Visual Analytics for Monitoring and Exploration of Blockchain Data With a Focus on the Bitcoin Blockchain

Petra Isenberg, Christoph Kinkeldey, Jean-Daniel Fekete

Visual Analytics for Monitoring and Exploration of Blockchain Data With a Focus on the Bitcoin Blockchain

Petra Isenberg
Inria
Palaiseau, France
petra.isenberg@inria.fr

Christoph Kinkeldey
Inria
Palaiseau, France
christoph.kinkeldey@inria.fr

Jean-Daniel Fekete
Inria
Palaiseau, France
jean-daniel.fekete@inria.fr

CHI, 2018, Montréal, Canada
ACM .

Abstract
The goal of our work is to devise visual analytics techniques to deeply explore data stored on blockchains. Given the novel data storage nature of blockchains, they offer a unique opportunity to study the evolution of the stored data as well as data storage patterns. Opposed to most existing approaches that analyze the raw data we focus on aggregated measures. In particular, we analyze long-term activities and monitoring of activities that led to building the blockchain. We currently focus on the Bitcoin blockchain as the most popular blockchain with a turbulent history.

Author Keywords
Blockchain; Bitcoin; visualization; visual analytics; monitoring; behavior

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Blockchain technology was originally invented to support recordkeeping of financial transactions that are sent and received without any central intermediary. The Bitcoin blockchain, for example, is the heart of the Bitcoin system. It represents a public ledger in which all transactions between pseudonymous users are registered, validated, main-

tained, and distributed across the entire network of users [3, 10]. It can only be added to, and past financial transactions stored in the blockchain cannot be changed. Given the inherent storage of all historical records, blockchains offer a unique opportunity to study the temporal evolution of the data as well as how actors add data to the blockchain.

The goal of our work is to devise visual analytics techniques to deeply explore data stored on blockchains. We focus on long-term activities and monitoring: effects that cover the entire blockchain or that are recent and different from what happened in the past, to try to understand what activities people or enterprises (=entities) exhibit on the blockchain. Long-term effects are more informative about behaviors of entities on the blockchain than behavioral patterns extracted from short temporal snippets. There are several questions and tasks that we want to target:

- How has the activity on the blockchain changed over the years?
- What is the general behavior of blockchain actors? E.g. can we identify groups of actors based on the stored data associated to them?
- How do external events influence the activity on the blockchain?
- How do internal events (e.g. security issues, fraud, etc.) influence the activity on the blockchain?
- How does activity on two different blockchains relate?

In particular, our work has begun with a focus on the Bitcoin blockchain as the most popular blockchain with a large enough dataset to provide an interesting analysis challenge. Currently, the raw data alone contains over 148 million transactions (>150 GB of raw data) and is constantly growing. Visualizing this many transactions requires techniques that work on multiple levels of data aggregation. We aim to provide complete overviews of the network and also to allow analysts to interactively drill down and see close details of transactions or individual actors on the network.

Past Work

We currently collaborate with economists working on the Bitcoin blockchain. The economists want to understand the activities on the blockchain and compare them with related economic activities with regular currencies. This is particularly important as 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 [3]. Thus, Bitcoin, and in particular users’ transaction activities are an important data source to study, as very little is known about the entities and their activities on the Bitcoin network. One particuliarity of the Bitcoin blockchain, however, is that users remain pseudo-anonymous (*pseudonymous*), i. e., no personal information is stored other than abstract sending and receiving addresses. 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.

Our data infrastructure involves collecting the raw data from the Bitcoin Core client [11] and to store it in a MongoDB database. We make use of Reid and Harrigan’s clustering heuristic [12] to combine addresses for individual entities. We store the resulting clusters together with other aggregate information in a MonetDB [9] database. This infrastructure is the basis of two specific tools we developed:

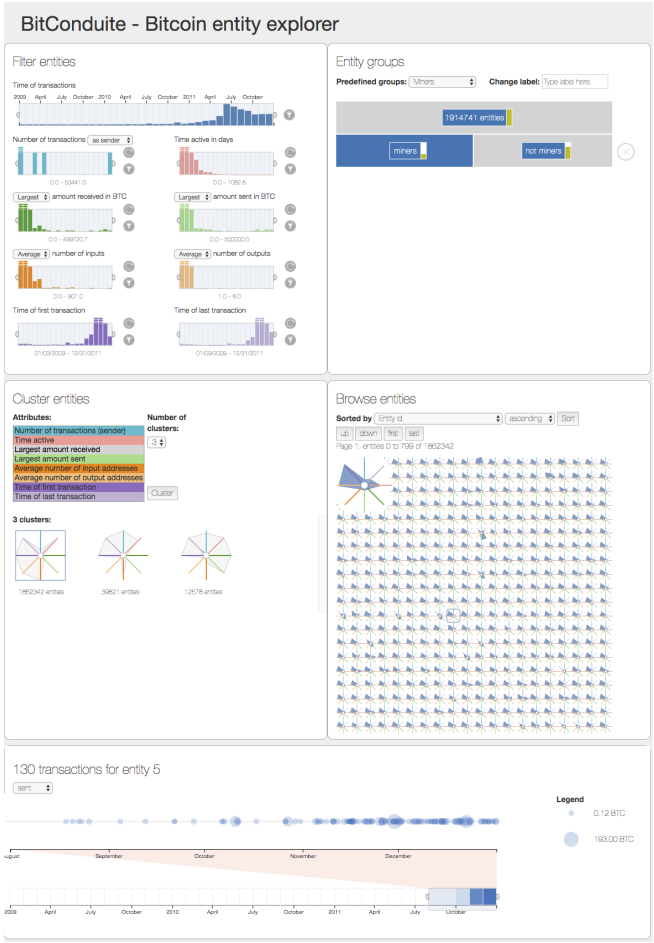

Figure 1: Bitconduite: The analysts can filter entities (top left), build a decision tree (top right), cluster entities by selected attributes (middle left), and compare individual entities and their transactions (middle right & timeline at the bottom).

Bitconduite

Bitconduite (Figure 1) [6] offers a visualization front end that supports a novel high-level view on entities' activity on the blockchain over time. In particular, it facilitates the exploration of activity through filtering and clustering interactions. The first step in the analysis is to narrow down the data to entities of interest using filters of derived blockchain activity metrics for each entity (Figure 1, top left). For example, we have calculated metrics such as number of transactions, time of activity, amount of Bitcoins received/sent, etc. For instance, it may be of interest to an analyst to compare miners to non-miners, or people who were early adopters to people who have joined recently. We include a decision tree (Figure 1, top right) that makes it possible to define reproducible classes of entities. Next, a set of chosen entities (e. g. miners) can be clustered according to a set of chosen blockchain attributes so that similar ones are grouped (Figure 1, middle left). For example, miners with similar maximum amounts of Bitcoin received can be grouped in order to make comparisons between very successful and less successful miners. Finally, each individual entity per group of entities (Figure 1, middle right) can be explored in a browser component according to the set of activity metrics we calculated. A timeline (Figure 1 bottom) gives a temporal distribution of transactions belonging to a selected entity and provides information on the transferred amounts per transaction.

In our future work we plan to extend the system to include outside information. For example, our partners in economics are asking for the price of Bitcoin to be added (in dollars) and to make annotations to the timeline with outside events (such as the breakdown of Mt. Gox) that may have influenced the behavior of entities.

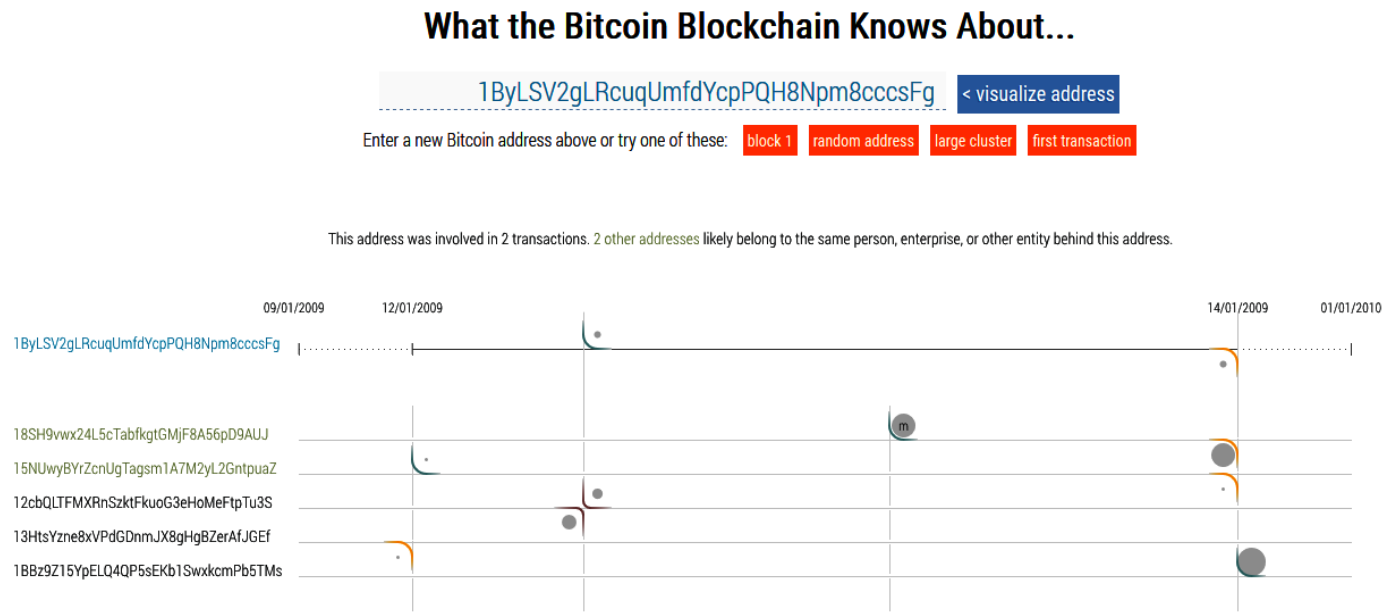

Figure 2: Overview of our Bitcoin Blockchain Entity Explorer. The data shows information on an input address (top) involved in a Bitcoin transaction with Satoshi Nakamoto (Bitcoin's founder). This entity was involved in 4 transactions (vertical lines) and also mined Bitcoins (third transaction from the left). The second line from the left represents a transaction in which Nakamoto sent Bitcoin to the input address at the top and returned the remainder of the transaction back to the original input address.

The Bitcoin Entity Explorer

Our second system, the Bitcoin Entity Explorer [5], allows to analyze the transactions of an entity in detail. It allows Bitcoin users and casual blockchain analysts to visually track their own or others' transactions based on a single input address. It differs from other visualization tools that show connections between transactions (e. g. [1, 2, 4, 7, 8]) by focusing on single entities and their connections to others.

Figure 2 gives an overview of our tool using an input address involved in a transaction with Satoshi Nakamoto, Bitcoin's mysterious founder [10]. The visualization shows an activity timeline for the chosen address and all other addresses belonging to the same entity in green. The solid portion of the timeline represents the period of time in which the entity was active while the endpoints connected by a dotted line show the temporal context of the whole dataset. In Figure 2 we loaded data from Jan. 9, 2009–Jan. 1, 2010. Each transaction on the timeline is represented by a vertical line connecting to horizontal lines emanating from the

displayed addresses. When an address is involved in a particular transaction, glyphs are drawn at the intersection of transaction and address lines. Transaction glyphs represent the amount of Bitcoin transferred with a circle of a size relative to the maximum amount transferred by this entity. In addition, input glyphs have an orange curve, output glyphs are blue, and addresses that served as both input and output have a red glyph with two curved lines. Some out-glyphs include an "m" standing for coins newly created through a process called mining. This interface communicates how an entity's addresses are connected internally and to external addresses.

One of the main challenges with the current design is scalability. In the data between 2009–2010, the largest entity already had 901 addresses. We expect that entity sizes grow rapidly especially for data from recent years where the process of creating new addresses has been extremely simplified in Bitcoin management software. We are therefore currently implementing a multi-scale data exploration approach where large entities are first displayed in an aggregated manner. We envision these aggregated clusters to be opened up into sub-clusters grouped by similarity. We will also use focus+context techniques on the timeline to allow for more detailed exploration of specific transaction clusters coupled with filtering to only the involved addresses.

Future Work

An immediate goal to our work is to integrate the Entity Explorer into the Bitconduite tool and to add information external to the Bitcoin blockchain into the tool. We have also conducted a first workshop with Bitcoin analysts using Bitconduite and plan to integrate further usability enhancements based on the outcome of this workshop.

Based on our work on the Bitcoin blockchain we plan to later explore other blockchains, such as Ethereum or consortium (half-private) blockchains when their data becomes available to us. We will reach out to consortium blockchain managers later in the project. The goal is to expand on our tools tailored to Bitcoin and offer visual analytics techniques that allow consortium blockchain managers similarly to understand how their blockchain evolves, according to their plans and how it compares to other blockchains.

Acknowledgements

This research was partially supported by Labex DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) operated by ANR as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02)

Biographies

Attending author:

Petra Isenberg is a Research Scientist at Inria. Her main research areas are information visualization and visual analytics with a focus on collaborative work scenarios, interaction, and evaluation. She has already advised two Master's theses and a PostDoc on the topic of Bitcoin visual analytics and is currently searching for a PhD student to continue the work. She is also the author of the Bitcoin Entity Explorer described above.

In addition to her work on Bitcoin visual analytics, Petrae is interested in exploring how people can most effectively analyze data sets on novel display technology such as small touch-screens, wall displays, or tabletops. She hopes to extend her work on Blockchain visual analytics to large wall-sized displays. Petra received her Ph.D. in Computer Science from the University of Calgary in 2009. You can find additional information on her work at: http://petra.isenberg.cc/.

Remaining authors:

Christoph Kinkeldey is a postdoc at Inria working on the topic of visualizing and analyzing activities on the Bitcoin blockchain. His main research interests are visual analytics, uncertainty visualization, and geovisualization. He received his doctorate in Geoinformatics from the HafenCity University Hamburg in 2015.

Jean-Daniel Fekete is Senior Research Scientist at Inria and scientific leader of the project team AVIZ. His main Research areas are Visual Analytics, Information Visualization and Human Computer Interaction. He is the chair of the IEEE Information Visualization Conference Steering Committee, member of the IEEE VIS Executive Committee, member of the Eurographics EuroVis Steering Committee, and member of the Eurographics publication board. He is also an ACM Distinguished Speaker.

REFERENCES

1. G. D. Battista, V. D. Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia. 2015. Bitconeview: Visualization of flows in the bitcoin transaction graph. In *Proc. Symposium on Visualization for Cyber Security (VizSec)*. 1–8. DOI: http://dx.doi.org/10.1109/VIZSEC.2015.7312773
2. Bitbonkers.com. ~2017. Website. (~2017). http://bitbonkers.com, visited 03/2018.
3. Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, and Edward W. Felten. 2015. SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In *IEEE Symposium on Security and Privacy*. 104–121. DOI: http://dx.doi.org/10.1109/SP.2015.14
4. Luc Hendriks. ~2017. Website. (~2017). http://bitcoin.interaqt.nl/, visited 03/2018.
5. Petra Isenberg, Christoph Kinkeldey, and Jean-Daniel Fekete. 2017. Exploring Entity Behavior on the Bitcoin Blockchain. In *Posters of the IEEE Conference on Visualization*. https://hal.inria.fr/hal-01658500
6. Christoph Kinkeldey, Jean-Daniel Fekete, and Petra Isenberg. 2017. BitConduite: Visualizing and Analyzing Activity on the Bitcoin Network. In *Eurographics Conference on Visualization (EuroVis), Posters Track*. https://hal.inria.fr/hal-01528605
7. Maximilian Laumeister. ~2017. Website. (~2017). http://www.bitlisten.com/, visited 03/2018.
8. D. McGinn, D. Birch, D. Akroyd, M. Molina-Solana, Y. Guo, and W. J. Knottenbelt. 2017. Visualizing dynamic Bitcoin transaction patterns. *Big Data* 2, 4 (2017), 109–119. DOI: http://dx.doi.org/10.1089/big.2015.0056
9. MonetDB. 2017. Database. (2017). http://www.monetdb.org/
10. Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. Published online.. (2008). http://bitcoin.org/bitcoin.pdf
11. Bitcoin Project. 2017. Bitcoin core client. (2017). http://bitcoin.org/en/download
12. Fergal Reid and Martin Harrigan. 2013. *An Analysis of Anonymity in the Bitcoin System*. Springer New York, New York, NY, 197–223. DOI: http://dx.doi.org/10.1007/978-1-4614-4139-7_10