Crypto Data and Finance

Igor Makarov
LSE

Columbia, October 21, 2022
Agenda

- Crypto data
- Blockchain analysis of the Bitcoin market
Crypto data

- Off-chain: Not recorded on the blockchain
  - CEX and other centralized intermediaries

- On-chain: Recorded on the blockchain
  - Coin transfers
  - dApps (smart contracts automating financial services)
Off-chain data

- Crypto CEX operate similar to traditional exchanges
  - Limit order book
  - Spot and derivatives contracts
  - Investor level data are proprietary

- Current book level and transaction data are typically freely available via exchange APIs

- Historical data
  - Exchange-on-exchange basis
  - Public web-sites, e.g., http://api.bitcoincharts.com/v1/csv/
  - Specialized data providers, e.g., Kaiko

- Exchanges are non-integrated: A unique laboratory for studying arbitrage and price formation, see e.g., Makarov and Schoar (2020)
On-chain data

- Blockchain technology introduces new ways of storing data and new set of intermediaries

- Data
  - Transactions are public
  - Pseudonymous cryptocurrency addresses

- New intermediaries
  - Validators (miners)
  - dApps
Data challenges

• Crypto addresses are easy to generate ⇒ potentially many addresses belong to the same entity

• Link anonymous addresses to real-life entities

• Processing smart contract data

• Cross-chain analysis
Blockchain analysis of the Bitcoin market

- Systematic analysis of the Bitcoin market using blockchain data
  - We build a novel Bitcoin database and methodology for identifying information about the main market participants

- Three major pieces of analysis:
  - **Network structure:** Analyze the transaction volume and network structure of the main market participants
    - Exchanges are central entities; 75% of volume is linked to exchanges
  - **Ownership concentration:** Document the ownership concentration of the largest bitcoin investors
    - High concentration: top 0.01% investors control 24% of bitcoins in circulation
  - **Miners:** Study the concentration and regional composition of miners who ensure the integrity of the Bitcoin blockchain
    - High concentration: 50 miners often control 50% of the total market
Data construction (1)

- Use the BlockSci program to convert raw data into a database

- Clustering: The Bitcoin community developed clustering heuristics to assign addresses to the same entity. We start with the most conservative Union-Find algorithm
  - All transactions that share inputs belong to the same entity

- Lower bound on the size of individual entities, since with a bit of effort a user can ensure any address is used only once and not mixed with other user’s addresses in the same transaction
Data construction (2)

- Link addresses to real-life entities using public and proprietary sources
  - Scrape cryptocurrency blogs and websites, such as Reddit, Blockchain.info, bitcointalk.org, etc
  - State-of-the-art database of crypto entities from Bitfury Crystal Blockchain, one of the leading providers of anti-money-laundering tools
- Most complete information about crypto entities that have been used in academic research
  - We cover 3,500+ of the largest entities
  - 700+ exchanges and OTC desks, 113 gambling sites, 57 on-line wallets, 120+ mining pools and miners, 400+ scammers, 1700+ dark net marketplaces and illegal services
Data construction (3)

- We develop methodology to obtain information about other important Bitcoin entities:
  - To make sure that we are not missing major players on the blockchain, we analyze the top 10,000 unknown clusters with the largest Bitcoin volume
    - We use the observed transaction patterns of known exchanges and OTC brokers to infer which unidentified clusters are likely exchanges or OTC brokers
  - To analyze ownership concentration we need to differentiate between addresses belonging to individual investors and those belonging to intermediaries
    - We use graph analysis and examine utilization pattern to separate intermediary and individual accounts
  - We identify miners by tracing rewards distribution of the largest mining pools to individual miners
Transaction Volume and Network Structure
Spurious Volume

- The UTXO design and the preference of many of users for anonymity lead to spurious volume: transactions where an address sends its balance to itself or to another address controlled by the same entity.

- The number of new addresses or nominal volume might not be meaningful statistics.

Makarov, Crypto Data and Finance
Real Volume Decomposition

- Majority of volume on the Blockchain is for trading activity
Off-chain and one-chain exchange volume

- Trading volume of BTC against the four major fiat currencies and two stablecoins: USD, EUR, JPY, KRW, USDT, and USDC
- The weekly correlation is 44%
Bitcoin Network

- Exchanges are central entities and are highly interconnected

- Ex: Network of entities that receive > 500K BTC over 2018-2020
  - 18 exchanges, 3 online wallets, 2 unknown entities — likely large OTC trading desks
  - Almost a complete graph
Illegal Transactions (Hydra Market)

• A small share of total volume (< 3%) but not a trivial amount $\sim$ $2.4B in 2020

• Exchanges with lax KYC norms with serve as a gateway for money laundering and other gray activities
  • E.g. LocalBitcoins, Bitzlato, Binance
  • Once the flows arrive at these exchanges they get mixed with other flows and become virtually untraceable, and so can be sent anywhere afterwards

Ex: Hydra network: Retain only nodes that send >1000 BTC within the network
Ownership of Bitcoin
Ownership of Bitcoin

• Important to understand ownership and concentration of Bitcoin holdings
  • Determines who will benefit most from wider adoption. A select few investors or the general public?

• A challenging task:
  • More than just tracing “rich list” of addresses with large balances
  • Many addresses belong to exchanges and other intermediaries that hold bitcoins on behalf of many investors

• We use graph analysis and examine utilization pattern to separate intermediary and individual accounts
Intermediary Ownership

- As of Dec 2020, exchanges and other intermediaries held 5.5M BTC
Individual Ownership

- As of Dec 2020, individuals held 8.5M BTC
- High concentration of ownership:
  - top 1000 investors control around 3M BTC
  - top 10,000 - 5M BTC
Individual Ownership: Lost coins

- Some people might have lost their private keys
- Check when an address was used last time

- The balance of the smallest cluster in the 1000 list is 900 BTC and in the 10,000 list is 93 BTC
- Bitcoin reached $1000 for the first time in 2014 and stayed above $1000 since 2017.
Benchmarking Concentration

- Saez and Zucman (2020) show as of 2020, the wealth share of the top 1% households in the US is more than 35% of wealth, and the top 0.1% hold about 16%.

- Estimates from Crypto.com suggest 71 million holders of Bitcoin as of January 2021. Top 7,100 holders (0.01% of all Bitcoin holders) hold 4.5 million bitcoins, which is 24% of all bitcoins in circulation.

- Thus, the concentration of Bitcoin wealth is significantly higher than the wealth distribution of the US population.
Miners
Miners: Provide Verification of Transactions

- Mining is done in pools
  - Provide coinsurance by pooling capacity of miners
  - Highly concentrated
  - Majority of pools are registered in China
- But mining pools are not miners!
- Pools’ power depends on the size distribution of miners
- We identify miners by analyzing pool distributions ~ 250K miners
Mining Pool Distribution: AntPool Example
## Mining Pool Distribution: AntPool Example

<table>
<thead>
<tr>
<th>Input (1)</th>
<th>100.00000000BTC</th>
<th>Output (101)</th>
<th>99.99956991BTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3MtjFGtgtpwy3fRrUzZQMAJfPQ-4AMNgp</td>
<td>0.00172563</td>
<td>1y55YmkecvU6k306zWUQ19YqyXIn38N8cRus</td>
<td>0.00175333</td>
</tr>
<tr>
<td>15BR7zy77vB4v751y4yFm65p4ySVl5wU</td>
<td>0.00177196</td>
<td>1C4eLb2yb4G659Wu1bzq72tL3FGsKxide</td>
<td>0.00160417</td>
</tr>
<tr>
<td>1KBy6KpYe65p3T3j3ALAL6j3y7Xk3MqWcg</td>
<td>0.00172848</td>
<td>16ALt8twtwujh5h5j3J2FhA3n49q9qYg</td>
<td>0.00174608</td>
</tr>
<tr>
<td>125SwFvAbt3Lh3A5ZvW9z7Dk6jph74TfLhskh</td>
<td>0.0016176</td>
<td>19VYOLbsx5sd5DFH9h3dXk-jYePAFlq68b</td>
<td>0.00172531</td>
</tr>
<tr>
<td>82نوالو0الو8الو9ق4ل96زك7د5ق3jqFJv3k2x</td>
<td>0.0016536</td>
<td>141j19kqWFBbsx84J018F3FoP9h2qg1YChx</td>
<td>0.00174628</td>
</tr>
<tr>
<td>372xoC0gA64m6b3585shphBFrEl1p72bBQ</td>
<td>0.00165418</td>
<td>3dhaMuUI5G54dennccxx3c08d38mlvJ5dqb</td>
<td>0.00162476</td>
</tr>
<tr>
<td>3dPvP7.G05s15r4yly2u34pcQV8R2qCno</td>
<td>0.00172631</td>
<td>31xxm6x3c3gogvy3c0G6nWAWJ4jUbK65gK</td>
<td>0.00170177</td>
</tr>
<tr>
<td>1Q14m2h3br-crzD7n6o6f5FXvFwMuK6bqok</td>
<td>0.00163813</td>
<td>bc10m0q5jwaw7chitscaklatoomqnoes8qug6lc4df7</td>
<td>99.78074499</td>
</tr>
<tr>
<td>3JR2YZV5UIwT3uy65eb8f3l2f2zd0WWLJ</td>
<td>0.00165463</td>
<td>99.78074499BTC</td>
<td>99.78065393BTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input (1)</th>
<th>99.78074499BTC</th>
<th>Output (20)</th>
<th>99.78065393BTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GJN4u76oqzFPvH451fKnet8BFtUapp6rBQX</td>
<td>0.00123170</td>
<td>3KvQYAp6vKvCvoJb5Jr8wzy8D mistakes Z6eja0XG</td>
<td>0.00144571</td>
</tr>
<tr>
<td>169KjQhxxZ98z69Pox9A99qppc2pS6Z3A</td>
<td>0.00196242</td>
<td>1NCv5hPgyWAFzX39tAF4xot5B8If87vK6GW</td>
<td>0.00148628</td>
</tr>
<tr>
<td>17YknwzvGKS5il6x940Goonaie1ntk-vbhtmr</td>
<td>0.00123892</td>
<td>35v2hEJzBEm8Sb39GCF8p875y77mP4toGc</td>
<td>0.00123629</td>
</tr>
<tr>
<td>198J102taztuxQP5S5Q20aANDPuy1pWNk</td>
<td>0.00123892</td>
<td>3JdXz8H1Fyp4J436Y1QD.H0p8jz12v5k7t2</td>
<td>0.00123479</td>
</tr>
<tr>
<td>15pa0vny1Fy3u6j620vH3j4a09B97NLMsw9K</td>
<td>0.00244746</td>
<td>3BzjFvl43J3Pe1912xYR3qav4gpxzr183Qw</td>
<td>0.00104654</td>
</tr>
<tr>
<td>33Tfyl5SgAv'CrokQ5059OFeZqSS2RNPcF</td>
<td>0.00171861</td>
<td>1ALHyP9MMK3h07G2U6kipetxhxy vaccinations Dake</td>
<td>0.00121873</td>
</tr>
<tr>
<td>187yd6JUkDv37rrr7RvFJk6YUNwaF5PBAYEP</td>
<td>0.00123833</td>
<td>1LQ7hN8YbT44yLr3nWfHf4Q9j6gK9x0K</td>
<td>0.00122980</td>
</tr>
<tr>
<td>1JqGJFkZc63hj9A6ipK18v3d7588P5Kvy7wv6J</td>
<td>0.00125456</td>
<td>3Q2n6hFf66w4xoz-gJ8BmsrW6j3b5A14Gg</td>
<td>99.75468391</td>
</tr>
</tbody>
</table>

Makarov, Crypto Data and Finance
Concentration of Mining Capacity

- Mining capacity is concentrated
- Concentration varies with the Bitcoin price
- High concentration increases systemic risk and makes it easier to collude to ensure that miners continue to earn rents in equilibrium
Main Takeaways

- The majority of Bitcoin volume is for trading activity
- Exchanges are central entities on the blockchain
- The current regulation has limited effect on preventing tainted flows from entering into circulation
- The Bitcoin eco-system is still dominated by large and concentrated players:
  - Bitcoin ownership is concentrated
  - Mining industry is concentrated
Thank You!