An empirical analysis of pool hopping behavior in the Bitcoin blockchain

Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, Petra Isenberg

An Empirical Analysis of Pool Hopping Behavior in the Bitcoin Blockchain

Natkamon Tovanich
IRT SystemX and Université Paris-Saclay, LISN
Palaiseau, France
natkamon.tovanich@irt-systemx.fr

Nicolas Soulié
Université Paris-Saclay, Univ Evry, IMT-BS, LITEM
Evry-Courcouronnes, France
nicolas.soulie@imt-bs.eu

Nicolas Heulot
IRT SystemX
Palaiseau, France
nicolas.heulot@irt-systemx.fr

Petra Isenberg
Université Paris-Saclay, CNRS, Inria, LISN
Gif-sur-Yvette, France
petra.isenberg@inria.fr

***Abstract*—We provide an empirical analysis of pool hopping behavior among 15 mining pools throughout Bitcoin's history. Mining pools have emerged as major players to ensure that the Bitcoin system stays secure, valid, and stable. Individual miners join mining pools to benefit from a more predictable income. Many questions remain open regarding how mining pools have evolved throughout Bitcoin's history and when and why miners join or leave mining pools. We propose a heuristic algorithm to extract the payout flow from mining pools and detect the pools' migration of miners. Our results showed that payout schemes and pool fees influence miners' decisions to join, change, or exit from a mining pool, thus affecting the dynamics of mining pool market shares. Our analysis provides evidence that mining activity becomes an industry as miners' decisions follow classical economic rationale.**

***Index Terms*—Bitcoin, Bitcoin mining, mining pools, pool hopping, visual analytics**

I. Introduction

Bitcoin mining as a term refers to Bitcoin's *proof-of-work protocol*. Nakamoto proposed the protocol to solve the *double-spending* problem in digital currencies and to prevent individuals from tampering with the blockchain [1]. Bitcoin miners compete to solve a computation-intensive task to propose a new block in the network. They receive financial rewards for each block they successfully *mine*. These rewards consist of the *block reward* fixed by the protocol and *transaction fees* from transactions in the mined block. Practically, individual miners receive a reward only occasionally, relative to their computational power. As more miners join the network [2] and with faster mining hardware available [3], the total computational power of miners (called the *hash rate*) has been growing rapidly [4]. The *mining difficulty* is set by the protocol relative to the hash rate [5] so the probability of mining a new block with the same hardware becomes lower as hash rate increases. Therefore, the expected reward of an individual miner diminishes with more competition. In order to overcome this problem, mining pools have emerged in which miners combine computational resources to gain a more stable and predictable income.

This research work has been carried out under the leadership of the *Institute for Technological Research SystemX*, and therefore granted with public funds within the scope of the French Program *Investissements d'Avenir*.

This is the author's version of the article that will appear in 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC).

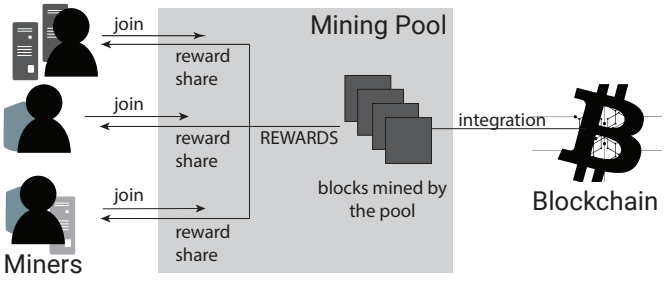

Fig. 1. The overview of Bitcoin mining activity. The diagram shows the interactions between individual miners, mining pools, and the Bitcoin protocol.

Nowadays, mining pools dominate the Bitcoin mining activity as known mining pools contribute ≈99% of the total hash rate [6]. Bitcoin mining is an economic activity with three main agents: mining pools, individual miners, and the Bitcoin network as shown in Fig. 1. First, *mining pools* are in competition with one another to attract more miners in order to gain a higher market share and have a better chance to obtain a mining reward. Mining pools distribute the rewards they collect to their individual miners and keep some amounts of rewards for their profits (*pool fee*). Second, *individual miners* invest in computational resources to receive a mining reward. They decide to enter a pool (*new miners*), move to another pool (*pool hoppers*), exit from mining activities (*dropout miners*), or even participate in multiple pools (*cross-pooling miners*). The migration of miners directly affects mining pools' market shares. Third, *the Bitcoin network* regulates the mining activity by automatically adjusting the mining difficulty every 2,016 blocks (≈2 weeks) according to the total hash rate in order to maintain the block discovery rate at 10 minutes. The network also pays the reward to the successful miners and sets the block reward. The block reward started with 50 BTC per block from *the genesis block* (block 0) and is reduced by half every 210,000 blocks (≈4 years).

Documenting the evolution of mining pools is of major interest for evaluating the future of Bitcoin, and more generally, of blockchain-based technology. A better understanding of the impacts on pools' growth or decline of both external (e. g., market price, regulation) and internal (e. g., payout scheme, pool fee, share of transaction fee) factors is critical to assess

the viability of this industry, and thus the sustainability of *proof-of-work* cryptocurrencies such as Bitcoin.

Mining pools are major players in Bitcoin mining that influence how stable, secure, and trustworthy the currency is. Joining or leaving a pool is a decision miners make not only to increase their short-term mining rewards but also to counteract the possible domination of pools in the network. In return, migration flows affect how mining pools set their policies (e. g., payout schemes and pool fees) to compete in the market. The combination of these activities and behaviors related to mining are not yet well understood and few methods exist that allow to study mining pools. We contribute to a new approach to detect miners' migration (called *pool hopping*) among mining pools. First, we developed an algorithm to extract reward payout flows based on the concept of *transaction purity*. Then, we obtained a list of miners for each payout flow and detected the miners that migrated between pools. Next, we derived quantitative measurements to evaluate miners' migration flows (e. g., *pool hopping* and *cross-pooling*). Finally, we visualize miners' migration among 15 mining pools throughout Bitcoin's history. We highlight the existence of regular patterns of miners' entry, hopping and cross-pooling behaviors in comparison to different payout schemes and pool fees.

II. Related Work

Some previous theoretical work related to pool hopping analysis has applied game theory to explain the motivation behind pool selection and miners' migration [7].

Lewenberg et al. [8] showed that miners are motivated to switch between pools to increase their expected rewards due to non-linear reward payout incentives and communication delays between mining pools. Schrijvers et al. [9] compared the payout schemes between proportional and Pay-Per-Last-N-Shares (PPLNS) in an optimized incentive compatibility condition. They showed that only PPLNS is incentive compatible. Liu et al. [10] considered the hash rate and the block propagation delay as metrics in their analysis. They found that miners' strategies will converge at the market equilibrium when there is a dominating strategy and no miner can switch pools. A recent work by Altman et al. [11] studied non-cooperative game competition over mining resources with constrained resource allocation. Their model suggests that only two major mining pools would dominate the network, unless the market is not stabilized or miners are not fully rational.

In contrast, our work relates more closely to other empirical work that has attempted to detect miners in mining pools and to analyze miners' migration patterns among pools in the Bitcoin network.

Belotti et al. [12] investigated pool hopping between KanoPool and SlushPool between April 6–20, 2016. The authors found that a few miners tried to exploit the time difference of reward payout between two pools with diverse strategies to gain a small profit gain. Romoti et al. [13] presented reward payout flow patterns of three pools: BTC.com, AntPool, and ViaBTC, between block 510,000 and 514,032 (≈4 weeks) and detected overlapped miners in those pairs. They detected high cross-pooling between BTC.com and AntPool probably because both pools are owned by the same company, Bitmain. Xia et al. [14] developed a visualization tool showing the internal address networks of mining pools and the estimated number of pool hoppers. This past work is closely related to ours but we deviate in several areas. We propose a miners' migration flow model and measurements to detect different types of pool hoppers over long time intervals. Our extraction method is also less computationally expensive approach although it shares a similar underlying concept to detect miners.

Overall, compared to the majority of past work, we look empirically at a higher number of mining pools across Bitcoin's mining history. Moreover, we compare our result with external information (e. g., market shares, payout schemes, and pool fees) to help explain pool hopping behavior.

III. Pool Hopping Detection

Our research approach involved extracting members of mining pools and measuring pool hopping behavior. Our process includes three steps: (A) we obtain the mining reward for each block and attribute it to a known mining pool; (B–C) for each coinbase transaction, we extract the reward payout flow and detected pool members (miners) of the mining pool; and (D) we identify miners who participated in 15 mining pools over time and migrated between pools. We provide the data about mining pool attribution and miner's migration at [15].

A. Mining pool attribution, market shares, and characteristics

Mining pool attribution. First, we identified the mining pool that mined each block in the Bitcoin blockchain. When a mining pool mines a block, it receives the mining reward from the *coinbase transaction* of the block. A coinbase transaction combines the block reward from the Bitcoin network and transaction fees from every transaction in the mined block. It also includes a *coinbase string* inserted by the miner. For each coinbase transaction, we attributed the mining pool based on the address matching or coinbase string pattern. We initially used the dataset from Romiti et al. [13] that compiled known mining pool tagging until block 556,400 (2018-12-31). After this block, we continued their procedure and tagged pools until block 650,731 (2020-09-30) with the datasets from Blockchain.info [16] and BTC.com [17]. We labeled the blocks that did not match any known mining pool as "unknown." We stored this data as an "attribution table." Table I shows the total number of blocks we found for each mining pool.

Mining pool market shares: We define the market share of a mining pool as the percentage of the blocks it mined compared to the total blocks mined in a month. We used our attribution table to calculate each pool's monthly market share.

Mining pool characteristics: We obtained information about pool characteristics, in particular payout schemes and pool fees, from the Bitcoin Wiki page [18] on the topic. We downloaded the page's edit history and manually cleaned the data for each month by comparing it with the information from

TABLE I
LIST OF MINING POOLS THAT RECEIVED MORE THAN 1,000 BLOCK REWARDS. ROWS HIGHLIGHTED IN GREEN COLOR ARE THE 15 MINING POOLS WE STUDIED IN THIS WORK.

Mining Pool	Blocks	Total Rewards	First Block	Last Block
F2Pool	58,174	1,101,804	2013-05-05	still active
AntPool	51,211	883,692	2013-12-07	still active
SlushPool	33,640	755,910	2012-01-26	still active
BTC Guild	32,936	1,010,779	2011-11-10	2015-06-30
BTC.com	31,814	403,201	2016-09-05	still active
DeepBit	31,107	1,508,254	2011-02-25	2013-11-28
GHash.IO	23,083	579,128	2013-08-04	2016-10-24
BitFury	20,901	420,027	2013-11-14	2020-03-25
ViaBTC	18,640	239,882	2016-06-05	still active
BTCC Pool	18,036	363,493	2014-10-21	2018-09-25
BTC.TOP	15,896	211,808	2016-12-11	still active
Poolin	15,142	175,971	2018-07-02	still active
BW.COM	12,733	250,044	2015-01-29	2018-11-09
Eligius	11,430	338,236	2011-06-14	2017-11-22
50BTC	7,859	198,651	2012-12-18	2014-06-02
KnCMiner	7,477	185,427	2014-02-25	2016-09-12
BitMinter	6,464	205,382	2011-11-07	2019-08-12
EclipseMC	6,024	212,395	2012-02-10	2016-03-02
Huobi	5,904	65,243	2015-06-01	still active
Bixin	5,753	80,640	2016-06-12	2019-09-01
BitClub Network	5,672	88,892	2015-04-02	2019-11-06
OzCoin	4,845	187,123	2011-12-30	2014-09-24
ASICMiner	3,146	79,279	2013-05-20	2014-04-09
1THash&58COIN	3,067	30,635	2019-08-25	still active
okpool.top	2,750	27,820	2018-12-27	still active
Bitcoin.com	2,465	32,944	2016-09-21	still active
KanoPool	2,432	46,359	2014-10-14	2020-07-26
GBMiners	2,093	28,980	2016-08-30	2018-04-15
DPOOL	1,918	24,398	2018-03-31	2019-05-11
1Hash	1,895	29,268	2016-03-04	2017-12-07
Telco 214	1,830	40,109	2014-12-19	2017-08-23
CloudHashing	1,824	45,745	2013-10-04	2015-02-14
21 Inc.	1,508	37,996	2015-04-09	2016-03-01
WAYI.CN	1,306	16,364	2018-02-14	still active
Polmine	1,290	32,365	2013-04-12	2015-06-07

the Bitcoin Forum [19]. As a result, we constructed panel data that includes all changes in pool characteristics over time.

Finally, we merged the mining pool market share data with the Wiki data. For many mining pools in the attribution table, the Bitcoin Wiki did not contain additional data. Therefore, we selected the top 15 mining pools that we found in both the Wiki data and the attribution table for our study on pool hopping behavior (highlight in green color in Table I).

B. Transaction flow and transaction purity

We introduce transaction flow graphs and the transaction purity definitions before applying them to our payout flow model and heuristic algorithm.

Definition 1. A *transaction flow* is a directed graph of Bitcoin transactions from a seeding transaction. Each node represents a transaction tx in the transaction flow. A transaction has a timestamp attribute $time$. Each directed edge corresponds to a value transfer from a transaction to another. Therefore, whether it is the input or the output of a transaction depends on the direction of the edge. An edge contains the information about the amount of transferred $value$, and the public-key address of the $owner$. Each edge contains references to the receiving transaction node $receive$ and spending transaction node $spend$.

We adopted a *transaction purity* measure to determine how much Bitcoin value in the transaction is received from the seeding transaction. This measure is commonly used for taint analysis in Bitcoin (e. g., [20], [21]).

Definition 2. Let $tx.in$ and $tx.out$ be sets of receiving (inputs) and spending (outputs) edges of a transaction tx respectively. The *transaction purity* is recursively defined as being the average purity of the input transactions weighted by their respective values. The purity of a transaction tx can be expressed as follows:

$$purity(tx) = \frac{\sum_{e \in tx.in} purity(e.receive) \cdot e.value}{\sum_{e \in tx.in} e.value} \quad (1)$$

The purity of a transaction without inputs is 1 because it is the root transaction in the transaction flow.

C. Mining pool payout flows

After a mining pool receives the mining reward from a coinbase transaction, the pool has to distribute the reward to pool members. Even though mining pools distribute the reward to individual miners in different patterns [13], [22], we introduce the payout flow model as a transaction graph consisting of four transaction types: coinbase (● $tx_{coinbase}$), payout (● tx_{payout}), intermediate (● tx_{inter}), and miner (● tx_{miner}). Examples of reward payout flows are shown in Fig. 2.

1) A mining pool receives mining rewards from coinbase transactions ● $tx_{coinbase}$ and collects them in a payout transaction ● tx_{payout} before distributing it to miners.
2) A mining pool distributes the reward from ● tx_{payout} to intermediate transactions ● tx_{inter} before splitting rewards to pool member (miner) addresses.
3) Pool members receive the reward from ● tx_{inter} and spend it in a transaction we call miner transaction ● tx_{miner}. We assumed that pool members receive the reward from this flow and then combine it with other Bitcoin values outside the flow to spend in ● tx_{miner}. Therefore, the purity of ● tx_{miner} is < 1.

Based on this model, the *reward payout flow* is the Bitcoin transaction flow from a payout transaction ● tx_{payout} to pool members ● tx_{miner} . We considered ● tx_{payout} as the seeding transaction because it collects every mining reward and distributes it to pool members.

Extracting reward payout flows. We devised Algorithm 1 to automatically extract payout flows from the coinbase transactions in the Bitcoin blockchain. We used the BlockSci API [23] to access the transaction data. We initiated the list of ● tx_{payout} from all outputs of ● $tx_{coinbase}$ as inputs to the algorithm. For each ● tx_{payout}, we traversed the transaction graph from ● tx_{payout} which has $purity = 1$ until the transaction has $purity_{tx} < 1$ (i. e. ● tx_{miner}). The algorithm returns a directed edge list that represents the payout flow.

We added two additional termination criteria $valid(tx)$ that stop following the current transaction tx flow: (1) when the

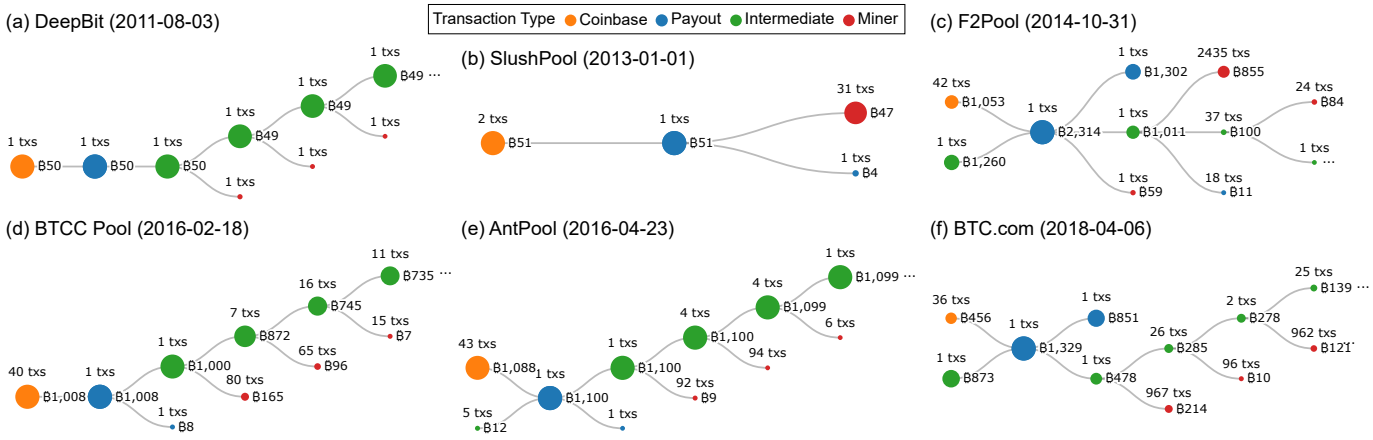

Fig. 2. Representative examples of reward payout flow patterns from the mining pools in our study. The flows were extracted using our algorithm and are represented as node-link diagrams. Here, we sampled the payout flow in the month where the mining pool had the highest market share. Each node represents a transaction type with branches of similar patterns grouped together. The color of the node indicates the transaction type. The total value of transactions in each node is encoded by circle size in proportion to the tx_{payout} value. The number of transactions and their combined value are the top and the right labels for each node respectively. We omitted labels for combined values below 1 BTC.

Algorithm 1: Reward payout flow extraction

Input : tx_{payout} is a payout transaction as a seeding node of the payout flow.
Output : $edges$ is the edge list of the the payout flow.

```
queue ← PriorityQueue([tx_payout]);
edges ← List();
while queue is not empty do
    tx ← queue.pop();
    if purity(tx) = 1 and valid(tx) = True then
        for edge in tx.out do
            edges.append(edge);
            queue.append(edge.spend);
        end
    end
end
```

time difference between ● tx_{payout} and tx is more than one day and (2) when the $tx.value$ is $<$ 0.001 BTC—as most mining pools have a minimum payout value [12], [14].

Identifying individual miners. For each edge list obtained from Algorithm 1, we constructed a payout flow graph using the NetworkX library [24]. Representative payout flow patterns that we obtained from the algorithm are shown in Fig. 2. Next, we extracted the ● tx_{miner} and derived the list of miners from each payout flow graph.

Definition 3. *Miner transaction* ● tx_{miner} is a transaction in the payout flow graph that does not have any output in the payout flow graph $|tx_{miner}.out| = 0$. We tagged all input edge(s) of ● tx_{miner} as *owner* edges. The list of miners who received the reward from ● tx_{payout} is defined as $M_{tx_{payout}}$.

Some ● tx_{miner} transactions may be connected to the pool wallet to keep the represented value as profits for the pool or as deposits for the next payout, as illustrated in Fig. 2 (c) and (f). We detected ● tx_{miner} input edges that have the same *owner* addresses as the mining pool and assigned them as ● tx_{payout} to extract further reward payout flows.

Payout flow patterns: We visualized payout flow patterns for all mining pools in our study and show representative flows patterns in Fig. 2. Early mining pools, operated from 2011 until mid-2015, distributed the reward directly after every block it mined. We found two payout patterns in this period: (1) the long chain of payout flows distributing the reward to a single miner at each step, e. g., DeepBit (a), BTC Guild, and GHash.IO; and (2) the direct payout to miners after receiving the mining reward, e. g. SlushPool (b) and Eligius.

After mid-2015, most mining pools tended to collect the mining rewards in their wallet and to distribute the reward to miners regularly (i. e. daily). We also observed two payout patterns (1) the chain of payout flow distributing the reward to multiple miners at each step, e. g. AntPool (e), BTC.com (f), BTCC Pool, and Poolin; and (2) the indirect payout to miners on an ● tx_{inter}, e. g. F2Pool (c). We also noticed that F2Pool (c) and BTC.com (f) usually send half of their payout back to their addresses as a reserve to pay miners in the next payout.

D. Miners' migration between mining pools

To analyze miner migration between pools, we compared the list of miners who received rewards from each mining pool in a set time interval and calculated the intersection of miners between pools. We set the time interval to months to be able to analyze detailed patterns for the entire mining pool history.

Definition 4. Let t be a time interval where $t \in T = \{t_0, ..., t-1, t, t+1, ..., t_n\}$. The set of *miners in the mining pool* M^t_{pool} is the summation of the miner list $M_{tx_{payout}}$ for all payout transactions of a mining pool *pool* at time t.

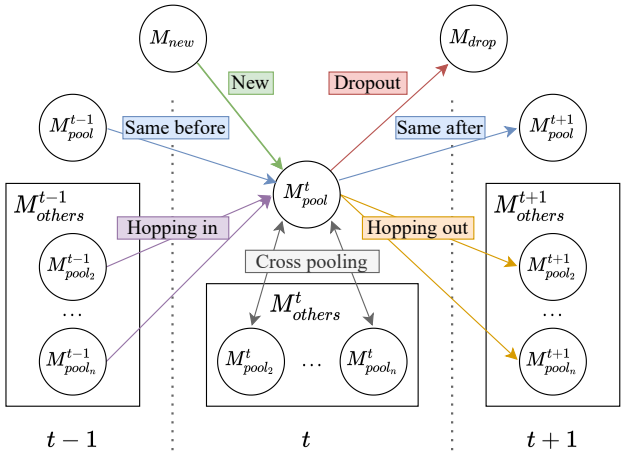

Fig. 3. The miners' migration flow model of *pool* at time interval t. $t-1$ ($t+1$) is the time interval before (resp. after) t. M_{new} (M_{drop}) is the list of miners not in any pool at $t-1$ (resp. is not found in any mining pool at $t+1$). The union of the list of miners from other pools is M_{others}.

The miner's migration flow is modelled as a diagram in Fig. 3. For each time interval t, the list of miners that migrate from/to a mining pool *pool*, M_{pool}^{t}, is divided into 7 *miner groups* as follows:

- *New (Dropout)* miners are miners that enter (exit) the mining activity at time t, annotated as $M_{new|pool}$ ($M_{drop|pool}$).
- *Same before (Same after)* miners are in M_{pool}^{t} but are also in M_{pool}^{t-1} (M_{pool}^{t+1}).
- *Hopping in (Hopping out)* miners are in M_{pool}^{t} but move from (to) other pools $M_{others|pool}^{t-1}$ ($M_{others|pool}^{t+1}$).
- *Cross-pooling* miners are in M_{pool}^{t} but also receive a reward from other pools at the same t ($M_{others|pool}^{t}$).

We estimated the quantity of miners' migration as the percentage of the total value for each miner group. We report the percentage of value rather than the number of addresses because it gives more weight to miners that have a high contribution to the pool and therefore the measure is more robust regarding small or occasional miners.

Definition 5. The *percentage of the total value of miners* (X) is the total value of M_{pool}^{t} associated with M_x, where x is a set of miners from miner groups. We defined this measure as:

$$X(M_{pool}^{t}, M_x) = \frac{\sum_{m \in M_{pool}^{t} \cap M_x} m.value}{\sum_{m \in M_{pool}^{t}} m.value} \quad (2)$$

For example, the percentage of *hopping in* (*hopping out*) miners is annotated as $X(M_{pool}^{t}, M_{others|pool}^{t-1})$ (resp. $X(M_{pool}^{t}, M_{others|pool}^{t+1})$).

For each mining pool, we obtained the monthly percentage of miners' migration for each miner group. To understand the flow of miners in a mining pool, we summarized *miners' migration flows* into a net gain or loss metric for the pool from different flow types with (1) *New and dropout flow:* the percent difference between new and dropout miners; (2) *Hopping in and out flow:* the percent difference between hopping in and hopping out miners; and (3) *Cross-pooling:* the percentage of cross-pooling miners.

Additionally, we calculated *the percentage of cross miners' rewards from the pool* as the total reward that cross miners received from the pool divided by the total reward that cross miners received from all mining pools. A higher percentage implies that miners dedicated more computational resources to this particular pool. It also indicates the attractiveness of the pool compared to other pools at the same time interval.

IV. Discussion on the Pool Hopping Detection

In this section, we discuss the validity and quality of our approach and compare it to related work. All confidence intervals are 95% bootstrap CIs.

A. Assumptions and limitations of the approach

Our approach rests on the assumption that individual miners who receive a reward share will spend it in a transaction that includes input transactions from outside the flow. The algorithm will make a false classification when a miner simply forwards the reward using a transaction without further inputs. In this case, the algorithm will calculate that the transaction purity is 1, assign it as tx_{inter}, and follow all outputs from $tx_{inter}.out$.

To detect migration patterns, we also assume that miner addresses are reused. We are aware, however, that miners can always generate new addresses. As a result, our percentage of new and dropout miners is an upper bound. In our data, each miner address received a reward from a pool on average 18.1 [15.4, 21.4] times. We summarize the *miners' migration flows* into net gain or loss metrics to reduce the impact of miners who change their addresses within the same month.

Although the basic address clustering method [25] is an effective method to group the addresses that are likely to belong to the same entity [26], we found that it led to false-positive clusters. For example, the method may group different miners in the same cluster because they used the same exchanges or mixing services. We expect that miners would participate in 1–2 pools at a time. We report the average number of mining pools that miners participated as the average weighted by their total reward. During the first halving (second halving) period, we found that miners receive the reward from 3.92 [3.12, 4.74] (3.06 [2.47, 3.68]) different pools per month compared to 1.46 [1.39, 1.56] (1.30 [1.24, 1.37]) pools per address. The percentage of cross-pooling per month using address clustering is on the average of 25.9% [24.2%, 27.6%] (31.6% [29.7%, 33.7%]) higher than using solely miner addresses, with pairwise comparison for the same pool and month. Therefore, we decided to use miner addresses to avoid adding errors from the address clustering to the results.

B. Miners addresses association with known entities

Since there is no ground truth to evaluate the identity of individual miners, we indirectly validated whether our approach can identify individual miners correctly. We posited

TABLE II
THE PERCENTAGE OF ADDRESSES AND TOTAL BITCOIN VALUES ASSOCIATED WITH KNOWN ENTITIES FROM 2013-01-01 TO 2016-12-31

Type	Addresses			Total Value		
	Input	Miner	Output	Input	Miner	Output
Unknown	96.1	84.8	91.5	44.0	84.8	68.6
Mining pool	1.38	0.555	7.43e-2	45.9	6.02	0.371
Exchange	1.42	8.36	4.80	0.35	6.52	18.0
Wallet	0.428	4.38	2.67	7.62	2.48	12.6
Marketplace	0.665	1.19	0.567	2.14	0.162	0.398
Gambling	3.46e-2	0.609	0.347	1.86e-4	3.25e-2	5.07e-2
Mixer	1.15e-3	5.51e-2	3.69e-2	1.16e-5	6.57e-3	3.57e-2
Lending	5.03e-3	4.54e-2	2.86e-2	3.68e-5	1.50e-3	1.51e-3

the assumption that miners should receive a mining reward (input address) from the mining pool and keep it in their wallet (miner address) before spending it (output address) on services (e. g. exchange, mixer, or marketplace). We used a known entity dataset from WalletExplorer.com with entity type classification from Zola et al. [27]. We report the percentage of addresses and Bitcoin values for each entity type in Table II. We studied the payout flows that spent between 2013-01-01 and 2016-12-32 because the website stopped updating more known entities from 2016 [28]. In summary, we found:

1) Miners detected from our algorithm mostly cannot be associated with any known address ("unknown" type in Table II) (84.8%) as well as input and output address (96.1% and 91.5% resp.). However, when we measured the total value for each entity type, we found that 84.8% of miner rewards are from unknown addresses, compared to 44.0% for input and 68.6% for output addresses. Therefore, we show that our algorithm can detect individual miners because they are largely not associated with any known Bitcoin entities.
2) Miners tend to receive a reward from known mining pool addresses (45.9% of the total value) followed by unknown addresses (44%). This result aligns with our assumption that miners should receive the money from tx_{inter} of the mining pool. For unknown addresses, mining pools may use external addresses that are undetected in the known entity dataset to pay miners.
3) Miners spent 68.6% of their total value using unknown output addresses. We also detected that some miners spent their reward on exchanges (18% of the total value) and wallet services (12.6%). This result provides evidence that regular miners convert mining rewards to fiat currencies or deposit them to their Bitcoin wallets.

Our approach differs from Xia et al.'s work [14] as we do not filter out known entities after we extract the payout flow. We have three reasons for this choice: (1) Xia et al. focus on only a 1-month time frame. The WalletExplorer dataset, however, includes 30,167,518 labeled addresses. It is computationally expensive to linearly scan for addresses in every transaction; (2) WalletExplorer does not update new entity labels after 2016 [28]. Hence, it cannot be applied to recent reward payout flows; and (3) Our measurements based on the percentage of value are tolerant to possible misclassification of miners.

V. ECONOMIC ANALYSIS OF POOL HOPPING BEHAVIOR

We calculated miners' migration statistics for 15 pools that adopted three main payout schemes: Proportional, Pay-Per-Share (PPS), and Pay-Per-Last-N-Shares (PPLNS). We explain miners' behavior in the Bitcoin network based on rational behavior in economic theory and provide visual evidence that some characteristics of mining pools (e. g. market shares, payout schemes, pool fees) affect miners' mobility.

Bitcoin mining has become an industry where miners gather into pools to maximize their investment in mining devices [29]. Choosing a pool becomes a strategic economic decision for miners as a pool's characteristics greatly affect a miner's income. First, we focus on the competition between pools based on payout schemes and transaction fees. Then, we investigate market entry and the expected revenue of new miners. Finally, we analyze miners' cross-pooling behavior that helps to diversify income and risks.

A. Pools' competition, fees and pool hopping

In the competition to attract miners, payout schemes and pool fees are major pool characteristics that directly impact miners' income. We illustrate that fee and payout schemes exhibit the usual economic evolution observed in the competition context in Fig. 4 (a). Our previous work [30] showed that the Proportional payout scheme was used at the beginning of Bitcoin. Over time, mining pools switched progressively to PPS and PPLNS payout schemes. As PPS and PPLNS are more robust to pool hopping than the proportional reward [31], these payout schemes are more attractive for pool managers. Our empirical result is in line with prior work as Proportional payout scheme disappeared in 2013. After that, PPS and PPLNS became the dominant payout schemes among the large pools.

A second explanation for the growing use of PPS and PPLNS relies on their different but complementary risk/return ratios. PPS pools pay miners in proportion to their contribution to the pool and thus provides risk-free, low income. All the risk is supported by the pool, which needs then to create a reserve of money to be able to pay the miners during 'bad luck' periods. In comparison, PPLNS pools pay only those miners who contributed to the last N shares in a given time window. Miners who contribute but leave the pool before a block has been mined might not get any reward. Therefore, PPLNS lefts all the risk to the miners, and the expected reward variance is higher compared to PPS [31][1]. These two payout schemes can be viewed then as two different *financial assets*. For this purpose, it is noticeable that the fees applied to these two financial assets follow the classical *two-parameter financial asset pricing model* [32]. In financial markets, risky assets

[1]Following Rosenfeld's article [31], the expected rewards of PPS and PPLNS are both equal to $(1-f)pB$, where B is the block reward, f is pool fee, and p is the probability of a share to be a valid one. However, as we showed in this paper, the PPLNS fee is lower than the PPS fee. PPLNS tend then to generate, in the long run, higher income than PPS. Moreover, PPLNS reward variance can be approximated following [31] by $\frac{pB^2}{N}$, using previous notations and N is the total number of share in a round, while each share sends to the pool in the PPS scheme is rewarded by a fixed amount, leading to no or insignificant reward variance.

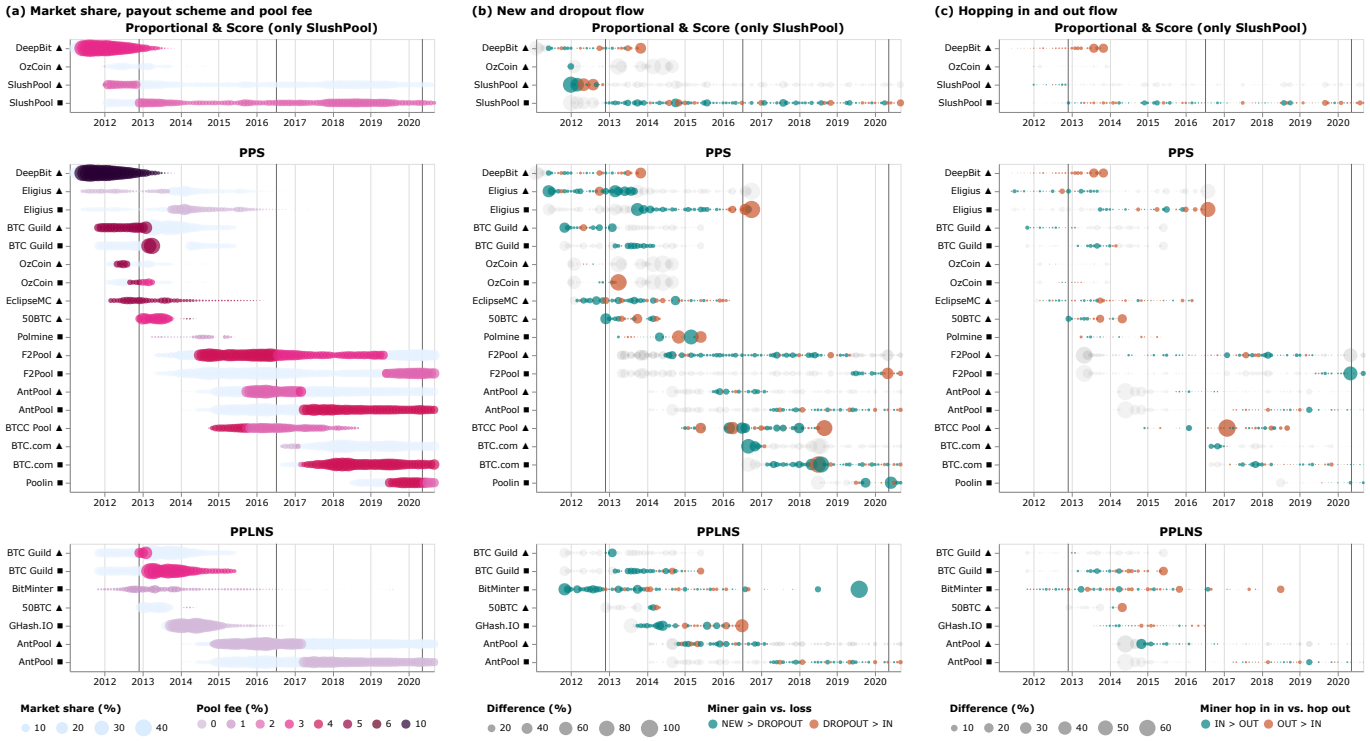

Fig. 4. Mining pool characteristics and miners' migration statistics over time. We divided into three main payout schemes: Proportional (and Score only for Slush Pool), Pay-Per-Share (PPS), and Pay-Per-Last-N-Shares (PPLNS). Each row in a graph represents the mining pool and the shape encodes whether the pool kept transaction fees for itself (squared ■) or shared with its miners (triangle ▲). We separated the same mining pool in different rows and facets but provided the shadow colors (blue in a, grey in b and c) to highlight the continuity of the pool with different payout schemes and transaction fee policies. (a) Market shares and payout schemes. The market share of the mining pool is represented as the size of the circle for each month. Pool fees are encoded as the color scale. (b) New and dropout flow. The size of the circle represents the absolute difference between new and dropout miners. The positive (or negative) flow of new miners are encoded as green (or red) color. (c) Hopping in and out flow. The size of the circle represents the absolute difference between hopping in and hopping out miners. The color indicates whether hopping in miners are more than hopping out miners (green) or vice versa (red). Three grey vertical lines in each chart indicate halving days on 2012-11-28, 2016-07-09, and 2020-05-11.

must have a higher expected return to be attractive. In the case of Bitcoin mining, Fig. 4 (a) is consistent with this scheme as the more risky asset (PPLNS) is likely to have a lower fee (≈0%) compared to the risk-free one (PPS, ≈2-3%).

Pool fees are used as a competitive advantage for mining pools. Within each payout scheme type, new pools tend to apply a lower fee than the incumbents. For instance, DeepBit applied a relatively high fee for PPS (10%) as the first dominant mining pool between 2011–2012. In 2012, mining pools, such as BTC Guild or OzCoin, applied lower PPS fees (5%) to attract new miners (Fig. 4 (b)) and hopping-in miners (Fig. 4 (c)), probably from DeepBit which had more hopping-out miners in the same period. We see the same pattern in 2013 when F2Pool (4%, named Discus Fish at the time) or 50BTC appeared (3%), then in 2014 with AntPool (2.5%) or BTCC (2%), and in 2016 with BTC.com (1.5%). This competition led to a decrease in the average PPS fees implemented by pools which stabilized around 2% from 2016. The same dynamics occurred for PPLNS pools. While BTC Guild has applied a 3% fee since 2011, 50BTC created in 2012 applied a lower fee (2.5%). This trend got stronger with GHash.io (0%) in 2013 or AntPool (0%) in 2014. When these pools appeared with lower fees, new miners were attracted by those pools (Fig. 4 (b)) and hopped out from older pools (Fig. 4 (c)).

Summary: The market share of mining pools is a confounding factor with miner flows. Mining pools that gain market share tend to attract new and hopping-in miners. Miners drop out and hop-out from pools that lose market share. This feedback loop probably explains the domination of a few mining pools at a time. The main driver of pool-hoping we observe in this article is the gap between pools fee for a given reward scheme. New successful pools adopted lower fees to attract miners while the older ones declined or stopped operating if they did not follow this trend. After 2015, pool fees tended to converge for each reward scheme, and pool-hoping decreased.

B. *Bitcoin value and mining market entry*

Another evidence for the economic rationale of mining activities comes from the incoming flow of miners during bitcoin's high valuation periods. Fig. 4 (b) provides information about the new miners joining Bitcoin mining pools. Bitcoin experienced 8 local high valuation periods before 2020: 06/2011 ($19), 04/2013 ($130), 12/2013–03/2014 ($800), 06/2014 ($600), 12/2015 ($420), 06/2016 ($630), 05–12/2017 ($15,000) and 06–12/2019 ($10,500) [33]. These periods are characterized by many new miners entering pools and even

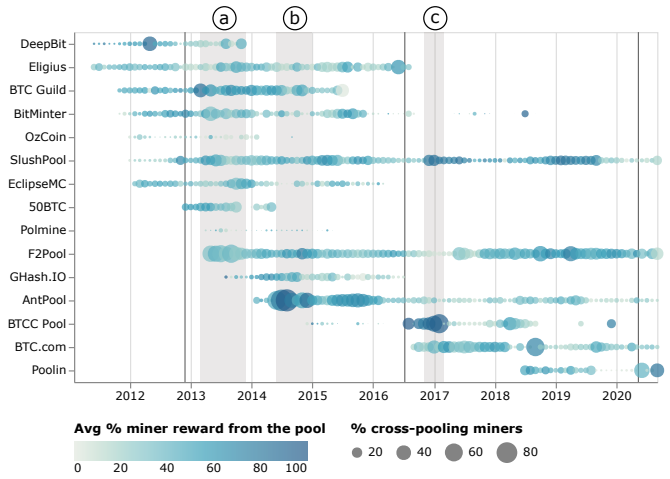

Fig. 5. Cross-pooling miners. Each row represents the percentage of cross-pooling for each mining pool over time. The percentage of reward that cross-pooling miners obtained from the pool is encoded as the color scale. Three selected periods that we focused on are highlighted in the grey background.

pool creations. Bitcoin mining became an economic investment as demonstrated in Prat and Walter [29] and acquisition of mining hardware tends to increase when its expected return rises. The corollary is that miners exit the market when Bitcoin value decreases. Fig. 4 (b) shows indeed large miner dropout close to halving days, which correspond to periods of sharp decrease in mining revenue.

C. Bitcoin values, income optimization and cross-pooling

Cross-pooling allows to diversify risks and optimize income. Fig. 5 (c) provides interesting insights into cross-pooling practice. Three periods of intense cross-pooling can be observed: 04/2013–11/2014, 06/2014–12/2014, and 11/2016–01/2017. These three periods exhibited an economic rationale where miners seem to diversify their risk between risk-free pools (PPS) and more risky ones (PPLNS).

In particular, the first two periods corresponded to a similar pattern, which is the rise of cross-pooling from PPS pools toward PPLNS ones. The first period (04–11/2014) corresponded to the switch from PPS toward PPLNS scheme for BTC Guild. BTC Guild applied PPS until February 2013, then proposed PPS and PPLNS until March 2014, and after offered uniquely PPLNS. Before that time, cross-pooling was very low. It seems that the availability of PPLNS reward has attracted miners operating on pools using PPS (F2Pool or Eligius) or Score (SlushPool) as shown in Fig. 6 (a).

The second period (06–12/2014) followed a similar pattern, except for AntPool (a newly created pool). Before the creation of AntPool, cross-pooling was limited. Once AntPool launched the PPLNS payout scheme (in addition to PPS), cross-pooling rose considerably. Fig. 6 (b) shows that cross-pooling occurred between the main PPS (Eligius and F2Pool) and PPLNS pools (AntPool, BTC Guild and GHash.IO).

The last period of intense cross-pooling (11/2016–01/2017) followed the same rationale but the timing was different. In this period, neither the creation nor the switch toward a PPLNS pool led to cross-pooling. However, the apparition of a large PPS pool (BTCC Pool) generated a lot of cross-pooling with an already existing PPLNS pool (AntPool). Fig. 6 (c) shows that a large flow of cross-pooling existed between AntPool and BTCC Pool, and also other PPS pools (BTC.com or F2Pool).

In all the cases reviewed above, we demonstrated that cross-pooling is used to diversify miners' risk and leads them to combine mining in risk-free pools (PPS) and more risky ones (PPLNS). In this respect, individual miners seem to act as a portfolio manager who optimizes their income concerning the risk associated with each type of asset.

VI. Conclusion and Future Work

We contribute a new approach to extract reward payout flows of mining pools and to detect individual miners from the payout flow. We provide rationale and evidence that our approach based on the general payout flow model can be used to extract payout flows with different payout patterns. We also propose the miner's migration flow to measure the mobility of miners who enter, exit, hop, or cross between mining pools.

Based on these algorithms and metrics, we provide an analysis of miners' migration among 15 mining pools across Bitcoin's history. The visualizations allow us to highlight regular patterns of miners' entry, hopping, and cross-pooling behaviors. These regularities are consistent with classical economic behaviors under competition. Our work provides new empirical evidence that miners and mining pools behave as typical economic agents, seeking to maximize their profits. We show in particular that pools' competition is based especially on pool fee and payout schemes. The most popular current payout schemes, namely Pay-per-Share (PPS) and Pay-per-Last-N-Shares (PPLNS), can be seen as two different financial assets from a miner's viewpoint: a free-risk scheme (PPS) and a more risky one (PPLNS). Consistent with the economic rationale, the more risky one is associated with more expected returns due to the lower pool fee associated with PPLNS. Moreover, miners tend to perform cross-pooling with pools applying different reward schemes seemingly for risk diversification and income optimization. This is especially the case during high Bitcoin value periods where expected income from mining gets higher, in particular for the PPLNS payout scheme. These periods are then associated with important levels of new miner entries and cross-pooling from PPS pools toward PPLNS ones. Additionally, pool fees are major drivers of the pool's competition. We show that new pools tend to apply a lower fee with respect to incumbents. It leads to a decrease in the average pool's fee along time and is correlated with important pool hopping decisions toward new pools. Our result provides an insight into the dynamics of mining pools, which is crucial for improving regulations and policies in cryptocurrencies.

The data obtained from our approach provides rich and detailed information on miners' migration among pools. The data is publicly available at [15]. For future research, we are planning to develop a quantitative model to explain miners' decisions and dynamics of mining pool competitions. In

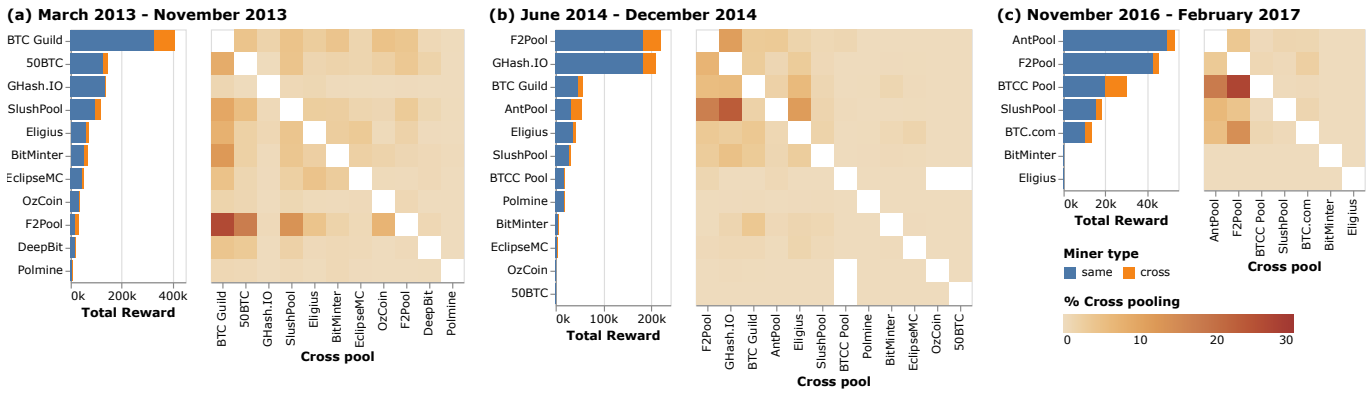

Fig. 6. Cross-pooling among mining pools during three periods: (a) March–November 2013, (b) June–December 2014, and (c) November 2016–February 2017. The stacked bar chart shows the total reward distributed from each mining pool during the period as a proxy for the pool size. Mining pools are sorted from the highest to the lowest value. Each bar is separated by the total amount of non-cross-pooling miners (in blue) and cross-pooling miners (in orange). The heatmap shows the percentage of the total cross-pooling miners from mining pools on the y-axis to other pools on the x-axis.

particular, the econometric analysis might be useful to check if the trends highlighted in this empirical work still hold when controlling for confounding variables. On the other hand, we are developing a visual analytics system to explore and monitor Bitcoin's mining activities in multiple coordinated views. The tool will allow mining pool managers, individual miners, and Bitcoin data analysts to analyze the miner's migration flow in granular details, and to relate the information to pool characteristics and other relevant Bitcoin indicators.

Acknowledgment

We thank Antoine Durand, Lorenzo Candeago, Suprawee Mekrungruangkul, Lucas Khornelord, Kahlifa Toumi, and Kalpana Singh for their valuable feedback. We also express our gratitude toward Malte Möser who maintains BlockSci and Aleš Janda (WalletExplorer), Marc Jourdan, and Francesco Zola for providing the known entity dataset.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Tech. Rep., 2008. [Online]. Available: http://bitcoin.org/bitcoin.pdf

[2] L. Wang and Y. Liu, "Exploring miner evolution in bitcoin network," in *Passive and Active Measurement*. Springer, 2015, pp. 290–302.

[3] M. B. Taylor, "The evolution of bitcoin hardware," *Computer*, vol. 50, no. 9, pp. 58–66, Sep 2017.

[4] Blockchain.com, "Total hash rate of the bitcoin network," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://www.blockchain.com/charts/hash-rate

[5] BTC.com, "Bitcoin mining difficulty," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://btc.com/stats/diff

[6] ——, "Mining pool statistics," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://btc.com/stats/pool

[7] Z. Liu, N. C. Luong, W. Wang, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "A survey on applications of game theory in blockchain," *arXiv preprint arXiv:1902.10865*, 2019.

[8] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, "Bitcoin mining pools: A cooperative game theoretic analysis," in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '15. International Foundation for Autonomous Agents and Multiagent Systems, 2015, pp. 919–927.

[9] O. Schrijvers, J. Bonneau, D. Boneh, and T. Roughgarden, "Incentive compatibility of bitcoin mining pool reward functions," in *Financial Cryptography and Data Security*. Springer, 2017, pp. 477–498.

[10] X. Liu, W. Wang, D. Niyato, N. Zhao, and P. Wang, "Evolutionary game for mining pool selection in blockchain networks," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 760–763, 2018.

[11] E. Altman, D. S. Menasche, A. Reiffers, M. Datar, S. Dhamal, C. Touati, and R. El-Azouzi, "Blockchain competition between miners: a game theoretic perspective," *Frontiers in Blockchain*, vol. 2, p. 26, 2019.

[12] M. Belotti, S. Kirati, and S. Secci, "Bitcoin pool-hopping detection," in *IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2018, pp. 1–6.

[13] M. Romiti, A. Judmayer, A. Zamyatin, and B. Haslhofer, "A deep dive into bitcoin mining pools: An empirical analysis of mining shares," *arXiv preprint arXiv:1905.05999*, 2019.

[14] J.-z. Xia, Y.-h. Zhang, H. Ye, Y. Wang, G. Jiang, Y. Zhao, C. Xie, X.-y. Kui, S.-h. Liao, and W.-p. Wang, "SuPoolVisor: a visual analytics system for mining pool surveillance," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 4, pp. 507–523, 2020. [Online]. Available: https://doi.org/10.1631/FITEE.1900532

[15] N. Tovanich, N. Soulié, N. Heulot, and P. Isenberg, "Dataset: An empirical analysis of pool hopping behavior in the bitcoin blockchain (version 1)," Dec. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4342747

[16] Blockchain.info, "Blockchain known pools," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://raw.githubusercontent.com/blockchain/Blockchain-Known-Pools/master/pools.json

[17] BTC.com, "Blockchain known pools," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://raw.githubusercontent.com/btccom/Blockchain-Known-Pools-BCH/master/pools.json

[18] Bitcoin Wiki contributors, "Comparison of mining pools," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://en.bitcoin.it/wiki/Comparison_of_mining_pools

[19] Bitcointalk contributors, "Bitcoin mining pools discussion forum," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://bitcointalk.org/index.php?board=41.0

[20] G. Di Battista, V. Di Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia, "Bitconeview: visualization of flows in the bitcoin transaction graph," in *Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2015, pp. 1–8. [Online]. Available: https://doi.org/10.1109/VIZSEC.2015.7312773

[21] M. Ahmed, I. Shumailov, and R. Anderson, "Tendrils of crime: Visualizing the diffusion of stolen bitcoins," in *Graphical Models for Security*. Springer, 2019, pp. 1–12. [Online]. Available: https://doi.org/10.1007/978-3-030-15465-3_1

[22] X. F. Liu, X.-J. Jiang, S.-H. Liu, and C. K. Tse, "Knowledge discovery in cryptocurrency transactions: A survey," *IEEE Access*, 2021.

[23] H. Kalodner, M. Möser, K. Lee, S. Goldfeder, M. Plattner, A. Chator, and A. Narayanan, "Blocksci: Design and applications of a blockchain analysis platform," in *USENIX Security Symposium (USENIX Security)*, Aug 2020, pp. 2721–2738.

[24] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the Python in Science Conference*, 2008, pp. 11–15.

[25] F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in *Security and Privacy in Social Networks*. Springer, 2013, pp. 197–223.

[26] M. Harrigan and C. Fretter, "The unreasonable effectiveness of address clustering," in *Proceedings of the Conference on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*. IEEE, 2016, pp. 368–373.

[27] F. Zola, J. L. Bruse, M. Eguimendia, M. Galar, and R. Orduna Urrutia, "Bitcoin and cybersecurity: Temporal dissection of blockchain data to unveil changes in entity behavioral patterns," *Applied Sciences*, vol. 9, no. 23, p. 5003, Nov 2019.

[28] A. Janda, "Information about walletexplorer.com," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://www.walletexplorer.com/info

[29] J. Prat and B. Walter, "An equilibrium model of the market for bitcoin mining," CESifo, Tech. Rep. 6865, 2018. [Online]. Available: https://www.cesifo.org/en/publikationen/2018/working-paper/equilibrium-model-market-bitcoin-mining

[30] N. Tovanich, N. Soulié, and P. Isenberg, "Visual analytics of bitcoin mining pool evolution: on the road toward stability?" in *International Workshop on Blockchains and Smart Contracts (BSC)*, Apr 2021. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02902465v2

[31] M. Rosenfeld, "Analysis of bitcoin pooled mining reward systems," *arXiv preprint arXiv:1112.4980*, 2011.

[32] W. F. Sharpe, "The sharpe ratio," *The Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, 1994.

[33] CoinMarketCap, "Bitcoin market price," 2020, Accessed on: Dec 12, 2020. [Online]. Available: https://coinmarketcap.com/currencies/bitcoin/