BlockSci: a platform for blockchain science and exploration

Harry Kalodner, Malte Möser, Steven Goldfeder, Alishah Chator, and Arvind Narayanan
Princeton University

Why we need analytics

- Motivate scaling solutions
- Discover new areas for improvement
- Categorize demand by use cases

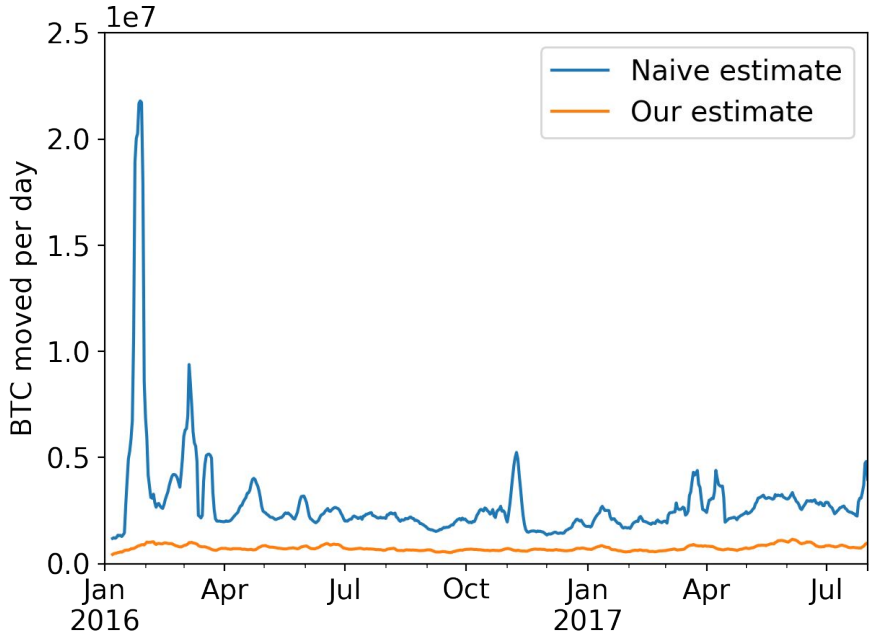
Can we tell the difference between organic and artificial demand?

Differentiating demand

- Real spending or self churn
- Store of value or medium of exchange
- Organic demand or malicious spam

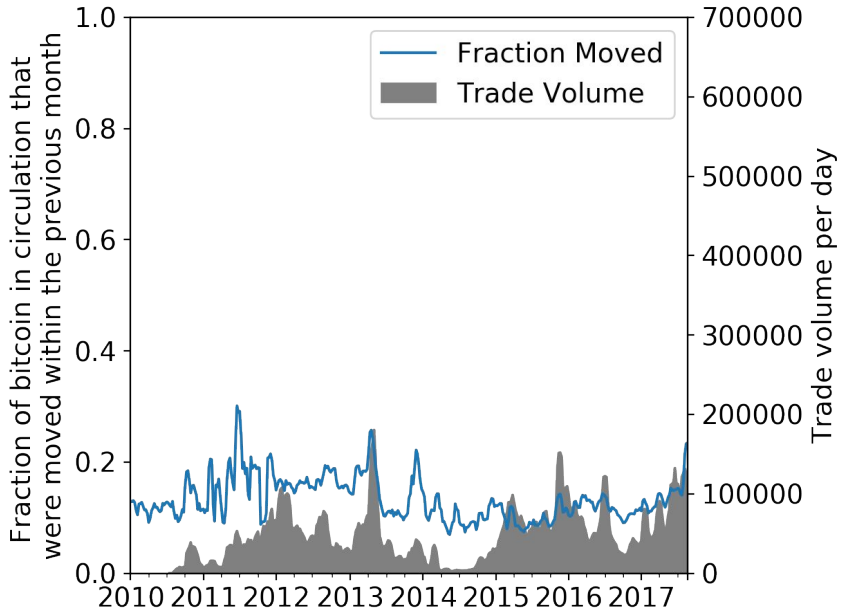
The velocity of Bitcoin

- Understand usage
- Discount self churn
- Ignore false flags



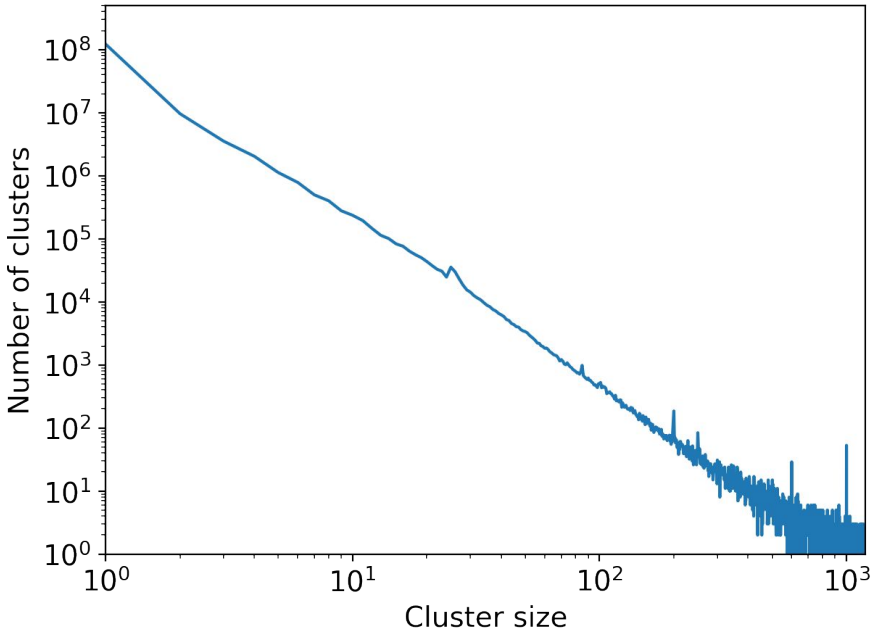
Bitcoin as a store of value

Most spikes in activity on the blockchain correlate with spikes in trade volume



Attempting to understand wallets

- Link addresses together to form wallets
- Allow users to easily and rapidly try out different heuristics

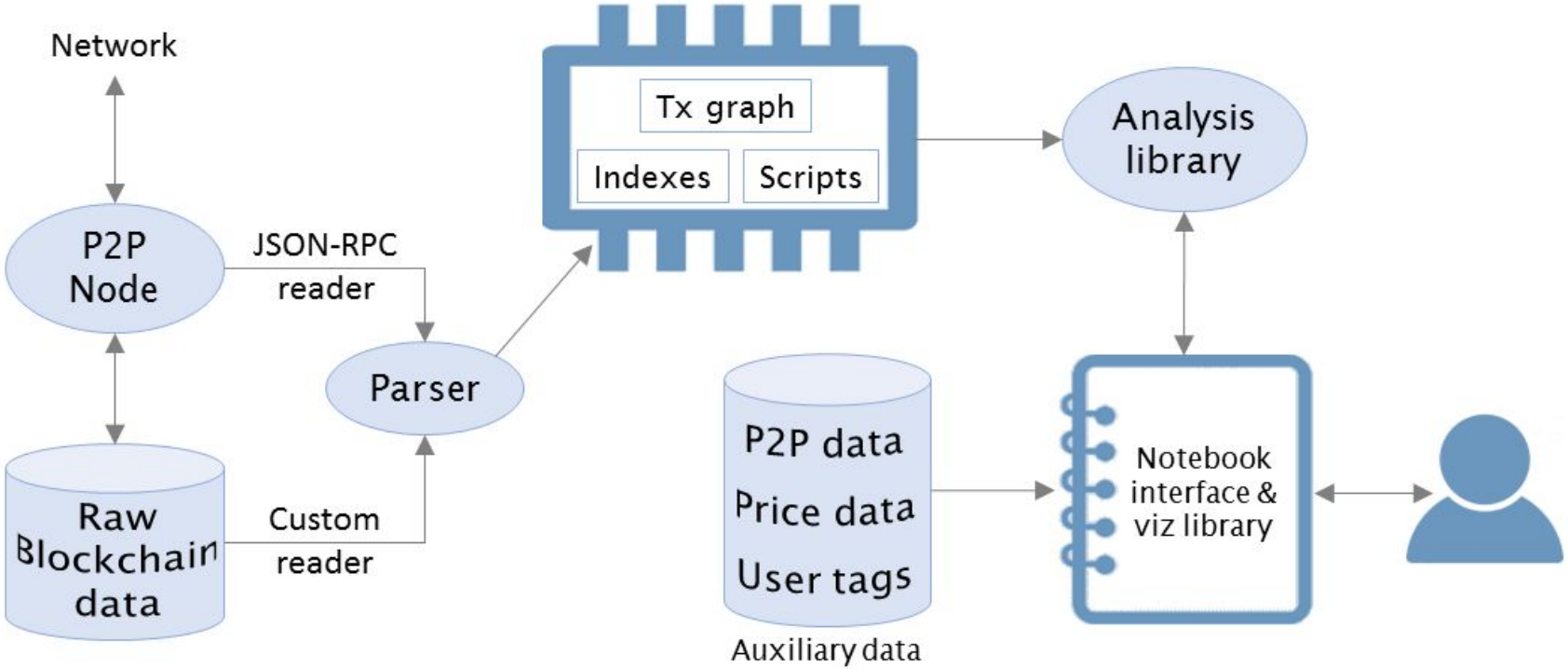


Aren’t there existing tools for this?

- Closed source
- Limited functionality
- Insufficient performance

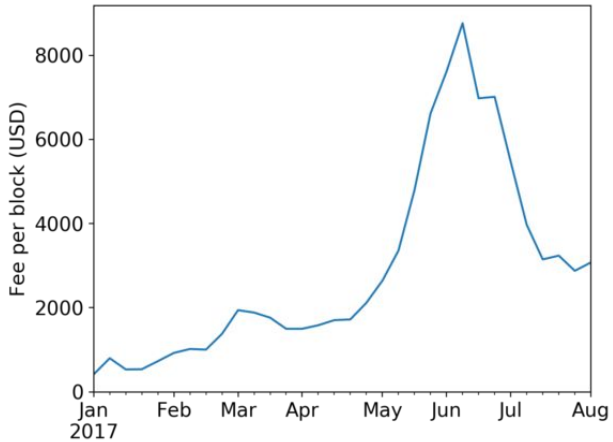
The solution

BlockSci


Network
P2P Node
JSON-RPC reader
Parser
Raw Blockchain data
Custom reader
Tx graph
Indexes
Scripts
Analysis library
P2P data
Price data
User tags
Auxiliary data
Notebook interface & viz library

BlockSci

```
fees = [sum(block.fees) for block in chain.range('2017')]
times = [block.time for block in chain.range('2017')]
converter = blocksci.CurrencyConverter()
df = pandas.DataFrame({"Fee":fees}, index=times)
df = converter.satoshi_to_currency_df(df, chain)
```



Broad Applicability

- Python interface for easy usage
- Easy incorporation of external data feeds
- Full support for all standard script types

Performance

Iterating Over	Single Threaded	Multithreaded
Transaction headers	13.1 sec	3.2 sec
Transaction outputs	27.9 sec	6.6 sec
Transaction inputs & outputs	46.4 sec	10.3 sec

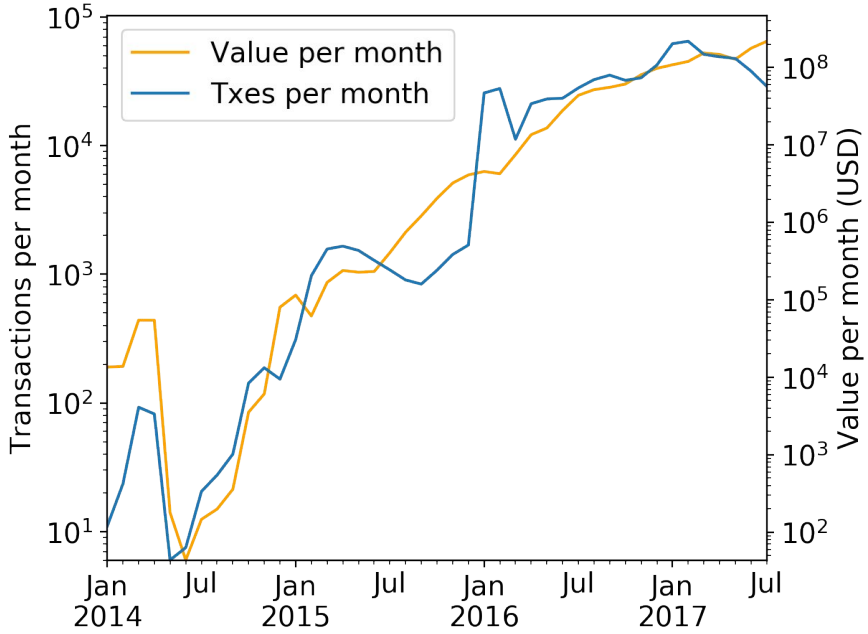
*Up through block 478,449

How do we achieve this?

- A customized data format optimized for locality of information
- Coded in C++ for maximal efficiency
- Uses memory mapping to allow easy scalability

Effects on privacy

- Valuable multisig addresses undergo a partial keyset change
- Organizational information is leaked



Existing work

- Goldfeder, Steven, et al. "When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies." arXiv preprint arXiv:1708.04748 (2017).
- Kalodner, Harry, et al. "BlockSci: Design and applications of a blockchain analysis platform." arXiv preprint arXiv:1709.02489 (2017).

Thank you!

https://github.com/citp/BlockSci