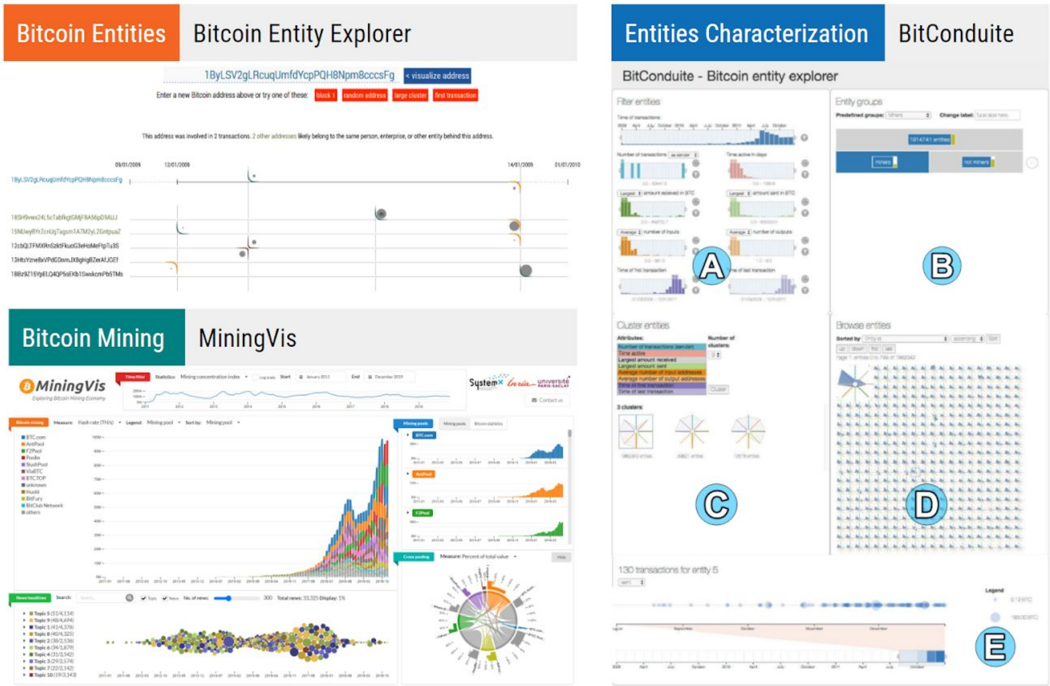
Visual Analytics of bitcoin Blockchain Data

Advisors

Natkamon Tovanich, natkamon.tovanich@inria.fr
Petra Isenberg, petra.isenberg@inria.fr



This address was involved in 2 transactions. 2 other addresses likely belong to the same person, enterprise, or other entity behind this address.

Screenshots of our works on visual analytics of Bitcoin blockchain data in different task domains

Topic

Bitcoin is the first and so far the highest valued cryptocurrency blockchain. It was introduced by Satoshi Nakamoto in 2008. The Bitcoin blockchain is a distributed peer-to-peer network storing append-only transaction data in which all transactions between pseudonymous users are registered, validated, maintained, and distributed across the entire network of users. The advantage of the technology lies in the decentralized system governed by autonomous logic, in contrast to a centralized system controlled by the government, bank, or any organization.

As the Bitcoin data is constantly growing (> 300 GB of raw data), it offers a unique opportunity to study the evolution of the transaction data as well as the interactions of users in the network. The Bitcoin network involves diverse groups of users (e.g., individuals, enterprises, miners, and exchanges), and their activities are influenced by multiple factors from both internal (e.g., Bitcoin protocol, frauds, and cyber-attacks) and external (e.g., news and market price) historical events. The visual analytics approach allows us to interactively explore Bitcoin data in different levels of aggregation, detect the changes in collective activities, and characterize different patterns on the network.

Topic 1: Characterization of Entities in the Bitcoin Blockchain.
The goal of this master's thesis topic is to devise visual analytics techniques to deeply explore data stored on the Bitcoin blockchain. We will focus on exploring and monitoring different types of users *(entities)* to better understand what is going on in the Bitcoin blockchain. There are several questions and tasks that the research tools will target:

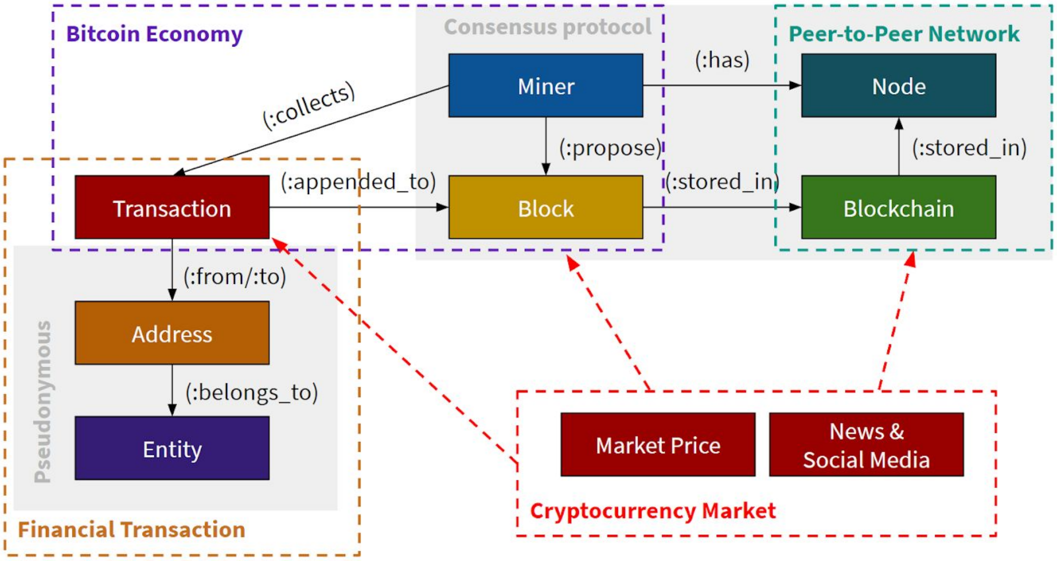
- How to identify the entity (set of addresses that belong to the same users) from the large Bitcoin transaction data?
- What are the different types of Bitcoin users? Can we identify groups of entities based on the transaction activities and network associated with them?
- What is the general behavior of the entity types in Bitcoin? How has the activity of each entity type changed over the years?
- How to analyze transaction activities and relationships of entities over time? How do transaction activities on two or more different entities relate?

The student will conduct the master's thesis in collaboration with researchers working on the ANR project BITUNAM (Bitcoin User Network Analysis and Mining).

Topic 2: Analysis and Visualization of the Bitcoin Forum
This master's thesis focuses on analyzing textual contents in Bitcointalk.org, the first and one of most active discussion forums among Bitcoin users. The student will work on crawling data from the forum website and applying text mining techniques to analyze discussions in the forum. The goal is to develop a visual analytic tool to help economists studying the evolution of discussions in the Bitcoin forum and related the information with the activities inside the Bitcoin network.

The student will conduct the master’s thesis in collaboration with the economist from LITEM, Institut Mines-Télécom Business School who is working on research in the field of cryptocurrency and blockchain economic analysis.

Background



An overview of Bitcoin blockchain components

The Bitcoin blockchain records all *transactions* in a public ledger (i.e. database) that is copied and stored among peers in the distributed network. Transactions record the Bitcoin value transfer from the input address(es) to the output address(es). An *address* is represented as a long string with cryptographic properties that can be validated by its owner with the private key. All the value of input addresses is sent to the output addresses. The owner can send the change back to any of his or her addresses if he or she wants to transfer less than the total value of inputs. The owner also pays transaction fees as the difference between input and output values.

Transactions are publicly available on the blockchain but the owner of the address cannot be inferred directly from the address, hence they are referred to as *pseudonymous*. A common practice is the owners should regularly change addresses for privacy and security purposes. Yet, we can still trace the activities of *entities* from address clustering heuristics and publicly available datasets that provide a list of addresses that belong to well-known entities (e.g. WalletExplorer).

The problem of a decentralized system without a central authority is that the owner may spend transactions twice while the recipient does not notice that the bitcoin has already been spent, the *double-spending* problem. Bitcoin solves this problem by adding new transactions to the chain of *blocks* that records all previous transactions, hence the term *blockchain*. Transactions are validated and appended into a block by a pool of people called miners. *Miners* perform a *proof-of-work* that involves running computationally expensive methods to append a new block to the ledger. The miner who successfully proposed a new block can reclaim a coinbase transaction that includes newly generated Bitcoin values and transaction fees from every transaction in a block. The coinbase reward is halved every 210,000 blocks. The difficulty of mining is decided by the hash rate, the total computation power in the blockchain network, which often adapts to reach the desired rate of adding a new block every 10 minutes.

Due to the rapid growth of computational mining power, nowadays, individual miners are hardly expected to receive the mining reward in the short term. They also need to bear the cost of purchasing specific hardware for mining and electricity costs. In practice, miners are sharing their computational resources to mining pools to receive a more stable and predictable income.

Because Bitcoin is a decentralized network, it is vulnerable to cyber attacks from criminals or dishonest users that exploit the *consensus protocol* to manipulate or disfunction the network. The attacks can happen at multiple levels. Some malicious users may try to double-spend their Bitcoin value (e.g. Finney attack and Brute force attack). Bitcoin money can be stolen from the security breach as it happened to the Mt.Gox exchange company. In the mining activities, if a few miners control more than 50% of the total mining power, they can perform 51% attacks to alter the record of transactions. The peer-to-peer network is also vulnerable to network attacks (e.g. Distributed Denial-of-Service (DDoS) and Sybil attacks).

Apart from internal components in the Bitcoin blockchain, a cryptocurrency exchange market provides platforms to exchange the Bitcoin value to the currency (e.g. US Dollar) and so forth. Thus, it determines the *market price* of Bitcoin. The exponential growth of Bitcoin value in recent years is partially explained by the spread of narratives from *news and social media* to a wider public. These external elements of the Bitcoin blockchain have some influences on internal activities in Bitcoin which are worth further empirical studies.

References

Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. "Visualization of Blockchain Data: A Systematic Review." *IEEE Transactions on Visualization and Computer Graphics* (2019). https://hal.archives-ouvertes.fr/hal-02426339v1

Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse. "Knowledge Discovery in Cryptocurrency Transactions: A Survey." *arXiv preprint arXiv:2010.01031* (2020). https://arxiv.org/abs/2010.01031

Matthias Lischke, and Benjamin Fabian. "Analyzing the bitcoin network: The first four years." *Future Internet* 8, no. 1 (2016): 7. https://www.mdpi.com/1999-5903/8/1/7

Marc Jourdan, Sebastien Blandin, Laura Wynter, and Pralhad Deshpande. "Characterizing entities in the bitcoin blockchain." *IEEE International Conference on Data Mining Workshops* (2018). https://arxiv.org/abs/1810.11956