

BLOCKCHAIN ANALYSIS OF THE BITCOIN MARKET



Antoinette Schoar

BLOCKCHAIN ANALYSIS OF THE BITCOIN MARKET



Antoinette Schoar
MIT Sloan, NBER, Ideas42

MOTIVATION

- Cryptocurrencies are no longer a niche market



B Bloomberg.com
Bitcoin (\$BTC USD),
Ether (\$ETH) Lead
Crypto to \$3 Trillion
Market Cap



US Public Pension Fund for Firefighters Adds Bitcoin and Ether Worth \$25 Million to Portfolio



- Many calls for even wider Bitcoin adoption
- But many open questions about the utilization of Bitcoin, its major players, potential risks, and spillover effects on the real economy

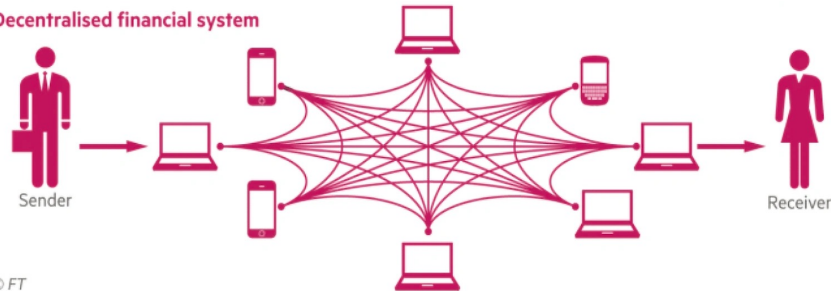


“TRUSTLESS TRUST ARCHITECTURE”

Traditional financial system



Decentralised financial system



© FT

- **Blockchain technology offers the possibility of a different financial architecture where record keeping is decentralized and access to the system is anonymous and unrestricted**
- **Key building blocks**
 - Blockchain: Open-source and permissionless ledger that provides decentralized record keeping
 - Smart contracts: Self-executing algorithms embedded in the blockchain



WHAT CAN WE LEARN FROM BLOCKCHAIN DATA?

- **Systematic analysis of the Bitcoin market using blockchain data**
 - Makarov and Schoar (2021)
 - A novel Bitcoin database and methodology to identify information about main market participants
- **Three major pieces of analysis:**
 - **Network structure:** Analyze transaction volume and network structure of 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 1K investors control 3M BTC, top 10K - 5M BTC
 - **Miners:** Study the concentration of miners who ensure the integrity of the Bitcoin blockchain
 - High concentration: 25 miners often control 50% of the total market



DIGITAL FOOTPRINT OF BITCOIN TRANSACTIONS

- **Bitcoins are stored in Bitcoin addresses**
- **Bitcoin transactions record how bitcoins move between Bitcoin addresses**

tx:a8178a7223372414ac060b4bba4b33b8b4847a756fa76a715af7fd11bfd143d5 1,388.19 ⁷⁸⁴⁰⁵⁹ BTC Fee: 0.00 ¹ BTC			
←prev tx	17A16QmavnUfCW11DAApiJxp7ARnxN5pGX	-1,388.19 ⁸⁸⁴⁰⁵⁹ BTC	3QKAn2B1uDquujLZnoynVoq1M9uac66Ysr 0.00 ⁷⁹⁵⁷⁵⁹ BTC
			1F8fDpYbMLMaz1tBEehqPJSN8XTL6t5TDz 0.01 ²⁴¹⁰⁰⁶ BTC
			17A16QmavnUfCW11DAApiJxp7ARnxN5pGX 1,388.17 ⁷⁴⁷²⁹⁴ BTC
tx:efb3f60304532ebc80163b5f375fa8a94a39a8b0807b99703b6b646c1f7af5bf 0.71 ⁹⁸¹⁸⁶⁸ BTC Fee: 0.00 ¹¹⁷⁴ BTC			
←prev tx	3EocBKm4AgtX6Bi7P8HokjoVhCTZNWNP5q	-0.32 ²⁵⁰²¹⁷ BTC	3JGTTdXUwfsDmb6mWptD4CaXYm6KTdfTPc 0.00 ⁸⁸²⁹⁹⁸ BTC
←prev tx	3KA8RU8rmQEMVwXSQPSBKqqewZw63Upfwx	-0.30 ⁸⁹⁷⁹⁹⁹ BTC	3HLUD8s8C4wKfaNTj1n9D3NKpNMLMgjWkk 0.71 ⁰⁹⁸⁸⁷ BTC
←prev tx	33QJtiYPkQzYf5BNbCztHWowhco77sg18g	-0.08 ⁵⁹²⁰³³ BTC	
←prev tx	14J6bdZhJPeiJPEwcTGd7RWXRjUsMoqfxN	-0.00 ³¹⁰³³⁴ BTC	
←prev tx	3DqmyW4NLHgNQPT7T3i2gstNe1sx88rKyaB	-0.00 ⁰⁴⁸⁶⁸⁵ BTC	

- **Two main challenges:**
 - Bitcoin addresses are easy to generate → potentially many addresses belong to the same entity
 - Link anonymous Bitcoin addresses to real-life entities



DATA

- **Obtain blockchain data using the open-source software of Bitcoin Core and use the BlockSci program to convert raw data into a database**
 - As of June 28, 2021, there are about 650 million Bitcoin transactions and 850 million unique addresses
- **Link addresses to real-life entities using public and proprietary sources**
 - Scrape cryptocurrency blogs and websites, such as Reddit, Blockchain.info, bitcointalk.org, walletexplorer.com, and Matbea.com
 - State-of-the-art database of crypto entities from Bitfury Crystal Blockchain, one of the leading providers of anti-money-laundering tools



DATA, CONTINUED

- **Most complete information about crypto entities that have been used in academic research**
 - We cover 1,043 of the largest entities
 - 393 exchanges, 86 gambling sites, 39 on-line wallets, 33 payment processors, 63 mining pools, 35 scammers, 227 ransomware attackers, 151 dark net marketplaces and illegal services



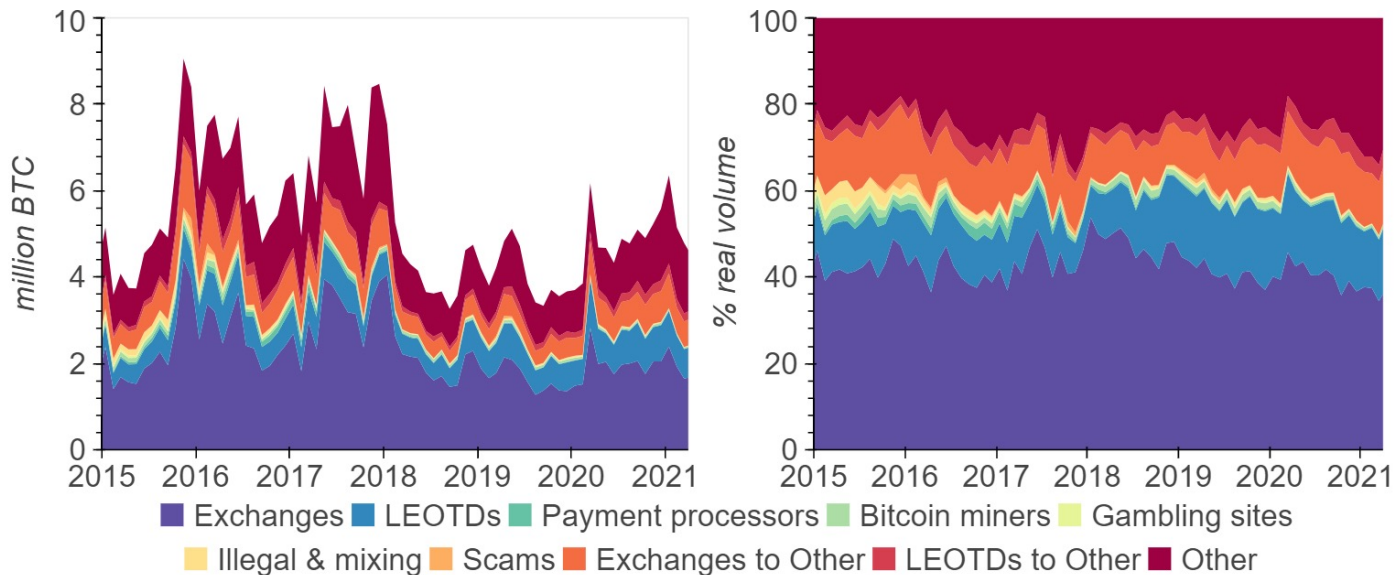
A black and white photograph of the MIT Sloan School of Management building. The building is a large, multi-story structure with a grid of windows. In the foreground, there is a stone wall with the text "MIT SLOAN SCHOOL OF MANAGEMENT" engraved on it. A person is walking in the foreground, blurred. A large red rectangular overlay covers the right side of the image, containing the text "TRANSACTION VOLUME AND NETWORK STRUCTURE" in white, bold, uppercase letters.

TRANSACTION VOLUME AND NETWORK STRUCTURE

MIT SLOAN SCHOOL OF MANAGEMENT

REAL VOLUME DECOMPOSITION

- Majority of volume on the Blockchain is for trading activity



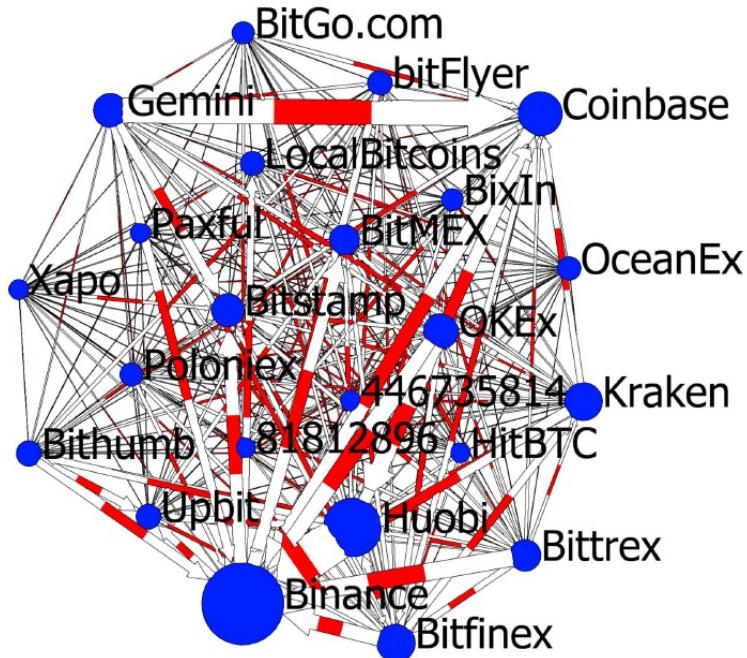
BITCOIN NETWORK

- **Trace the interconnectedness of the largest wallets on the network**
 - Focus on 10,000 highest volume clusters, which account for more than 55% of the total Bitcoin flows
 - Use snapshot from 2018 to the middle of 2021
- **To represent network, use a directed weighted network graph**
 - A node i corresponds to cluster i and an edge (or link) from node i to j corresponds to the total Bitcoin flows over the period 2018-2020 from cluster i to cluster j
 - For convenience only graph addresses that received more than 500,000 BTC



LARGEST ENTITIES

- Exchanges are central entities on the blockchain

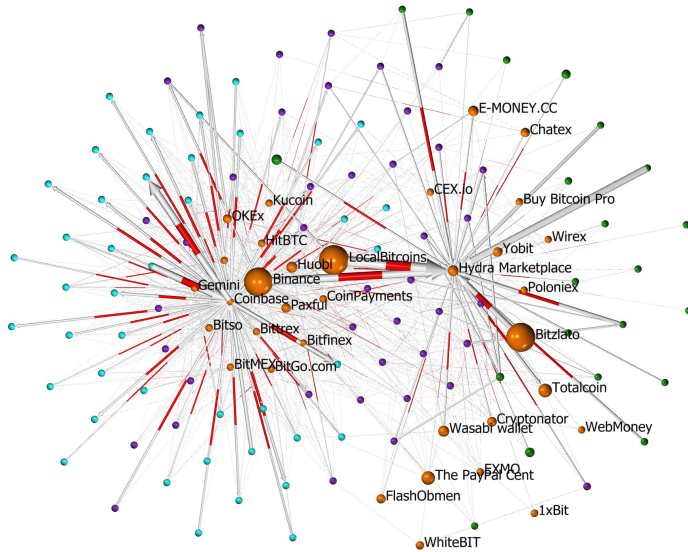


- Ex: Network of entities that receive
- > 500K BTC over 2018-2020
 - 18 exchanges, 3 online wallets, 2 unknown entities — likely large OTC desks
 - Almost a complete graph
 - Large volume between KYC (e.g., Coinbase and Gemini) and non-KYC (e.g., Binance) exchanges



ILLEGAL TRANSACTIONS

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



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

- Highest volume entities interacting directly with dark net marketplaces are non-KYC exchanges
 - E.g. LocalBitcoins, Bitzlatto, Binance
 - But once the flows arrive to these exchanges they get mixed with other flows and become virtually untraceable, and so can be sent anywhere afterwards
- Direct interactions of dark net marketplaces with exchanges that enforce KYC norms are small, but their interaction with the neighboring “mixing” clusters is large



IMPLICATIONS FOR KYC REGULATION

- **Current situation: KYC entities are allowed to accept flows from entities with lax-KYC norms**
- **Digital footprint has limited effect on preventing tainted flows from entering into circulation**
 - Even if KYC entities were restricted to deal exclusively with other KYC entities, preventing inflows of tainted funds would require placing severe restrictions on who can transact with whom
 - Every transaction would have to be subject to the approval of Bitcoin analytics companies e.g. Bifury Crystal Blockchain and Chainalysis. They would be de facto trusted parties
- **The wider the adoption of Bitcoin is, the easier it will become to use it for illegal transactions or tax evasion without ever having to touch regulated entities**





OWNERSHIP OF BITCOIN

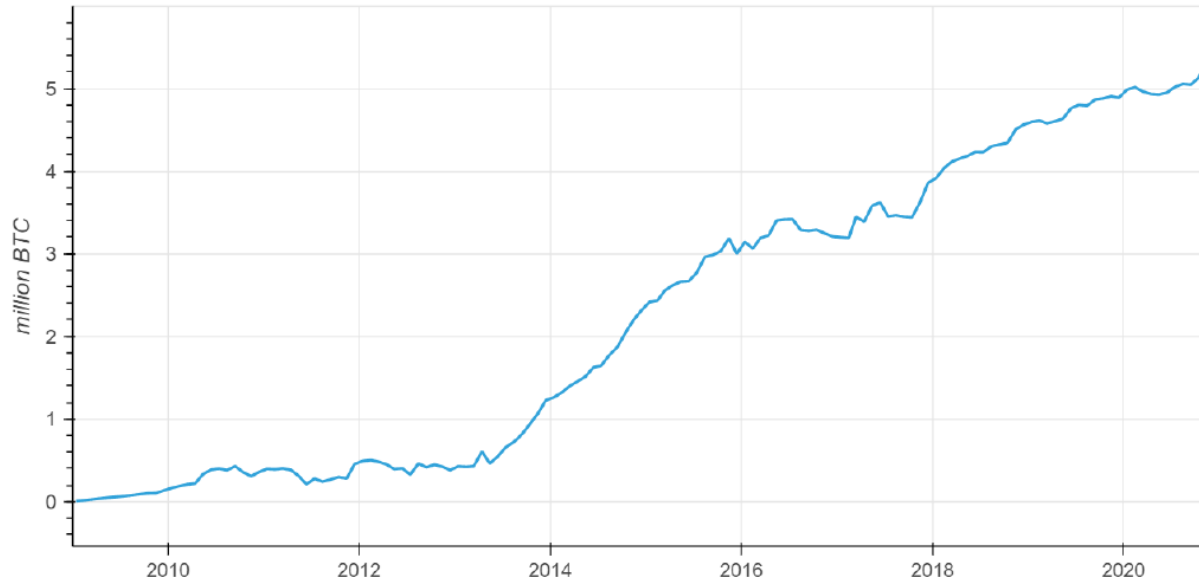
MIT SLOAN SCHOOL OF MANAGEMENT

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**



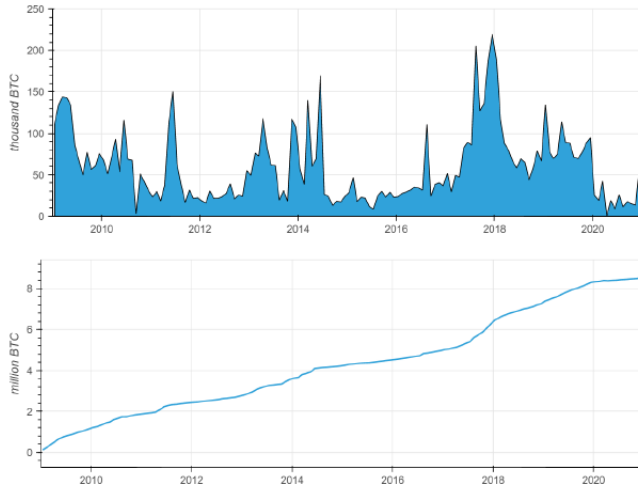
INTERMEDIARY OWNERSHIP



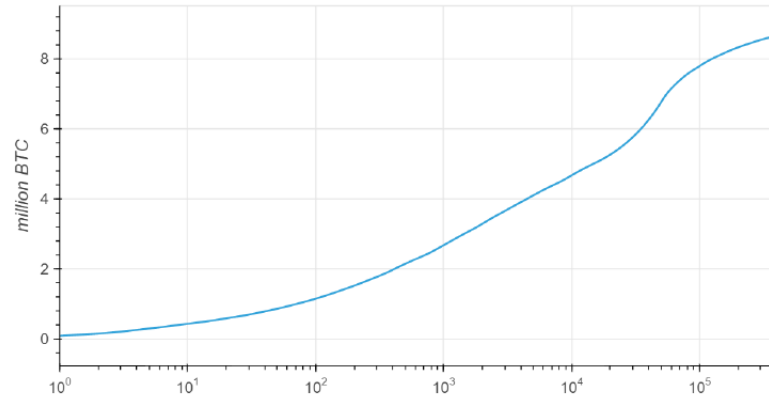
- **As of Dec 2020, exchanges and other intermediaries held 5.5M BTC**



INDIVIDUAL OWNERSHIP



early vs late adopters



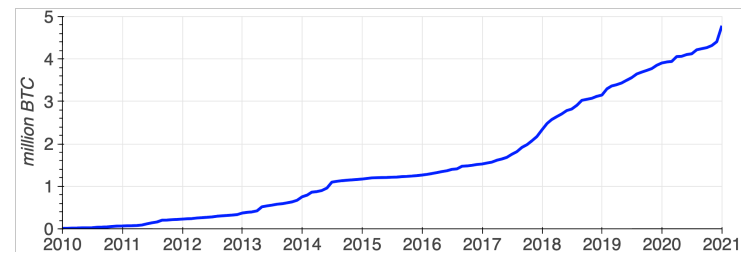
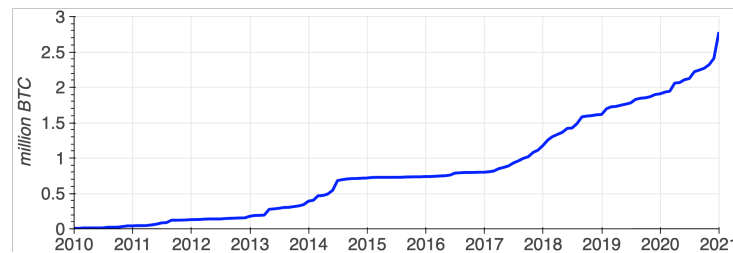
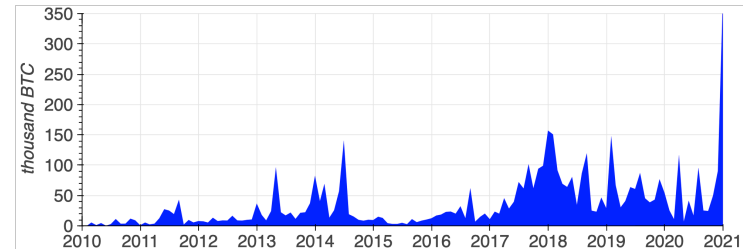
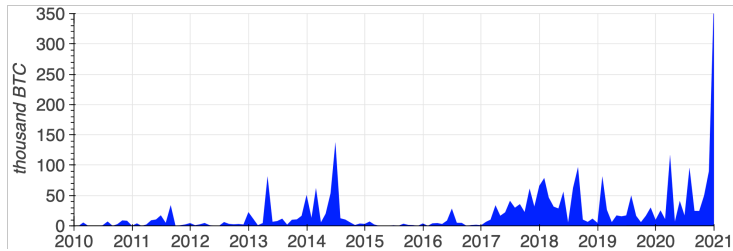
ownership concentration

- 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



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 10,000 holders are about 0.014% of Bitcoin holders. Hold 4.8 million bitcoin, which is 26% of Bitcoin wealth.**

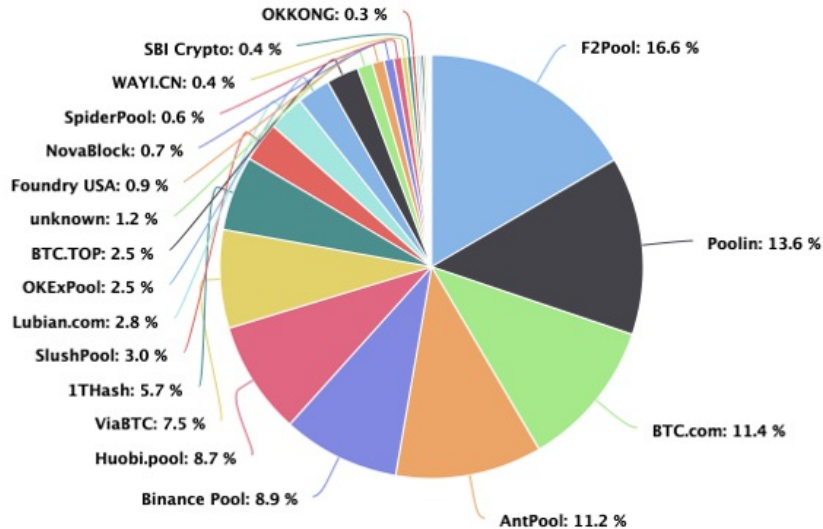




MINERS

MIT SLOAN SCHOOL OF MANAGEMENT

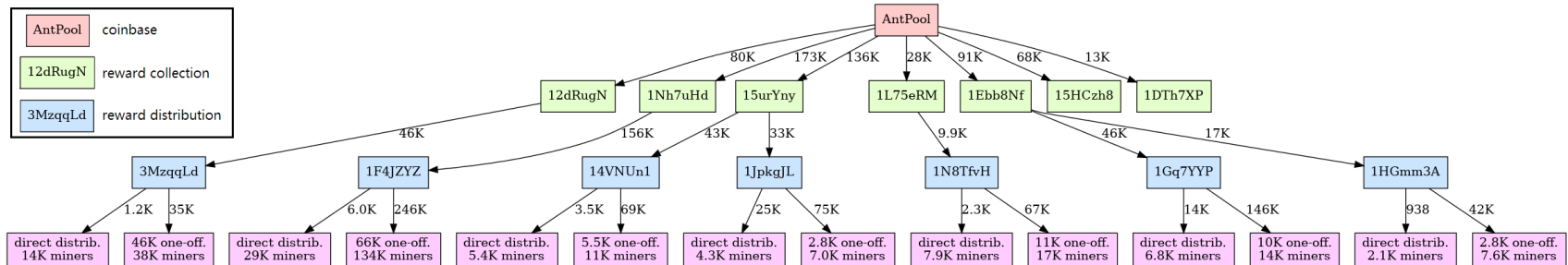
MINERS: PROVIDE VERIFICATION OF TRANSACTIONS



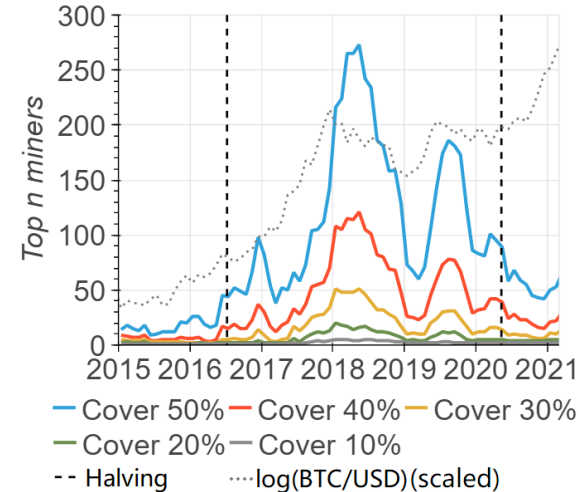
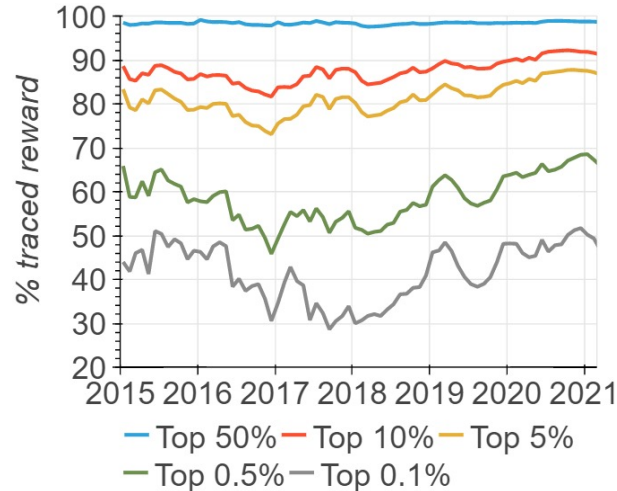
- **Mining is done in pools**
 - Provide coinsurance by pooling capacity of miners
 - Highly concentrated
 - Majority of pools are registered in China
- **But mining pools are not miners!**
- **Pools' power depends on the size distribution of miners**
- **We identify miners by analyzing pool distributions 250K miners**



MINING POOL DISTRIBUTION: ANTPOOL EXAMPLE



CONCENTRATION OF MINING CAPACITY



- Mining capacity is concentrated
- Concentration varies with the Bitcoin price
- We estimate that about 70% of miners are located in China based on the geographic location of exchanges where miners cash out their rewards (pre June 2021)
 - Use Xinjiang mining incidence (April 17-18, 2021) to verify our results



MAIN TAKEAWAYS

- **The majority of Bitcoin volume is for trading activity**
- **Exchanges are central entities on the blockchain**
- **Bitcoin ownership is concentrated**
- **Mining industry is concentrated**
- **The current KYC regulation has limited effect on preventing tainted flows from entering into circulation**





MIT
MANAGEMENT
SLOAN SCHOOL

THANK YOU

Antoinette Schoar

aschoar@mit.edu

in @ f t y w