are restricted to the respective subsets of transactions whereas in BitConduite we allow an exploratory analysis of transactions of all types of actors. On a more detailed level, BitCone View [8], displays the traces of specific transactions in a Gantt chart to support an analyst in detecting suspicious mixing of Bitcoins through the blocks (taint analysis). Other than BitConduite, t is tailored to one special task and provides insights on the transaction level only. Another visual approach of this kind is BlockChain Vis [9] that shows node-link diagrams of transactions and enables analysts to filter by block, number of

2. BitConduite System

Bit Conduite consists of a back end for data preparation and management as well as for high performance data access and a front end with a graphical user interface (GUI) that comprises five linked views (Fig. 1). We collaborated with three cryptography researcher from our institution who regularly provided feedback to inform the system's development. BitConduite's design is based on a data-first methodology [12] triggered by real-world data and high-level exploratory analysis tasks supported by the data source. Next, we describe the components of the system in detail.

2.1. Activity Measures

To systematically describe an entity activity we defined eight simple statistical measures together with the experts who emphasized the need for measures that are simple and easy to understand.

To ensure that the measures are sufficiently expressive we designed them to facilitate explanation of entity groups from the literature (such as A they et al.'s [13] user model). The eight measures are listed in Table 1. With this set of measures we are able to describe an entity's activity related to number, time, amount and type of transactions. In all views of the GUI, the colors shown in Table 1 consistently represent the activity measures. In the future, BitConduite will also be extended to include additional activity measures when other types of activities are analyzed.

2.2. Data Acquisition and Preparation

We downloaded the raw data of blocks and transactions, imported them into a MongoDB database and then extracted the transaction data into a column-oriented Monet DB database. The latter enables fast aggregation of the data needed for computing the activity measures. Due to Bit- Conduite's exploratory nature it requires computationally expensive re-computation of the measures for any time range on the fly. To accelerate this process, we further wrote the entity-related data into HDF5 files that are loaded into memory where the server software can access them quickly. We opted for an in-memory solution that uses the pandas (Python data analysis) library for fast data processing. We base all our analyses on high-level entities rather than on Bitcoin addresses. To do so, we aggregated all transaction addresses using the Reid and Harrigan [14] heuristic that has been shown to be effective [15]. First, we exported address pairs that appear together as inputs of a trans- action. From them we constructed a graph with the addresses as nodes and their co-occurrence (being input address of the same transaction) as links using the Network X library in Python. A Union Find algorithm yielded all the addresses that are linked to the entities, following the heuristic. The result is a list of addresses for each entity. We downloaded and scraped lists of known addresses from public sources such as Wallet Explorer [16]. With this information we were able to tag over 70,000 addresses that added context to the analysis with **BitConduite.**

2.3. Workflow

The workflow for exploratory analyses with BitConduite (Fig. 2) includes the following high level tasks: Overview. Inspecting an overview of all activity measures related to the whole dataset or to subsets. Filter. Specifying dynamic queries to filter data and focus on regions of interest over time and activity measures. Group. Defining and organizing groups of entities with similar activity. Cluster. Automatic grouping of entities across activity measures to determine suitable value ranges for creating meaningful groups. Details. Exploration of entities' characteristics in **detail.**

2.4. Graphical User Interface

BitConduite's GUI (Fig. 1) consists of five linked views: filter view, tree view, cluster view, entity browser and transaction view. They are integrated into a single page web application. All five views are dynamically updated with every change and can be manipulated independently for iterative exploration of the data.

In the following, we describe the five views and provide more details in the use cases section. 2.4.1. Filter View. The filter view (Fig. 1-A) is a dashboard that provides an overview on the temporal distribution of transactions as well as histograms for the activity measures listed in Ta- ble 1. Initially, the time and value distributions are displayed for all entities in the current data set. The analyst can filter entities on any of the histograms using brushing or a text input field. Pressing the filter button confirms the selection and filters the current set of entities. For some activity measures (those related to all transactions of an entity) the analyst can switch between smallest, average and largest value per entity.

2.4.2. Tree View. The tree view (Fig. 1-B) shows a hierarchy of partitions representing groups of entities, visualized as an icicle tree for its compactness. Initially, only the root node, rep- resenting the group of all entities, is visible and automatically marked as active. Every time a filter is executed in the filter view, it is applied to the currently selected node. A new row is added in the tree below its parent with two new sets of entities: those that fulfill the filter condition and the remainder set. In Fig. 3 we see the whole set of entities in the tree view (Fig. 1-B). The filter view histogram for "largest amount received" (Fig. 1-A) shows a range from 0–90,000 BTC. We apply a range filter of 0–10 BTC. A new row with two new groups of entities appears in the tree: those that fulfill the filter criterion (left) and the remaining ones (right). The small bars next to the labels signify the relative number of entities per class and tooltips provide detail-in-context. We show the nodes with equal widths and add a glyph to represent the number of entities inside each node. In an initial design, we scaled the size of each node by the number of entities but often groups were small and their nodes became hardly visible. Clicking on a node selects it as the current

context and updates the visualizations. Labeling

Principle of the entity group tree. Filtering (middle) the set of 108,540 entities (top) yields two subsets in the entity group tree (bottom): entities meeting the condition (left) and the remainder (right) the nodes and deleting rows is possible as well. Using the tree view, the analyst can iteratively build up entity group trees and switch between the groups to compare their characteristics. An export function saves the entity group tree to a file for archiving and sharing..

6. Conclusion and Future Work

The main contribution of Bit Conduite is to make in-depth exploratory analysis of Bitcoin entity activity possible, lowering the threshold for analysts without the technical background to prepare and handle the data. An important part of its workflow involves systematic and reproducible grouping of entity activities using filtering coupled with a tree representation. Clustering can be used to reveal new groups of entities with similar activity. Starting with large scale analyses it is possible to drill down and retrieve detailed information on single entities and display their transactions on a timeline. In two use cases we demonstrated how BitConduite can help characterize entity activity to answer questions relevant to Bitcoin experts. During a workshop with Bitcoin experts we learned that several research questions could easily be answered using Bit Conduite (e. g. about trends and outliers in activities or mining behaviour). Questions regarding temporal trends could not be answered (e. g. seasonality in the use) and pointed out a limitation of the approach. Ratings concerning BitConduite's usability (confidence, ease-of-use, learnability) were predominantly positive. Overall, five out of six experts said they would like to use BitConduite more frequently. Limitations of our approach stem from the fact that aggregation of addresses to entities provides a new perspective but adds uncertainty that cannot be reliably quantified. To decrease uncertainty one could include more external information such as tagging of entities, requiring deanonymization, which was not our goal in this project. The workshop showed that the most important extension of our work would be a more convenient comparison of temporal patterns. An additional view could be integrated, e. g. a radial chart or similar to facilitate comparison of activity patterns over time. Another useful extension would be similarity search, i. e. suggestion of entities similar to a specific entity of interest or search for anomalies, i. e. entities with abnormal activity. Lastly, future work will be to add the capability to track addresses and individual entities by integrating the functionality we demonstrated in a separate approach called the Block chain Entity

Explorer [20].

7. Acknowledgements

We thank the anonymous reviewers for their feedback. This research was partially supported by Labex Digi Cosme (project ANR-11- LABEX- 0045-DIGICOSME) operated by ANR as part of the program "Investisement d'Avenir" Idex Paris- Saclay (ANR-11-IDEX-0003-02). Special thanks are due to Daniel Augot from Inria Saclay / LIX for constantly supporting our work this project with expertise and ideas.

REFERENCES

1. S. Nakamoto, "Bit coin: A peer-to-peer electronic cashsystem," 2008. [Online]. Available: http://bitcoin.org/bitcoin.pdf/

2. A. Narayanan, J.Bonneau , E. Felten , A. Miller, and

S. Goldfeder, Bitcoin and Cryptocurrency Technologies:A Comprehensive Introduction. Princeton University Press, 2016.

3. "Wikipedia: Legality of bitcoin by country or territory." [Online]. Available: https://en.wikipedia.org/wiki/Legality of bitcoin by country or territory

4. M. Oppermann and T. Munzner, "Data-first visualization design studies," in Workshop on Evaluation and Beyond- Methodological Approaches to Visualization (BELIV),2020, pp. 74–80.

5. "Blockchain.info." [Online]. Available: https://blockchain.info

6. J.-z. Xia, Y.-h. Zhang, H. Ye, Y. Wang, G. Jiang,Y. Zhao, C. Xie, X.-y. Kui, S.-h. Liao, and W.-p.Wang, "SuPoolVisor: a visual analytics system for mining pool surveillance," Frontiers of Information Technology & Electronic Engineering, vol. 21, no. 4, pp. 507–523, Apr. 2020. [Online]. Available: http://link.springer.com/10.1631/FITEE.1900532

7. X. Yue , X. Shu, X. Zhu, X. Du, Z. Yu, D. Papadopoulos, and S. Liu, "Bit extract: Interactive visualization for extracting bit coin exchange intelligence," IEEE TVCG,vol. 25, no. 1, pp. 162–171, 2018.

8. G. Di Battista, V. Di Donato, M. Patrignani, M. Pizzonia,V. Roselli, and R. Tamassia, "Bitconeview: Visualization of flows in the bit coin transaction graph," in Symposium on Visualization for Cyber Security. IEEE, 2015.

9. S. Bistarelli and F. Santini, "Go with the—bitcoin—flow, with visual analytics," in Conference on Availability ,Reliability and Security. ACM, 2017, pp. 38:1–38:6.

10. D. Mc Ginn, D. Birch, D. Akroyd, M. Molina-Solana,Y.Guo, and W. Knottenbelt, "Visualizing dynamic bitcoin transaction patterns," Big Data, vol. 4, no. 2, pp. 109–119,2016.

11. N. Tovanich, N. Heulot, J.-D. Fekete, and P. Isenberg,"A Systematic Review of Online Bit coin Visualizations,"EuroVis 2019 - Posters, 2019, poster.

12. M. Oppermann and T. Munzner, "Data-first visualization design studies," CoRR, vol. abs/2009.01785, 2020.[Online]. Available: https://arxiv.org/abs/2009.01785

13. S. Athey, I. Parashkevov, V. Sarukkai, and J. Xia, "Bitcoin pricing, adoption, and usage: Theory and evidence,"Stanford University Graduate School of Business,Tech. Rep. Research Paper No. 16-42, 2016. [Online]. Available: https://ssrn.com/abstract=2826674

14. F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in Security and Privacy in Social Networks. Springer, 2013.

15. M. Harrigan and C. Fretter, "The unreasonable effective-ness of address clustering," in Conference on Advanced and Trusted Computing. IEEE, 2016, pp. 368–373.

Output : ScreenShots: