BitConduite: Visualizing and Analyzing Activity on the Bitcoin Network

Christoph Kinkeldey, Jean-Daniel Fekete, Petra Isenberg

Eurographics Conference on Visualization (EuroVis), Posters Track (2017), pp. 1–3
A. Puig and T. Isenberg (Guest Editors)

BitConduite: Visualizing and Analyzing Activity on the Bitcoin Network

C. Kinkeldey, J.-D. Fekete and P. Isenberg

Inria, France

Abstract

BitConduite is a system we are developing for the visual exploration of financial activity on the Bitcoin network. Bitcoin is the largest digital pseudo-currency worldwide and its study is of increasing interest and importance to economists, bankers, policymakers, and law enforcement authorities. All financial transactions in Bitcoin are available in an openly accessible online ledger—the (Bitcoin) blockchain. Yet, the open data does not lend itself easily to an analysis of how different individuals and institutions—or entities on the network—actually use Bitcoin. Our system BitConduite offers a data transformation back end that gives us an entity-based access to the blockchain data and a visualization front end that supports a novel high-level view on transactions over time. In particular, it facilitates the exploration of activity through filtering and clustering interactions. We are developing our system with experts in economics and will conduct a formal user study to assess our approach of Bitcoin activity analysis.

Categories and Subject Descriptors (according to ACM CCS): Visual Knowledge Discovery, Data Clustering, Data Filtering, Business and Finance Visualization

1. Introduction

Bitcoin is a digital pseudo-currency based on cryptography (a *cryptocurrency*) and a payment system that challenges several notions of traditional banking and government-regulated currencies and transactions [Nak08]. Using Bitcoin, traditional centrally governed payment systems can be bypassed. It is a comparably cheap and fast way to transfer money globally while the users remain pseudo-anonymous (*pseudonymous*), i. e., no personal information is stored other than abstract sending and receiving addresses. A large amount of "real" money has already been invested in infrastructures and global ecosystems around Bitcoin. Its economic value is estimated at more than 10 billion dollars at the time of writing [BMC*15].

Thus, Bitcoin, and in particular users' transaction activities are an important data source to study, as very little is known about the actors and their activities on the Bitcoin network. Unfortunately, the study of Bitcoin is hampered by its relatively complex pseudonymous data structure and overall transaction count—more than 200 million transactions in early 2017—and so far, few visual analytics tools exist that can help in the high-level analysis of Bitcoin.

We report on our progress developing BitConduite, a system to help study the different types of activities and actor profiles on the Bitcoin network. BitConduite focuses on identifying individual *entities* (actors on the network that may be individual users or different kinds of organizations or services such as currency exchange platforms or stores) and characterizing different groups of entities on the network by the kind of transactions they carried out over time.

2. Challenges

The heart of Bitcoin is the principle of a public ledger—the *blockchain*—in which all transactions between users are registered, distributed across the entire network, validated, and maintained. Thus, data on all Bitcoin transactions is publicly available and the data is valid because the system makes sure that transactions are not fraudulent. Despite all this, it is challenging to make sense of the Bitcoin data because entities (i. e., any kind of individual person, company or institution) on the network can send and receive Bitcoins using hundreds of pseudonymous addresses. Our approach to analyzing Bitcoin is to offer an entity-based view in the data by first finding out which addresses belong to the same entity and then building visualizations based on an entity's set of aggregated addresses. In this approach, we have to tackle several fundamental challenges:

Finding Entities: As stated above, Bitcoin users are pseudonymous: only abstract addresses without any personal information are used that typically change with every transaction. As a consequence, the raw data does not reveal which transactions belong to which entity and thus it is not possible to analyze entity-based activity without further data preparation. That is why methods are needed that use the network's topology to estimate which of the addresses may belong to an entity.

Classification of Entity Activity: Bitcoin is used in diverse ways, e. g., as investment ("digital gold"), for legal and illegal shopping, or for gambling, and its usage has changed a lot since its launch in 2009 [APSX16]. We want to classify entities by their activity patterns to shed light on how Bitcoin is used for and how. One strategy is to utilize decision trees like the one defined by Athey et al. [APSX16] to classify entities into types such as "Long-term infrequent transactor". Yet, existing approaches, to our knowledge, perform the classification based on a-priori hypotheses, whereas our goal is to facilitate explorative classification of entities more directly from the data.

Impact of External Events on Activity: Activity on the Bitcoin network is likely impacted by "real world" events. An example is the 2012/2013 Cypriot financial crisis during which people invested in Bitcoins as an alternative to gold in order to escape from their bank system [Cox13]. There are rumors about the role of Bitcoin during such events but it is difficult to measure them since long-time analysis of activity is still hardly possible. We envisage a tool that provides an overview on Bitcoin activity over long time periods to better be able to estimate the impact of external events.

3. Methodology and BitConduite Tool

To tackle the above-mentioned challenges we currently develop BitConduite, a visual analytics tool to facilitate exploratory analysis of activity of Bitcoin users over large-scale time periods. Contrary to existing approaches that visualize transactions on a detailed level for rather short time periods [BDP*15] [MBA*17] we provide a long-time and aggregated overview on entity activity by deriving high-level measures such as frequency of usage and amounts transferred. During the development we are guided by regular meetings with experts in economics with an interest in analyzing Bitcoin for their own research. The system's back end is implemented in Python and the front end in JavaScript/D3.

3.1. Data Processing

Since our approach is to analyze Bitcoin data at different levels of aggregation instead of the raw data only, it is crucial to have an appropriate data infrastructure for data processing (Figure 1). We obtained the raw data from the Bitcoin Core client [Bit] and stored it in a MongoDB document database. For further data processing we use a column-oriented MonetDB database [Mon] particularly suited for data aggregation that provides fast access to the data used in BitConduite.

Using this infrastructure we derive entities from pseudonymous addresses using the input heuristics suggested by [RH13]. In order to group entities by similarity we use k-means clustering with a user-defined number of clusters.

3.2. Visual Analysis Interface

There are three major components of the BitConduite visual analysis interface: filter view, cluster view, and timeline view:

Filter view: The first step in the analysis is to filter out entities with certain attributes. For instance, it may be of interest to filter

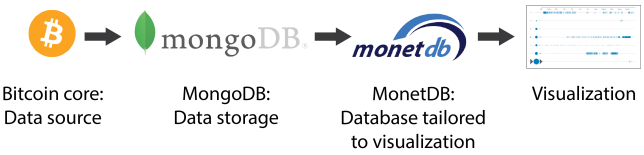

Figure 1: *Data workflow from raw Bitcoin data to aggregated data used for the visualization.*

out entities with only one transaction (one-time users) or entities that have not been active anymore for a long time. We are currently extending the existing filtering functionality by a decision tree that makes it possible to define reproducible classifications of entities.

Cluster view: The filtered set of entities can be clustered so that similar ones are grouped. For this, the analyst can select a number of attributes that are of interest, as well as the number of clusters. This way, entities with similar maximum amounts of Bitcoin transferred can be grouped in order to make comparisons between entities transferring high amounts of Bitcoin.

Timeline view: The main component of the tool consists of a number of horizontal timelines, representing transactions over time for each entity cluster (Figure 2). A bar and a glyph on the left show the number of entities per cluster and further information on each cluster (number of transactions, mean amount transferred, etc.). The analyst can retrieve the temporal distribution of transactions belonging to each cluster and get further information from tooltips by hovering over the glyphs or transactions in the timeline.

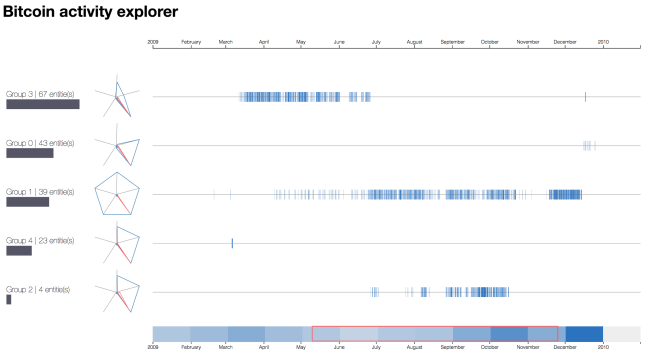

Figure 2: *Timeline view showing transactions grouped by entity clusters over time as well as attributes of each entity cluster.*

4. Outlook

The development of the BitConduite tool is ongoing work. The focus of our current work is to provide guidance for the analyst when exploring entity activity and making filtering and clustering parameters transparent and reproducible. This is planned to facilitate the definition and exchange of activity classifications between analysts. While the development of the visualizations is already informed by expert users, the final part of the development will be an end-user evaluation to assess what novel insights can be gained from the visual analytics tool compared to existing solutions.

References

[APSX16] ATHEY S., PARASHKEVOV I., SARUKKAI V., XIA J.: Bitcoin pricing, adoption, and usage: Theory and evidence. *Stanford University Graduate School of Business Research Paper No. 17-14* (2016). URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2826674. 2

[BDP*15] BATTISTA G. D., DONATO V. D., PATRIGNANI M., PIZZONIA M., ROSELLI V., TAMASSIA R.: Bitconeview: visualization of flows in the bitcoin transaction graph. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)* (Oct 2015), pp. 1–8. doi:10.1109/VIZSEC.2015.7312773. 2

[Bit] Bitcoin core client. URL: http://bitcoin.org/en/download. 2

[BMC*15] BONNEAU J., MILLER A., CLARK J., NARAYANAN A., KROLL J. A., FELTEN E. W.: Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *IEEE Symposium on Security and Privacy* (May 2015), pp. 104–121. doi:10.1109/SP.2015.14. 1

[Cox13] COX J.: Bitcoin bonanza: Cyprus crisis boosts digital dollars, 2013. URL: http://www.cnbc.com/id/100597242. 2

[MBA*17] MCGINN D., BIRCH D., AKROYD D., MOLINA-SOLANA M., GUO Y., KNOTTENBELT W. J.: Visualizing dynamic bitcoin transaction patterns. *Big Data 2*, 4 (2017), 109–119. doi:10.1089/big.2015.0056. 2

[Mon] Monetdb - the column store pioneer. URL: http://www.monetdb.org/. 2

[Nak08] NAKAMOTO S.: Bitcoin: A peer-to-peer electronic cash system. URL: http://bitcoin.org/bitcoin.pdf. 1

[RH13] REID F., HARRIGAN M.: *An Analysis of Anonymity in the Bitcoin System*. Springer New York, New York, NY, 2013, pp. 197–223. doi:10.1007/978-1-4614-4139-7_10. 2